1 **Representational dynamics during extinction of fear memories in the human**

2 **brain**

3 Daniel Pacheco-Estefan[1,2], Antoine Bouyeure[2], George Jacobs[2], Marie-Christin Fellner[2], Katia

4 Lehongre[3], Virginie Lambrecq[3,4,5], Valerio Frazzini[3,4,5], Vincent Navarro[3,4,5,6], Onur Güntürkün[7], Lu

5 Shen[8,9,10], Jing Yang[8,9,10], Biao Han[8,9,10], Qi Chen[8,9,10], Nikolai Axmacher[2]

6

7 Correspondence to: daniel.pacheco@uab.cat, nikolai.axmacher@rub.de, qi.chen27@gmail.com

8

9 [1] Department of Basic, Developmental and Educational Psychology, Faculty of Psychology, Autonomous
10 University of Barcelona, Spain
11 [2] Department of Neuropsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University
12 Bochum, 44801 Bochum, Germany
13 [3] Sorbonne Université, Paris Brain Institute – Institut du Cerveau, ICM, INSERM, CNRS, APHP, Pitié-Salpêtrière
14 Hospital, Paris France
15 [4] AP-HP, Département de Neurophysiologie, Hôpital Pitié-Salpêtrière, DMU Neurosciences, Paris, France
16 [5] AP-HP, Epilepsy Unit, Pitié-Salpêtrière Hospital, DMU Neurosciences, Paris, France
17 [6] AP-HP, Center of Reference for Rare Epilepsies, Pitié-Salpêtrière Hospital, Paris, France
18 [7] Department of Biopsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University
19 Bochum, 44801 Bochum, Germany
20 [8] Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, China
21 [9] School of Psychology, South China Normal University, Guangzhou 510631, China
22 [10] Center for Studies of Psychological Application, and Guangdong Key Laboratory of Mental Health and
23 Cognitive Science, South China Normal University, Guangzhou 510631, China

24

25 Extinction learning – the suppression of a previously acquired fear response – is

26 critical for adaptive behavior and core for understanding the etiology and treatment

27 of anxiety disorders. Electrophysiological studies in rodents have revealed critical

28 roles of theta (4-12Hz) oscillations in amygdala and hippocampus during both fear

29 learning and extinction, and engram research has shown that extinction relies on the

30 formation of novel, highly context-dependent memory traces that suppress the initial

31 fear memories. Whether similar processes occur in humans and how they relate to

32 previously described neural mechanisms of episodic memory formation and retrieval

33 remains unknown. Intracranial EEG (iEEG) recordings in epilepsy patients provide

34 direct access to the deep brain structures of the fear and extinction network, while

35 representational similarity analysis (RSA) allows characterizing the memory traces of

36 specific cues and contexts. Here we combined these methods to show that

37 amygdala theta oscillations during extinction learning signal safety rather than threat

38 and that extinction memory traces are characterized by stable and context-specific

39 neural representations that are coordinated across the extinction network. We further

40 demonstrate that context specificity during extinction learning predicts the

41 reoccurrence of fear memory traces during a subsequent test period, while

42 reoccurrence of extinction memory traces predicts safety responses. Our results

43   reveal the neurophysiological mechanisms and representational characteristics of

44   context-dependent extinction learning in the human brain. In addition, they show that

45   the mutual competition of fear and extinction memory traces provides a mechanistic

46   basis for clinically important phenomena such as fear renewal and extinction

47   retrieval.

48

49

50 *Introduction*

51 In a constantly changing world, the ability to update previously learned knowledge is

52 essential for adaptive behavior. The neural mechanisms that support this ability have

53 been extensively studied within the domain of extinction learning. Extinction learning

54 allows for the inhibition of previously acquired unconditioned responses (in particular,

55 fear responses) that have become dysfunctional; accordingly, deficits of extinction

56 learning are associated with numerous psychiatric conditions, most prominently

57 anxiety disorders and depression, and are targeted during exposure therapy

58 (Beckers et al., 2023; Bouton et al., 2021; Maren et al., 2013; McNally, 2007).

59 While early accounts of extinction learning suggested that it weakens the

60 association between a conditioned stimulus (CS) and an unconditioned stimulus

61 (US) (Rescorla & Wagner, 1972), more recent theories propose that it consists in the

62 formation of a new inhibitory memory trace that suppresses the initial fear response

63 (Bouton, 2004; Y. Liu et al., 2024; Maren et al., 2013; Maren & Quirk, 2004). This

64 new memory trace is highly context-dependent, as demonstrated by the

65 phenomenon of renewal, where fear responses re-emerge when tested either in the

66 initial (acquisition) context or in a new context (Bouton, 2004; Corcoran & Maren,

67 2004; Maren et al., 2013; Orsini et al., 2011; Orsini & Maren, 2012; Quirk & Mueller,

68 2008).

69 Human neuroimaging studies have provided a detailed understanding of the

70 circuits involved in fear learning and extinction, identifying a core network that

71 includes amygdala (AMY), hippocampus (HPC), and regions of prefrontal cortex

72 (PFC; Fullana et al., 2016, 2018). Among these regions, PFC and HPC appear

73 particularly relevant for the context specificity of extinction learning (Bouton, 2004;

74 Eichenbaum, 2017a, 2017b; Gilmartin et al., 2014; Maren et al., 2013; Maren &

75 Quirk, 2004; Orsini et al., 2011). Indeed, the PFC seems to be more involved in the

76 inhibition of fear responses (Gilmartin et al., 2014; Maren et al., 2013) and in the

77 contextual regulation of fear during and after extinction (Orsini et al., 2011) than in

78 the acquisition of novel fear memories. Furthermore, recent studies indicate that the

79 lateral PFC (lPFC) supports flexible, task-dependent updating of memory traces

80 (Pacheco-Estefan et al., 2024), consistent with its role in adaptive cognitive control

81 (Darna et al., 2024; Weber et al., 2023).

82   Research in rodents has provided important insights into the neural
83   mechanisms of fear acquisition, expression and extinction. Notably, several studies
84   have shown that the amplitude of theta (4-12Hz) oscillations in the prelimbic cortex,
85   dorsal anterior cingulate cortex (dACC), and/or amygdala increases during fear
86   acquisition (Paré & Collins, 2000) and expression (Courtin et al., 2014; Fenton et al.,
87   2014; Karalis et al., 2016; Lesting et al., 2011; Likhtik et al., 2014; Rahman et al.,
88   2018; Seidenbecher et al., 2003; Taub et al., 2018). Similarly, human EEG studies
89   have reported heightened frontocentral theta oscillations, particularly during fear
90   conditioning (e.g., Chen et al., 2021; Pirazzini et al., 2023) and fear recall (e.g.,
91   Bierwirth et al., 2021; Mueller et al., 2014; Sperl et al., 2019; but see (Bierwirth et al.,
92   2023). In addition, studies in rodents have shown that long range theta (2-12Hz)
93   connectivity across regions of the extinction network, including AMY, PFC and HPC,
94   has been associated with the discrimination of aversive and safe cues (Likhtik et al.,
95   2014), and with the extinction of fear associations (Lesting et al., 2011, 2013;
96   Nguyen et al., 2023).

97   In humans, the electrophysiological mechanisms underlying extinction
98   learning in AMY and HPC are largely unknown because these regions are difficult to
99   investigate non-invasively. Intracranial EEG (iEEG) recordings in epilepsy patients
100 provide invaluable data about neurophysiological processes in deep brain areas at
101 the highest temporal and spatial resolution possible (Axmacher, 2023; Parvizi &
102 Kastner, 2018). Indeed, iEEG research in the domain of episodic memory has
103 provided important insights into the mesoscopic neural mechanisms underlying
104 memory formation and retrieval (Burke et al., 2014; Fell et al., 2011; Foster et al.,
105 2012, 2015; Gattas et al., 2023; Lega et al., 2012; Miller et al., 2013; Seger et al.,
106 2023), including evidence for a prominent role of theta oscillations (Herweg et al.,
107 2020). Other recent iEEG studies described the neurophysiological mechanisms in
108 AMY and HPC underlying the impact of aversive emotions on episodic memory
109 formation and retrieval (Costa et al., 2022, 2024; Qasim et al., 2023; H. Zhang et al.,
110 2024; Zheng et al., 2017).

111 Despite this progress in understanding episodic memory and its modulation
112 by aversive emotions, surprisingly few iEEG studies explored the neurophysiological
113 mechanisms of fear learning and extinction, and thus it is unclear whether and to
114 which extent they rely on the same neural machinery that supports episodic memory.
115 A notable exception is the study by Chen et al. (2021) who reported increased theta

116   oscillations in AMY and dorsomedial PFC following aversive stimuli during fear
117   learning. Whether AMY and PFC theta frequency oscillations play a role during
118   extinction learning remains unknown.

119   In addition to neural oscillations, rodent research has described how memory
120   traces of specific experiences – engrams – are built in AMY, HPC, and other areas
121   during fear learning and how they are suppressed or modified during extinction (X.
122   Liu et al., 2012; X. Zhang et al., 2020). In humans, Representational Similarity
123   Analysis (RSA; Kriegeskorte et al., 2008; Kriegeskorte & Diedrichsen, 2019) is
124   increasingly used to identify stimulus-specific memory traces, i.e. the neural
125   representations of unique episodes and events, at a meso- and macroscopic level
126   (Griffiths et al., 2019; J. Liu et al., 2021; Michelmann et al., 2016; Pacheco Estefan et
127   al., 2019; Staresina et al., 2016; Yaffe et al., 2014). Applied to iEEG data, this
128   approach typically relies on patterns of the power of neural oscillations across
129   frequencies and electrodes (Manning, 2023; Pacheco Estefan, 2023; Xie et al.,
130   2023). RSA can then either be used to calculate the similarity of representations
131   between encoding and retrieval (reinstatement, or encoding-retrieval similarity) or to
132   measure similarities between representations of different items during either
133   encoding or retrieval.

134   Episodic memory studies using RSA revealed fundamental differences in the
135   "representational formats" of memory traces of item-context associations as
136   compared to single items: While memories of individual items depend on broad
137   frequency ranges and on sensory regions – for visual stimuli, in lateral temporal
138   cortex (TMP) – memory traces of item-context associations exhibit a strong reliance
139   on theta frequency oscillations and the HPC (Pacheco Estefan et al., 2019;
140   Staresina et al., 2016). Because of the strong context dependency of extinction
141   learning, one may expect similar representational signatures as for episodic memory
142   traces of item-context associations. Furthermore, RSA studies using functional
143   magnetic resonance imaging (fMRI) demonstrated that the "stability" of item
144   representations – i.e., the similarity of neural activity patterns across repeated
145   exposures – correlates with successful memory formation (Lu et al., 2015; Xue et al.,
146   2010). Similar effects were observed during fear acquisition, where higher levels of
147   item stability predicted learning success (Visser et al., 2013).

148   Here we investigated the neurophysiological mechanisms and
149   representational characteristics of fear and extinction memory traces and their

150   impact on subsequent renewal. We developed a novel fear and extinction learning

151   paradigm that we conducted with epilepsy patients (N = 49) who were implanted with

152   iEEG electrodes across the fear and extinction network (Figure 1A). During

153   acquisition, patients were exposed to images of three electric devices, two of which

154   were paired with an aversive stimulus (CS+) while the third one was not (CS-).

155   During extinction, the contingency of one CS+ cue changed, resulting in three

156   different types of cues: $CS+_{acquisition}/CS+_{extinction}$ (CS++), $CS+_{acquisition}/CS-_{extinction}$ (CS+-

157   ) and $CS-_{acquisition}/CS-_{extinction}$ (CS--) (Figure 1A, top). During the final test phase,

158   none of the items was paired with a US. Critically, to study the specificity of context

159   representations, the CS items were presented within four different thematically

160   related context videos in each experimental phase (e.g., four snow landscape videos

161   during acquisition, four videos during extinction, and four videos during test),

162   corresponding to an "ABC" paradigm (Figure 1A, top). This paradigm allowed us to

163   track the memory traces of items and contexts during fear acquisition, extinction, and

164   renewal across key regions of the extinction network including AMY, HPC, and PFC

165   regions, as well as sensory processing areas in TMP (Figure 1B).

166

167   *Results*

168   *Behavioral results*

169   We first analyzed the behavioral responses to the different cue types across

170   experimental phases. A two-way ANOVA with "cue type" (CS++, CS+-, CS--) and

171   "experimental phase" (acquisition, extinction, test) as repeated measures revealed

172   significant main effects of both cue type ($F_{(2, 94)}$ = 19.02; p = 1.16e-07) and

173   experimental phase ($F_{(2, 94)}$ = 9.03; p = 2.58e-04) as well as a significant interaction

174   ($F_{(4, 188)}$ = 6.96; p = 3.02e-05). Importantly, cues with changing contingency (CS+-)

175   were perceived as significantly less threatening during extinction than during

176   acquisition ($t_{(47)}$ = 5.67; p = 7.44e-06, Bonferroni corrected), confirming that

177   participants correctly adjusted their expectations of the US.

178   We evaluated the progressive learning of trial contingencies during acquisition

179   and extinction. We sorted trials according to their temporal position (i.e., trial

180   number) separately in the three conditions (CS++, CS+-, and CS--), and compared

181   responses at every position using paired Wilcoxon signed rank tests, followed by

182   cluster-based permutation statistics to correct for multiple comparisons (Methods).

183   During acquisition, ratings of CS++/CS+- vs. CS-- cues started to differ at trials 10

184   and 13, respectively (both p = 0.001; brown and grey lines in Figure 1C), while

185   ratings did not differ between CS++ and CS+- items, which were both threatening

186   during acquisition (red and yellow line, all p > 0.72). During extinction, ratings started

187   to differ between CS++ and CS+- cues at trial 9 (p = 0.001), indicating that

188   participants gradually learned the new contingencies, while they differed immediately

189   between CS++ vs. CS-- cues (p = 0.001; Figure 1C, orange and brown line). Ratings

190   of CS+- and CS-- cues did not differ significantly (all p > 0.46). During the test phase,

191   ratings of CS++ and CS-- differed in the first 3 trials, while the remaining contrasts

192   did not show any significant differences. However, comparisons of average ratings

193   during the test revealed significant differences between CS++ and CS+- trials (t(47)

194   = 2.51, p = 0.015) and between CS++ and CS-- trials (t(47) = 3.1, p = 0.003),

195   whereas no significant difference was observed between CS+- and CS-- trials (t(47)

196   = 1.13, p = 0.26). Taken together, these results show that participants accurately

197   learned the task contingencies and their changes throughout the experiment. In

198   addition, while differences in ratings diminished during the test phase, participants

199   still rated CS++ items on average as more threatening than CS-- or CS+- items,

200   indicating that fear responses did not completely disappear.

201

202   *Amygdala theta oscillations signal safety during extinction*

203   Previous literature has established a key role of theta (4-12Hz) oscillations in AMY,

204   HPC and PFC for fear learning and extinction, in both rodents (Seidenbecher et al.,

205   2003) and humans (Chen et al., 2021). We thus focused on this frequency band in

206   our initial analysis and compared differences in theta power between CS+ and CS-

207   cues. Following previous studies (e.g., Chen et al., 2021), we specifically assessed

208   the time period from cue onset until US presentation (in reinforced CS+ trials, or

209   corresponding time point in the other trials), during both acquisition and extinction.

210   During acquisition, we found no significant power differences between CS+

211   and CS- cues in any of our regions of interests (ROIs; all p > 0.54; see Figure 2A, left

212   for the results in the AMY). During extinction, theta power was significantly higher for

213   CS- as compared to CS+ trials in the AMY ($p_{corr} < 0.005$; Figure 2A, right), while no

214   significant power differences were observed in any other ROI (all $p_{corr} > 0.24$). This

215   AMY effect occurred at a relatively late time period preceding the US presentation

216   (i.e., from 1.18 to 1.75s) and was confined to the theta (3-10Hz) frequency range

217   (Figure 2A, right). Follow up analyses revealed that theta power was significantly

218   lower in CS++ trials compared to both CS+- (p = 0.005; Figure 2B, top) and CS--

219   trials (p = 0.026; Figure 2B, middle), with effects occurring in overlapping time

220   frequency periods to those observed in the "current valence" contrast in Figure 2A,

221   right. No significant differences were observed between CS+- and CS-- trials (p =

222   0.38; Figure 2B, bottom). To determine whether the difference between CS+ and CS-

223   trials was specific to the theta band, we analyzed all frequencies across the 1-100Hz

224   range. Results confirmed that differences between conditions were localized to the

225   same cluster observed in Figure 2A, right, while no significant effects were observed

226   in other time-frequency bins (p = 0.004; Figure 2C, top). Further analyses in this

227   cluster revealed that effects were driven by theta power increases above baseline for

228   both CS+- ($t(31) = 3.21$, p = 0.003) and CS-- items ($t(31) = 2.3$, p = 0.027), as well as

229   theta power reductions below baseline for CS++ items ($t(31) = -3.9$, p = 0.0004;

230   Figure 2C, bottom). As a baseline, we chose the average across all trials in all

231   phases (see Methods). No differences between CS+ and CS- trials occurred in this

232   time-frequency cluster during acquisition ($t(31) = -0.69$, p = 0.5). Please note that

233   these latter analyses are not circular because the time-frequency cluster is defined

234   by contrasting CS+ and CS- trials during extinction, and not by comparing the three

235   different trial types (CS++, CS+-, CS--) to zero. Together, these results show

236   increases of AMY theta power during extinction for CS- as compared to CS+ trials.

237   They also reveal theta power in CS- trials during extinction were systematically

238   above baseline, which may reflect a safety signal.

239

240   *Stable cue representations of CS+ items in TMP and AMY during extinction*

241   Previous fMRI research has shown that the stability of neural representations across

242   repetitions predicts both fear learning (Visser et al., 2013) and episodic memory

243   formation (Lu et al., 2015; Xue et al., 2010). We employed RSA to compare the

244   stability of CS+ vs. CS- cue representations during acquisition and extinction (Figure

245   3A), again focusing on the time period of cue presentation. We built representational

246   feature vectors based on the pattern of iEEG power values across frequencies (44

247   values in the 1-100Hz range) and across electrodes in each ROI. These patterns

248   were extracted in windows of 500ms, incrementing in steps of 50ms (90% overlap),

249   and were compared across all pairs of trials using Spearman correlations (see

250   Contrast-based RSA, Methods).

251 During acquisition, item stability did not differ between CS+ and CS- items in
252 any ROI (all p_{corr} > 0.66). During extinction, we observed significant increases in item
253 stability for CS+ vs. CS- cues in the AMY between 1.25-1.5s (p_{corr} = 0.018),
254 overlapping in time with the period of theta power increases for CS- items (Figure
255 3B, right). Follow-up analyses within this time period revealed that item stability was
256 significantly higher for CS++ as compared to both CS+- trials (t(31) = 2.30, p =
257 0.028; Figure 3C, left) and CS-- trials (t(31) = 2.39, p = 0.023; Figure 3C, middle). No
258 differences were observed between CS+- and CS-- trials (t(31) = 1.3, p = 0.2; Figure
259 3C, right). In the TMP, item stability during extinction was significantly higher for CS+
260 as compared to CS- trials between 0.65-1s (p_{corr} < 0.005; Figure 3D, right). Within
261 this time window, TMP item stability was significantly higher for CS++ as compared
262 to CS+- trials (t(40) = 2.46, p = 0.018; Figure 3E, left), but no significant differences
263 were observed between CS++ and CS-- trials (t(40) = 1.64, p = 0.11; Figure 3E,
264 middle) or between CS+- and CS-- trials (t(40) = 0.52, p = 0.61; Figure 3E, right). No
265 significant differences in item stability were observed in other ROIs during extinction
266 (all p_{corr} > 0.31).

267 We next investigated whether item stability was coordinated across brain
268 regions within the extinction network, analogous to the coordination of item-specific
269 representations between HPC and TMP during episodic memory retrieval (Pacheco
270 Estefan et al., 2019). Specifically, we tested whether trial-level metrics of item
271 stability were correlated between AMY and TMP – where condition differences
272 between CS+ and CS- items were observed – and between AMY or TMP and any
273 other ROI. For each trial, ROI, and time point, we averaged the similarity of the
274 representation of a particular item with the similarity of the same item in all other
275 trials (Figure 3F top left, Methods).

276 Item stability was significantly correlated between TMP and AMY (Figure 3F,
277 bottom left): Trials with high item stability in TMP also showed high item stability in
278 AMY in two temporal clusters, at the beginning of the cue period in both regions (0-
279 0.95s, p_{corr} < 0.005), and shortly before the presentation of the US (0.9-1.75s, p_{corr} =
280 0.01). In addition, item stability was coordinated throughout the extinction network
281 (Figure 3F, right): TMP item stability correlated with item stability in HPC, lPFC, and
282 OFC (all p_{corr} < 0.005). AMY item stability correlated with item stability in HPC (p_{corr} =
283 0.005), but not in prefrontal regions (lPFC: p_{corr} = 0.375; OFC: p = 0.25). Notably,
284 most effects occurred around the diagonal in the temporal generalization map,

285 demonstrating that item stability was coordinated at matching time periods across
286 regions. However, the time windows at which item stability was coordinated varied
287 across regions: While TMP-OFC correlations occurred early after cue onset, TMP-
288 HPC and AMY-HPC correlations occurred across the whole time period of cue
289 presentation, and TMP-lPFC correlations occurred close to the presentation of the
290 US. In supplementary analyses, we tested whether item-stability differed between
291 CS+ and CS- trials, and between CS++ and CS-- trials during extinction, and
292 observed no significant differences between these conditions (Supplementary Note
293 1). All relationships between all metrics are shown in the summary Figure 8 (see
294 below).

295

296 *Context-specific representations in HPC and lateral PFC*
297 Rodent studies have described the crucial role of PFC, in coordination with HPC and
298 AMY, in mediating the context dependency of extinction learning (Gilmartin et al.,
299 2014; Maren et al., 2013; Orsini et al., 2011). We investigated context-specific
300 representations in these regions and across the extinction network, by comparing the
301 similarity of representations of same versus different contexts (Figure 4A, Contrast-
302 based RSA analysis, Methods). We analyzed both the time period when only the
303 context was shown and the subsequent time period when context and cue were
304 shown conjointly, during both acquisition and extinction (Figure 4).

305   During the time period when only the context was shown, we observed
306 significant context-specific representations in the HPC during acquisition (0-500ms;
307 $p_{corr} < 0.005$; Figure 4B, top left) and in the lPFC during extinction (50-300ms, p_{corr} =
308 0.0245; Figure 4B, bottom right). No significant context-specific representations were
309 observed in the other ROIs during acquisition or extinction (all $p_{corr} > 0.06$).

310   During the subsequent period when both contexts and cues were shown, we
311 did not observe any significant context-specific representations during acquisition in
312 any ROI (all $p > 0.125$). During extinction, however, we observed significant context-
313 specific representations in lPFC (0.8s-1.15s; $p_{corr} = 0.012$), partially overlapping with
314 the time period showing item stability in TMP (Figure 4C, middle). lPFC context
315 specificity was significantly higher during extinction than acquisition ($p = 0.001$;
316 Figure 4C, right). No other ROI showed context specificity in this time window during
317 extinction (all $p > 0.23$).

318   To investigate whether context specificity was coordinated across the
319   extinction network, we again performed a trial-level correlation analysis between the
320   HPC, the lPFC, and other ROIs, consistent with our analysis of item stability
321   coordination (Figure 3). Specifically, we computed a trial-level metric of context
322   specificity by calculating the average difference in correlations between a given trial
323   and other trials with the same context vs. other trials with a different context (Figure
324   4D, left; Methods). We focused on the regions and time periods where we observed
325   significant context-specific effects in our previous analysis: the HPC when only the
326   video was shown during acquisition, and the lPFC when both video and cue were
327   shown during extinction. We correlated the values observed in these two regions
328   with those found in our other ROIs.

329   Context specificity in the HPC was not coordinated with any other ROI (all p_{corr}
330   > 0.06). During extinction, we observed a significant coordination of context
331   specificity in lPFC and TMP during an early time period (0-1.3s) and between lPFC
332   and AMY during a late time period (0.5-1.55s; all p_{corr} < 0.005; Figure 4D, rigth). Both
333   periods overlapped with the time period showing significant context specificity in
334   lPFC, and the time period of lPFC-AMY coordination also overlapped with the AMY
335   theta power effect. Correlations between lPFC and other ROIs were not significant
336   (all p > 0.24). We did not observe significant differences between conditions in the
337   coordination analysis of context-specificity (Supplementary Note 1).

338   Taken together, we found significant context-specific representations during
339   both acquisition and extinction, but in different brain regions (HPC during acquisition,
340   lPFC during extinction) and at different time periods (during acquisition: when the
341   context was presented alone; during extinction: when the context was shown with
342   the cue). During extinction, the magnitude of context specificity was coordinated
343   across the extinction network between the lPFC and regions showing theta power
344   differences (AMY) and/or differential item stability between CS+ and CS- (AMY and
345   TMP). In a next step, we evaluated whether the different representational metrics
346   were related.

347

348   *Links between amygdala theta power, item stability in AMY and HPC, and context*
349   *specificity in lPFC*
350   Our results so far indicate that, during the extinction phase, both sensory TMP
351   regions and AMY represent CS+ cues in a more stable way than CS- cues. They

352 also show that the lPFC exhibits context-specific representations in an overlapping

353 time period. In order to uncover the relationship of AMY theta power (AMY_{THETA}),

354 TMP/AMY item stability (TMP_{ITEM} and AMY_{ITEM}), and lPFC context specificity

355 ($lPFC_{CONTEXT}$), we investigated their correlation across trials, and extended this

356 analysis to item stability and context specificity effects in other ROIs during extinction

357 (Figure 5A). We extracted a single-trial metric of AMY_{THETA} based on the cluster of

358 significant differences between CS+ and CS- items observed during extinction

359 (Figure 2; see as schematic depiction in Figure 5A, left). We separately z-scored this

360 metric for CS+ and CS- trials, in order to avoid any spurious correlation of single-trial

361 values driven by main condition differences. We first focused on the time period of

362 significant AMY theta power effects and averaged item stability and context

363 specificity across this time period in each trial and each ROI.

364 We observed that across trials, AMY_{THETA} correlated with both AMY_{ITEM} (t(31)=

365 -9.08; p_{corr} = 1.54e-9; Figure 5B, left) and HPC_{ITEM} (t(23) = -5.61; p_{corr} = 5.25e-5;

366 Figure 5B, right), but not with item stability in any other ROI (all t < 0.79; all p > 0.44).

367 Notably, AMY_{THETA} also correlated with $lPFC_{CONTEXT}$ (t(14) = -4.37; p_{corr} = 0.0032;

368 Figure 5C). Context specificity in none of the other ROIs correlated with AMY_{THETA}

369 (all t < 1.43, all p > 0.16). To assess the temporal specificity of these effects, we

370 performed the same analysis across the entire cue presentation period, focusing on

371 the ROIs where we observed significant effects (i.e., AMY and HPC for item stability

372 analysis and lPFC for context specificity). Notably, we observed significant

373 correlations during the whole time period of cue presentation in AMY, which were

374 most pronounced when the AMY_{THETA} condition difference effects were observed

375 (i.e., from 1.18-1.75s after cue presentation; p_{corr} < 0.005; Figure 5D, left).

376 Correlations between AMY_{THETA} and HPC_{ITEM} reached significance from 600ms to

377 1.65s (p_{corr} < 0.005; Figure 5D, right), and correlations of AMY_{THETA} with $lPFC_{CONTEXT}$

378 were observed in an intersecting time period (from 900ms to 1.5s after cue onset;

379 p_{corr} < 0.005; Figure 5E). In supplementary analyses, we tested whether correlations

380 of theta power and item-stability or context specificity differed between CS+ and CS-

381 trials, and between CS++ and CS-- trials during extinction, focusing on the regions

382 where we observed significant effects in the analysis including all trials. Our results

383 revealed no statistical differences between conditions (Supplementary Note 2).

384 Taken together, these results demonstrate that AMY_{THETA} was correlated with both

385   AMY and HPC item stability and lPFC context specificity during a late time period of

386   cue presentation.

387

388   *LPFC context specificity predicts reinstatement of fear memory traces in TMP*

389   Both rodent studies (Gilmartin et al., 2014; Maren et al., 2013) and clinical studies in

390   humans (Craske et al., 2014; Garfinkel et al., 2014; Milad & Quirk, 2012; Wang et al.,

391   2024) suggest that high context specificity during extinction predicts renewal, i.e., a

392   reoccurrence of fear memories in acquisition contexts or novel contexts. We thus

393   quantified the degree to which the representations of the three cue types during

394   acquisition and extinction re-appeared during the test phase. Specifically, we

395   computed the similarity of every trial during acquisition (or extinction) with all trials in

396   which the same item was shown during the test phase. Averaging these values

397   across trials separately for each cue type (CS++; CS+-; CS--) and subtracting

398   acquisition reinstatement and extinction reinstatement yielded a subject-specific

399   metric of fear reinstatement (REINST = $REINST_{ACQ}$ − $REINST_{EXT}$; Figure 6A). We

400   correlated these reinstatement values in AMY and TMP, where the main effects of

401   item stability were observed, with subject-averaged metrics of context specificity in

402   lPFC during the extinction phase (Figure 6B).

403   First, we analyzed possible relationships between lPFC context specificity and

404   reinstatement in AMY or TMP during the time period when significant context-

405   specific effects were observed during extinction. As hypothesized, we observed a

406   positive correlation between context specificity in lPFC and reinstatement in TMP,

407   indicating more pronounced reinstatement of acquisition than extinction memory

408   traces in participants with more pronounced context specificity during extinction.

409   Interestingly, this effect occurred specifically for CS+- trials whose contingencies

410   changed from acquisition to extinction (Rho(15): 0.57, p = 0.024; Figure 6C, middle),

411   but not for CS++ (Rho(14) = -0.13, p = 0.3; Figure 6C, left) or CS-- trials (Rho(15) =

412   0.4, p = 0.13; Figure 6C, right). LPFC context specificity did not correlate with AMY

413   reinstatement for any cue type (CS++: Rho(13) = -0.01 p = 0.98; CS+-: Rho(13) =

414   0.029, p = 0.92; CS--: Rho(13) = 0.28, p = 0.32). This effect was not observed when

415   we correlated lPFC context specificity with either TMP $REINST_{ACQ}$ or $REINST_{EXT}$

416   separately for any of our three trial types (Acquisition: all p > 0.229; Extinction: all p >

417   0.56), suggesting that it specifically affects the balance (or competition) between

418   reinstatement of acquisition vs. extinction memory traces (see Discussion).

Next, we extended the analysis of reinstatement to all time points per trial for our three trial types (Figure 6D). We correlated lPFC context specificity with averaged levels of TMP reinstatement in windows of 400ms (same size as the lPFC context specificity effect), incrementing in steps of 50ms, and correlated these two metrics across subjects for each time window. We corrected for multiple comparisons using cluster-based permutation statistics, by shuffling the subject id labels (Methods). We observed a very specific window where increases in correlations were observed at trend level and specifically in CS+- trials, from 700ms to 1.05s (p = 0.055; Figure 6D, middle). These results suggest that the relationship between lPFC context specificity and TMP reinstatement in CS+- trials occurred selectively during the time periods of significant context specificity in lPFC.

In complementary analyses, we evaluated whether overall levels of acquisition-to-test, extinction-to-test and differential (acquisition-to-test minus extinction-to-test) reinstatement differed across trials, and whether for each cue type, acquisition-to-test and extinction-to-test reinstatement differed in the TMP and the AMY. We observed a trend for higher acquisition-to-test reinstatement than extinction-to-test reinstatement in the AMY specifically in the CS+- trials, suggesting that fear memories of CS+- items might show slightly higher levels of reinstatement than the subsequent extinction memories of these items (Supplementary Note 3 and Supplementary Figure 1).

Reinstatement of extinction memory traces in TMP predicts safety responses

In our final analysis, we evaluated whether reinstatement of extinction memory traces during the test phase predicted subjective ratings of safety. For every item during the test phase, similarity was calculated with all instances of this item during extinction, and these values were correlated across trials with the ratings during the test phase. We found that reinstatement of extinction memory traces was positively correlated with subjective safety ratings for CS+- cues (t(18)= 2.3; p = 0.0078), while this was not the case for CS++ (t(20)= 0.35; p = 0.72) or CS-- cues (t(22)= -0.16; p = 0.87).

Discussion

Here we investigated the neurophysiological mechanisms and representational signatures of extinction learning in the human brain. CS- items during extinction were

453 associated with increases in theta power in the AMY and lower levels of item stability
454 in the TMP. Moreover, extinction learning showed higher context specificity in lPFC,
455 which predicted the reinstatement of fear memory traces in TMP. Further analyses
456 unraveled the distinct pattern and time-course of coordinated representations across
457 the fear and extinction network, which is summarized in Figure 8.

458

459 *A safety signal during extinction in the AMY correlating with representational*
460 *changes of contexts and items*
461 While a previous human iEEG study reported increases in theta power for CS+ vs.
462 CS- items during acquisition (Chen et al., 2021), our results showed higher theta
463 power for CS- vs. CS+ items during extinction. This apparent discrepancy may be
464 explained by the differing cognitive demands of acquisition and extinction, with the
465 latter requiring pronounced coordination between AMY and other brain regions
466 coding for contexts and cognitive control such as HPC and PFC. Indeed, increases
467 of AMY theta power during extinction were correlated with higher levels of item
468 stability in both AMY and HPC and with higher levels of context specificity in lPFC.
469 The lack of AMY theta power effects during acquisition in our study contrasts with
470 findings in animals and humans, which have typically reported theta increases during
471 the formation of cue-US associations across several regions of the fear and
472 extinction network (e.g., Chen et al., 2021; Pirazzini et al., 2023). Several factors
473 may explain this discrepancy, including differences in recording methodologies
474 (iEEG versus EEG), the distinct regions targeted (e.g., different subregions of the
475 PFC targeted in [Chen et al., 2021]), interspecies differences in theta oscillations
476 (Jacobs, 2014), and the use of a more "cognitive" paradigm in our study with a milder
477 aversive stimulus.

478 In our paradigm, both CS+- and CS-- trials signal safety during extinction –
479 either newly learned safety following extinction (for CS+- trials) or consistent safety
480 throughout the experiment (for CS-- trials). Thus, the successful extinction of
481 previous fear should result in neural activity and representations of CS+- items that
482 match those of CS-- items—in line with our findings on both AMY theta power and
483 AMY and TMP item stability.

484 Notably, the increase in theta power for CS- items observed during extinction,
485 coupled with the absence of such increases during acquisition, suggests that theta
486 power does not simply signal safety but rather safety within the specific context of

extinction. At least two key differences distinguish acquisition from extinction in our paradigm: (1) Contingencies changed for one of the cues (CS+-), and (2) contexts changed from A to B. While both of these factors may have influenced theta responses, future research is needed to determine whether theta power during extinction signals safety specifically when both context and contingencies change in the context of a multi-cue paradigm such as ours, or whether it would also be observed in extinction learning paradigms in which contexts do not change (e.g., AAB designs) or in those including only one cue (CS+-).

We note that in our AMY theta power analyses, some time-frequency bins observed early in the trial (~300-500ms) showed numerical increases in theta power for CS+ items at an uncorrected level. This might seem surprising, as it contrasts with the main effect observed later in the trial (1.18-1.75s) which reflects a decrease in theta power for threatening items. However, these numerical increases are within the range of what is expected by chance, as assessed by our cluster-based permutation analysis. While our primary analyses focused on theta frequency oscillations, humans are able to rapidly distinguish and respond to threats and safety cues (Méndez-Bértolo et al., 2016). This raises the possibility of a potential role of high-frequency neural activity in signaling threat and safety. Indeed, theoretical accounts suggest that fast cycle gamma oscillations might promote fast information transfer between brain regions (Fell & Axmacher, 2011), facilitating fear learning by modulating the timing of neuronal firing inducing plasticity (Cattani et al., 2024). In particular, gamma cycles align with several biophysical factors regulating excitatory input integration in basolateral amygdala principal neurons (Bocchio et al., 2017), and experimental studies in rodents have shown that amygdala gamma oscillations contribute to fear memory formation through theta-gamma coupling and spike-field coherence (Popescu et al., 2009; Stujenske et al., 2014). While we did not observe any effect of gamma frequency oscillations in our paradigm, we believe further research will be needed to fully assess their possibly role in signaling threat and safety and modulating fear learning in humans—for example, via phase amplitude coupling (Saint Amour di Chanaz et al., 2023). Interestingly, recent findings suggest that hippocampal representational patterns linked to amygdala gamma bursts reinstate the content of emotional memories (Costa et al., 2024), providing evidence for a potential role in the representation of fear in the interaction with the hippocampus

The high temporal resolution of iEEG recordings allowed us to shed light on the relative time courses of signals across regions of the extinction network. Our results reveal a temporal sequence in which an earlier representation of specific contexts in lPFC is followed by a later correlation of $\text{lPFC}_{\text{CONTEXT}}$ and $\text{AMY}_{\text{CONTEXT}}$ across trials, which, in turn, coincides with the safety-related theta power increase in AMY. This temporal order of representational and neurophysiological signals aligns with previous research showing that distributed areas of the extinction network, including the PFC, first coordinate the processing of contextual information, which subsequently influence the AMY (Holland & Bouton, 1999; Quinn et al., 2008).

Notably, the correlation of lPFC and AMY context-specificity overlapped with the period of lPFC context specificity (from 0.8 to 1.15s after cue onset). However, it also shows a delay between lPFC and AMY, with lPFC context specificity emerging ~1s earlier. This temporal lag seems to decrease during later trial periods. The main effect of lPFC context-specificity occurs immediately prior to the time period when the strongest coordination is observed with the AMY (at approximately 1.2s). Although the AMY did not show a main effect of context specificity—possibly because the similarity of same-context representations was attenuated by repetition suppression—the observed temporal offset suggests that the effect observed in the lPFC drives the AMY representations. Our findings underscore the relevance of dynamic and distributed context representations throughout the extinction network during fear extinction. More specifically, they suggest that AMY safety signals and representations during extinction depend on the preceding representation of specific contexts in lPFC.

The context-dependency of extinction learning shifts the balance in the competition between fear and extinction memory traces towards a renewal of fear

Our findings also shed light on the functional role of lPFC context representations in modulating the reinstatement of fear vs. extinction memories. Reinstatement of memory traces has previously been analyzed via encoding-retrieval similarity (Pacheco Estefan et al., 2019; Staresina et al., 2016; Yaffe et al., 2014). In the context of fear extinction, previous fMRI studies investigated the reinstatement of fear and extinction memory traces using multivoxel pattern similarity analysis (Hennings et al., 2022). Following a similar approach, here we implemented a novel metric of reinstatement that directly compared the reoccurrence of fear vs. extinction

555   memory traces, assuming an inhibition of these two memories during test (Bouton &
556   Swartzentruber, 1991; Lebois et al., 2019; Santini et al., 2008; Szeska et al., 2020).
557   Using this metric, we observed that higher fear vs. extinction memory reinstatement
558   in TMP correlated with context-specific representations in the lPFC – i.e., in
559   participants with elevated levels of context-signaling in the lPFC during extinction,
560   the balance between fear memory reinstatement and extinction memory
561   reinstatement was shifted towards the former (Figure 6). Importantly, this effect was
562   selectively observed for CS+- cues, but not for CS++ or CS-- cues, which are
563   unlikely to compete between acquisition and extinction. Our finding aligns with
564   clinical observations that high levels of context dependency during extinction reduce
565   the persistence of the safety responses learned during extinction, and favor the
566   recurrence of fear (Lebois et al., 2019). One may assume that if extinction contexts
567   are represented more specifically, the newly formed extinction memories are
568   considered as an exception, and the original fear memory is likely to come back. Our
569   results further show that aversive memories are extinguished through an interaction
570   between executive control (lPFC) and sensory regions (TMP) representing contexts
571   and items, respectively. Possibly, more pronounced context representations in lPFC
572   involve effortful inhibition processes similar to those during episodic memory control
573   (Engen & Anderson, 2018), which are difficult to maintain in the longer run.

574

575   *Fear learning, extinction, and episodic memory*

576   Our results highlight the crucial role of both HPC and lPFC in encoding
577   contextual information during fear learning and extinction, consistent with findings in
578   rodents (Bouton, 2004; Gilmartin et al., 2014; Maren et al., 2013). Notably, the lPFC
579   was selectively engaged in the representation of context-specific activity during
580   extinction and not during acquisition, again consistent with its involvement in the
581   suppression of fear memories via context-signaling. While we observed a context-
582   specific representation during acquisition as well (in the HPC), this effect did not
583   coincide with the time period of the cue presentation – suggesting that specific
584   contexts are not neglected during acquisition but exert a less prominent influence on
585   the memory traces of individual cues. While the lPFC effects align with the role of
586   this region in inhibiting fear expression (Gilmartin et al., 2014; Maren et al., 2013),
587   the HPC responses are consistent with the greater sensitivity of this region to views
588   of landscapes and scenes (Maguire & Mullally, 2013; Rolls, 1999) and the well

589 acknowledged role of the HPC in rapidly encoding new memory traces during

590 acquisition (McClelland et al., 1995). We note that several studies have shown that

591 fear acquisition is characterized by an overgeneralized and decontextualized fear

592 responses, while extinction engages PFC-dependent mechanisms more reliant on

593 context representation (for reviews, see Gilmartin et al., 2014; Maren et al., 2013).

594 We previously found that overgeneralization of representations predicted subsequent

595 intrusive memories (Kobelt et al., 2024), which are notorious for their

596 decontextualized nature and the fact that they can be easily and involuntarily

597 triggered by ubiquitous sensory cues.

598   Our findings suggest that extinction memory traces involve neurophysiological

599 patterns and representational characteristics that differ from those formed during

600 acquisition and are more reminiscent of episodic memory traces (see also Hennings

601 et al., 2022). First, CS- cues were associated with increased theta oscillations during

602 extinction. Previous studies on the role of theta oscillations for episodic memory

603 formation suggest that theta power increases particularly for associative – i.e.,

604 context-dependent – memories, while the encoding of item memories is commonly

605 associated with theta power reductions (Herweg et al., 2020). Second, the increased

606 levels of lPFC context-specificity during extinction are reminiscent of the formation of

607 new episodic memories, which are defined, among other factors, by their

608 dependence on context (Tulving, 2002; Yonelinas et al., 2019). These findings

609 support the notion of fear acquisition as a form of associative learning that lacks

610 contextual specificity, while fear extinction, similar to episodic memory, is more

611 strongly influenced by contextual representations. Third, memory traces of CS- items

612 during extinction were less stable and, therefore, more trial-specific in sensory

613 regions of the TMP, reminiscent of the unique and distinct representations during

614 individual trials observed in episodic memory paradigms. Indeed, while previous

615 studies have linked the stability of neural representations to both episodic memory

616 encoding (Lu et al., 2015; Xue et al., 2010) and the formation of fear associations

617 (Visser et al., 2013), others have suggested that item stability is modulated by

618 contextual factors and is disrupted when incongruent contextual information is

619 presented (Wu et al., 2023). To summarize, these findings point towards the flexible

620 and malleable nature of safety memory traces during extinction; and indeed, higher

621 level of context-dependency (and thus putatively higher "episodicity" of extinction

622 memories) predicted their relative weakening as compared to the more robust and
623 generalized initial fear memory traces.

624 While both extinction learning and episodic memory are tightly linked to
625 contextual information, these two forms of associative learning certainly differ in
626 other respects. Episodic memory is specifically defined by its connection to "mental
627 time travel" – the ability to mentally project oneself into the past or future and to
628 reconstruct specific experiences (Tulving, 2002) – a feature that is not evident in fear
629 extinction paradigms. Moreover, fear learning and extinction typically occur across
630 repeated trials, while episodic memory has been characterized as one-shot learning
631 (McClelland et al., 1995). Despite these differences, our data shows that extinction
632 learning shares representational signatures with item-context associations in
633 episodic memory (Eichenbaum, 2017b; Pacheco Estefan et al., 2019).

634

635 *Methodological considerations*

636 Cognitive impairments in implanted epilepsy patients are generally a concern
637 when conducting cognitive experiments using iEEG data. To facilitate the
638 interpretation of our findings, we provide in Supplementary Table 1 all available
639 demographic and clinical information of our patients, including global intelligence
640 quotient (IQ), memory quotient (MQ; assessed via the Wechsler Memory Scale;
641 WMS-III) and seizure onset zone (SOZ). While the IQ scores were within the normal
642 range, the patients in our study showed a lower-than-average MQ. Although our
643 paradigm did not directly probe the specific memory functions assessed by the
644 WMS, this limitation is inherent to intracranial EEG recordings in epilepsy patients
645 and should be acknowledged. While we cannot entirely rule out the influence of
646 memory impairments on our results, the simplicity of our task design—with only three
647 cues and a single instruction for all trials—, and the fact that learning curves align
648 with expected behavioral patterns partially alleviate this concern. Additionally, fear
649 associations can be formed in the absence of conscious awareness of the CS-US
650 contingencies (Knight et al., 2003, 2009), suggesting that explicit memory
651 impairments may not necessarily prevent the formation and extinction of fear
652 associations. Indeed, our trial-by-trial progressive learning analysis confirms that
653 patients successfully understood the task instructions and appropriately learned the
654 CS-US contingencies, adaptatively adjusting their responses to contingency changes
655 in CS+- trials (see Figure 1D).

656   We note that, similar to other fear conditioning paradigms conducted with

657   humans (e.g., Balooch et al., 2012; Bandarian Balooch & Neumann, 2011; Kinner et

658   al., 2016; Lissek et al., 2008; Lissek & Tegenthoff, 2024; Neumann et al., 2007;

659   Neumann & Kitlertsirivatana, 2010; Shiban et al., 2013), we chose to rely on self-

660   reported ratings of cue 'threat' and 'safety' as our behavioral measure of fear rather

661   than employing implicit physiological recordings. While explicit ratings have

662   limitations, we decided to employ them based on both ethical and clinical

663   considerations, particularly given the constraints of conducting research with

664   implanted epilepsy patients. Specifically, the use of a mild aversive stimulus (a

665   scream) rather than electric shocks in our study precluded the use of some

666   traditional implicit measures, such as skin conductance or heart rate responses,

667   which are more suited for capturing stronger autonomic reactions. Importantly,

668   research indicates that explicit ratings are more stable than physiological responses

669   across repeated measurements (Corneille & Gawronski, 2024), and several reviews

670   support their validity in fear conditioning research (Boddez et al., 2013; Constantinou

671   et al., 2021). Explicit self-reports have been shown to correspond with both fear

672   startle (Balooch et al., 2012; Lissek et al., 2008) and skin conductance responses

673   (Constantinou et al., 2021; Shiban et al., 2013; Taschereau-Dumouchel et al., 2020).

674   Moreover, self-reports might better capture deliberative processes than implicit

675   measures (Lonsdorf et al., 2017), which aligns with the more "cognitive" nature of our

676   paradigm (see also Kinner et al., 2016; Lissek & Tegenthoff, 2024). Although explicit

677   ratings are susceptible to demand effects or participant interpretation, such effects

678   are unlikely to have played a major role in our results due to the simplicity of the task

679   and the absence of explicit experimenter expectations beyond learning the correct

680   CS-US associations. Crucially, participants were instructed to rate the 'safety' of the

681   cue itself rather than their internal emotional state, minimizing subjective bias.

682   Furthermore, apart from analyzing the correlation of the ratings with reinstatement in

683   the TMP (Figure 7), we did not rely on the ratings to categorize stimuli or correlate

684   them with the neural data in other analyses, which minimizes their influence in the

685   overall interpretation of our findings. Finally, we highlight that the trial-by-trial learning

686   effects we observed based on the ratings align closely with what would be expected

687   for skin conductance responses (see for example, Ma et al., 2025).

688   Our study was not preregistered, and we acknowledge this as a limitation, as

689   preregistration enhances the transparency and reproducibility of research findings.

690 However, the exploratory nature of our study, which is the first to investigate context-
691 dependent extinction learning in humans via iEEG, made preregistration challenging.
692 Our research aimed to provide a first characterization of the representational
693 dynamics underlying the context-dependent nature of extinction learning—which
694 required the integration of methodologies from diverse domains and the
695 development of novel metrics. For example, we adapted RSA, an established
696 method in episodic memory research, to examine context specificity and item
697 stability within a classical ABC context-dependent extinction paradigm (Figure 2 and
698 3). We also developed a new measure to quantify fear reinstatement in the context
699 of classical conditioning (see Figure 6). Despite the pioneering and exploratory
700 nature of our study, we note that we formulated specific predictions in the
701 introduction and results sections.

702 Notably, some of our results were not in line with our predictions. Specifically, we
703 anticipated to observe increased theta power for CS+ as compared to CS- items
704 during acquisition in AMY, which we did not find. Instead, we observed a relative
705 decrease of power for these trials during extinction in this region (see section above:
706 *A safety signal in the AMY driven by context representations in the extinction*
707 *network*). We also expected context-specific signals both in lPFC and HPC during
708 extinction based on previous literature (Maren et al. 2013). However, context
709 specificity was only observed in the lPFC during extinction, while the involvement of
710 the HPC was only observed during acquisition (see section *Fear learning, extinction,*
711 *and episodic memory* above). Finally, we expected lPFC context specificity to drive
712 context representations in AMY and HPC, but while lPFC and AMY context
713 representations were indeed correlated across trials, we did not find a main effect of
714 context specificity during extinction in either AMY or HPC (see section *A safety*
715 *signal during extinction in the AMY correlating with representational changes of*
716 *contexts and items above*). Despite its exploratory nature, our work represents an
717 important first step in exploring the representations underlying context-dependent
718 extinction learning in humans using iEEG, providing a foundation for more targeted,
719 hypothesis-driven research in the future. In summary, our findings shed light on the
720 key neurophysiological mechanisms, representational characteristics, and functional
721 relevance underlying context-dependence fear extinction in the human brain. They
722 bridge the gap between animal and human studies, contribute first evidence to a
723 novel conceptual framework that integrates fear conditioning and episodic memory

724 research, and provide a mechanistic basis for understanding the renewal of fear in

725 clinical contexts.

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745 **Methods**

746

747 *Participants*

748 Forty-nine patients (22 females, 28.2 ± 8.18 years) with medically intractable

749 epilepsy participated in the study. Data were collected at the Pitié Salpêtrière

750 hospital, Paris, and the South China Normal University Hospital, Guangzhou, China.

751 The study was conducted according to the latest version of the Declaration of

752 Helsinki and approved by the institutional review board at the Pitié Salpêtrière

753 hospital (C11-16, C19-55 National Institute of Health and Medical Research

754 sponsor). All patients provided written informed consent.

755

756 *Experimental design*

757 Participants saw three different images of electrical devices (a hair dryer, a
758 ventilator, and a toaster) during acquisition, extinction, and test. During acquisition,
759 two out of these three images were associated with an aversive unconditioned
760 stimulus (US), and labeled as "CS+", while the remaining image was not associated
761 with the US and labeled "CS-" (Figure 1A, top). The association of the images with
762 the US was counterbalanced across subjects. Each image was presented 24 times
763 during acquisition and extinction in a pseudo-randomized fashion, in order to avoid
764 patients anticipating the images and the US (total number of trials in a block: 72).
765 Specifically, a maximum of 3 consecutive presentations of the same item was
766 allowed, irrespective of their particular contingencies. Every cue was presented in
767 one of many different contexts (i.e., videos), which were not related to the type of
768 trial and the occurrence of the US. Four unique context videos were presented
769 during acquisition and another four were presented during extinction. Critically,
770 during extinction, the CS type of one of the items was modified, while the other two
771 remained the same. This led to the distinction between three different types of items:
772 CS++, CS+-, and CS--. After the extinction period, participants performed a test in
773 which they again saw the three items during 16 trials each (total number of trials
774 during the test = 48), which were associated with eight new experimental contexts
775 (the number of contexts was doubled to match those presented during acquisition
776 and extinction). In the test phase, no US was presented for any of the items. The
777 association of the contexts and the cues was completely randomized in every
778 subject, so that contexts were not predictive of the CS. Please note that this
779 approach differs from conventional fear and extinction learning paradigms. Our
780 design was developed to leverage the power of RSA and iEEG to dynamically track
781 the representation of contexts across key hubs of the extinction network. On the
782 other hand, our paradigm aligns with conventional ABC paradigms as it contains
783 three different context categories across experimental phases (See Supplementary
784 Note 4). Similar to the cue-CS associations, the context videos were
785 counterbalanced across subjects and experimental phases. Following previous
786 literature, the reinforcement rate of the CS+ items was set to 50% both during
787 acquisition and extinction, in order to avoid fast associative learning and extinction of
788 the cue CS associations (Dunsmoor et al., 2014).
789 At every trial, in each experimental phase, participants first saw a fixation
790 cross for 0.5s, and then a context video which was presented for 2s. After the video

791   presentation, a cue was overlayed on top of the video for 1.75s, and then the US

792   was presented for 1s. Finally, a question was presented how much participants had

793   expected a US – i.e., how threatening of safe they had perceived the CS – and

794   participants provided their responses on a scale from 1 (threatening) to 4 (safe).

795   The US consisted of an unpleasant scream, which was delivered via the

796   computer speakers. The sound levels of the scream were at ~85 decibels, resulting

797   in a highly aversive but not painful amplitude. The duration of the scream was 1s.

798   Given the 50% reinforcement rate of the US, the scream was presented a total of 36

799   times in the experiment (CS++: 24 times across acquisition and extinction; CS+- 12

800   times only during acquisition).

801   The experiment was programmed in Presentation (Neurobehavioral systems,

802   California, USA), and was deployed on a laptop computer running Microsoft

803   Windows. Patients performed the experiment while sitting in their hospital beds and

804   responded to the memory test using the keyboard of the laptop computer.

805   Please note that our experimental design is grounded in theoretical

806   considerations from the human fear learning literature and has been adapted to meet

807   the specific requirements of conducting research in implanted epilepsy patients.

808   While its specific combination of features is novel and has not been fully validated,

809   we provide a detailed rationale for each of our design choices, highlighting their

810   similarities and differences to previous research in the field, in Supplementary Note

811   4.

812

813   *Behavioral analysis*

814   We conducted two main behavioral analyses. In the first analysis, we examined

815   average ratings across trials within each experimental phase in order to identify

816   differences in the ratings between phases. A two-way repeated measures ANOVA

817   was used with trial type and experimental phase as factors. Post-hoc tests with

818   Bonferroni correction were applied across conditions after confirming that both a

819   main effect and an interaction were observed (Figure 1). Due to a technical issue

820   during the test phase, one participant was excluded from this analysis, resulting in a

821   sample size of 48 subjects.

822   In the second analysis, we sorted trials according to their type (CS++, CS+-,

823   and CS--) and compared the ratings for each trial type at every position. Since the

824   comparisons were performed on single-trial data, group level distributions were not

825   normal. Thus, we employed the non-parametric Wilcoxon-signed rank test to assess
826   whether conditions differed at every trial. This was done in order to evaluate the
827   progressive learning of the trial contingencies during acquisition and extinction. In
828   this analysis, we sorted trials by type (CS++, CS+-, and CS--) and performed a
829   Wilcoxon signed-rank test across subjects at each trial to compare performance
830   across conditions. While sorting trials in this way disrupted the absolute trial
831   positions in the experiment, the sequential order of trials within each condition was
832   maintained. For a specific trial (e.g., trial 9), the CS type could be different across
833   subjects. However, after sorting by condition, the first CS++ in all subjects is always
834   compared with the first CS+- and the first CS-- trial, the second with the second, and
835   so on. Thus, the paired Wilcoxon tests were performed across subjects in individual
836   trials corresponding to the same ordinal position in each condition. We corrected for
837   multiple comparisons using cluster-based permutation statistics, shuffling the
838   condition labels at every trial.

839

840   *Intracranial EEG recordings*
841   In Paris, recordings were conducted using the ATLAS amplifier (Neuralynx,
842   Bozeman, MO) with a total of 160 channels and a sampling rate of 4096 Hz
843   (Lehongre et al., 2022). In Guangzhou, data was collected using a 256-channel
844   Nihon Kohden Neurofax 1200A Digital System, sampled at 2000 Hz. The number of
845   electrodes and their implantation sites were determined exclusively by clinical
846   requirements. For all analyses, data was downsampled to 1000 Hz. Bipolar
847   referencing was applied, where the activity at each contact point was subtracted
848   from that of the nearest contact on the same electrode, resulting in a total of N-1
849   virtual channels for an electrode with N channels after re-referencing.

850

851   *Channel Localization*
852   Stereotactically implanted electrodes varied in type, number of contacts, and inter-
853   contact distances. In Paris, two types of electrodes were used: standard macro
854   electrodes and macro-micro electrodes. The standard macro electrodes were depth
855   electrodes (Ad-Tech Medical Instrument Corporation, Wisconsin, USA), consisting of
856   4 to 12 platinum contact points, each with a diameter of 1.12 mm and a length of
857   2.41 mm, connected via nickel-chromium wiring. The center-to-center distance
858   between adjacent contacts was 5 mm. The macro-micro electrodes (also from

859 AdTech, Wisconsin, USA) were of the Behnke-Fried type, consisting of 8 platinum

860 contact points, each 1.28 mm in diameter and 1.57 mm in length, mounted on the

861 surface of a polyurethane tube with a hollow lumen. In addition, eight 40 µm

862 platinum-iridium microwires were embedded within the macroelectrode, extending 3-

863 6 mm beyond the tip into the surrounding cerebral tissue (data from these microwires

864 were not analyzed in this study). The inter-contact distance between contacts 1 and

865 2 was 3 mm, while the distance between subsequent contacts (e.g., 2 and 3, 3 and

866 4, etc.) was 7 mm.

867 In Guangzhou, depth electrodes were obtained from Sinovation Medical

868 Technology (Beijing), and consisted of platinum depth electrodes with 7 to 19

869 contacts each (0.8 mm diameter, 2 mm length, and 1.5 mm spacing) .

870 Electrode localization was conducted separately by our clinical teams in Paris

871 and Guangzhou using very similar methodological approaches. At both sites, post-

872 implantation computed tomography (CT) images were coregistered onto

873 preimplantation magnetic resonance imaging (MRI) data. Both sites used the same

874 software including Freesurfer (http://surfer.nmr.mgh.harvard.edu/), 3DSlicer

875 (https://www.slicer.org), and Statistical Parametric Mapping (SPM;

876 https://www.fil.ion.ucl.ac.uk/spm/).

877 In Paris, electrode coordinates were extracted using the EPILOC toolbox

878 developed on the STIM platform (Stereotaxy, Techniques, Images, Models; http://pf-

879 stim.cricm.upmc.fr). EPILOC automates several image processing steps using

880 Freesurfer, 3DSlicer, and SPM, complemented by Brainvisa (Rivière et al., 2011).

881 Anatomical models were generated from preoperative MRI data using Freesurfer.

882 Post-operative CT data within the patient's native space were coregistered onto the

883 preoperative MRI data and then normalized to a Montreal Neurological Institute

884 (MNI) template using SPM. Depth sEEG electrodes were automatically localized in

885 postoperative CT images by segmenting electrode artifacts and classifying them

886 based on their proximity to trajectories planned by stereotactic guidance devices.

887 In Guangzhou, channel locations were again identified by coregistering the

888 post-implantation CT images to pre-implantation MRIs acquired for each patient.

889 Patient specific anatomical surfaces were generated with Freesurfer and normalized

890 to MNI space using SPM. Electrode locations were determined using 3DSlicer.

891 A table containing the electrode coordinates in MNI space for all contacts

892 included in the study is presented in Supplementary Table 2. We note that since the

893 implantation scheme is solely determined by clinical criteria, density of sampling

894 within and across brain regions is inherently variable in iEEG research. Several

895 factors might affect the volume of neural tissue sampled, e.g., different electrode

896 configurations, individual implantation schemes, and the selection of ROIs. We

897 performed several control analyses to corroborate that the volume of neural tissue

898 sampled did not differ between our two patient cohorts (see Supplementary Note 5

899 and Supplementary Figure 2).

900

901 *ROI selection*

902 We recorded activity from several brain regions pertaining to the fear learning and

903 extinction network, including the amygdala, hippocampus, the lateral prefrontal

904 cortex (lPFC), the orbitofrontal cortex (OFC) as well as in various areas relevant for

905 processing of sensory stimulus features in the temporal cortex (TMP; Figure 1B).

906 Electrodes from both left and right hemispheres were included in each of the ROIs.

907 We employed the Desikan Killian atlas (Desikan et al., 2006) for label extraction.

908   Electrodes located at the following locations were labeled as TMP electrodes:

909 'inferiortemporal', 'middletemporal', 'superiortemporal', 'transversetemporal',

910 'fusiform', 'temporalpole', 'bankssts', 'parahippocampal', 'entorhinal'. This resulted in

911 a total number of 325 electrodes (46 subjects) in TMP.

912   Electrodes located at the following freesurfer locations were labeled as lPFC

913 electrodes: 'caudalmiddlefrontal', 'parsopercularis', 'parsorbitalis', 'superiorfrontal',

914 'parstriangularis', 'rostralmiddlefrontal', 'frontalpole'. This resulted in a total number of

915 74 electrodes (23 subjects) in lPFC.

916   Since the ROIs lPFC and TMP cover relatively large brain areas, we provide

917 information about electrode localizations in their specific subregions in

918 Supplementary Figure 3.

919   Electrodes located at the following freesurfer locations were labeled as OFC

920 electrodes: 'lateralorbitofrontal', 'medialorbitofrontal'. This resulted in a total number

921 of 35 electrodes (12 subjects) in OFC.

922   In the AMY, we recorded from 82 electrodes in 32 subjects. In the HPC, we

923 recorded from 115 electrodes in 34 subjects. A video depiction of AMY and HPC

924 electrode locations is provided in Supplementary video 1.

925   In all patients, we removed channels located in white matter, resulting in 631

926 clean channels across all patients (12.62 ± 6.66 channels per patient).

927

Preprocessing

929 Preprocessing was performed on the entire raw data using EEGLAB (Delorme &
930 Makeig, 2004), and included high-pass filtering at a frequency of 0.1Hz and low-pass
931 filtering at a frequency of 200Hz. We also applied a band-stop (notch) filter with
932 frequencies of 49-51Hz, 99-101Hz, and 149-151Hz. Artifact rejection was performed
933 using a previously published method (Staresina et al., 2015). This automatic
934 detection algorithm identifies epileptiform activity based on three key criteria: (1) the
935 amplitude of the time-series, (2) the gradient (i.e., the amplitude difference between
936 adjacent time points), and (3) the amplitude of the data after applying a 250-Hz high-
937 pass filter, which further enhances the detection of epileptogenic spikes. Each time
938 point in the raw iEEG data was converted into a z-score based on the participant-
939 specific mean and standard deviation of these three measures across the whole
940 experiment in each channel independently. A time point was marked as artifactual if
941 it exceeded a z-score of 6 in any measure or met a conjunction threshold of an
942 amplitude z-score of 4 combined with either a gradient or high-frequency z-score of
943 3.

944 We found that the combination of raw EEG amplitudes, EEG gradients, and
945 high-pass filtered EEG amplitudes at 250 Hz provided high sensitivity for detecting
946 epileptiform activity and other artifacts. This was further validated through manual
947 identification of artifacts in randomly selected subsets of the data. To ensure that all
948 epileptic activity was completely removed from our data, we not only excluded the
949 detected time points but also removed the 1,000ms preceding and following each
950 detected artifact sample.

951 After identifying and marking the segments of the data contaminated with
952 artifacts, we segmented the data into 7-second epochs (from -3 to 4 s) around the
953 presentation of each CS cue. Any epoch containing artifact segments detected in the
954 continuous (non-epoched) data was entirely removed from further analysis.
955 Additionally, we visually examined raw time-series plots and spectrograms to assess
956 the presence of artifacts in the frequency domain. The number of epochs removed
957 due to artifacts varied depending on the signal quality of each subject and was at
958 20.6 ± 19.2 epochs of cue presentation across subjects and channels. The number
959 of excluded trials in each condition of the experiment is presented in the
960 Supplementary Table 3.

Time-frequency analysis

Using the FieldTrip toolbox (Oostenveld et al., 2011), we decomposed the signal using complex Morlet wavelets with a variable number of cycles, i.e., linearly increasing in 29 steps between 3 cycles (at 1 Hz) and 6 cycles (at 29 Hz) for the low-frequency range, and in 15 steps from 6 cycles (at 30 Hz) to 12 cycles (at 100 Hz) for the high-frequency range (Pacheco Estefan et al., 2019; Staresina et al., 2016). The resulting time-series of frequency-specific power values were then z-scored by taking as a reference the mean activity across all trials in the experiment (Fellner et al., 2020). This type of normalization was applied to remove any common feature of the signal unrelated to the stability of the cue representations or the encoding of context-specific information.

We contrasted oscillatory power values in the frequency ranges of interest between CS+ and CS- trials during acquisition and extinction (Figure 2A). Note that in the analysis during extinction, the data for the CS- condition during extinction includes both CS+- and CS-- trials (i.e., CS- trials that were either CS+ or CS- during acquisition). This was done to focus on the current valence of the CS during extinction—which we assumed would be the primary factor influencing theta power— and to maximize statistical power by including all available trials in the CS- condition. The resulting time-frequency data (with the same decomposition parameters) was employed in our pattern similarity analyses (see below).

We only included subjects with at least 8 trials in the group comparisons, resulting in a total of N = 32 subjects in the analysis performed in the amygdala. The number of CS+ trials was 38.72 ± 8.48 (Mean ± STD) during acquisition and 19.46 ± 4.57 during extinction in this region.

Contrast-based representational similarity analysis

We employed temporally resolved Representational Similarity Analysis (RSA) to evaluate the dynamics of context and cue representations following previous work (Pacheco Estefan et al., 2019, 2021). We assessed the stability of neural representations by computing the within-item similarity of each cue (i.e., the average similarity for all repeated presentations of the same item, Figure 3A). We compared this metric between CS+ and CS- trials, i.e., tested whether the repeated visual presentation of a cue associated with a US (CS+ items) elicited more stable

995 response patterns as compared to the repeated presentation of non-predictive cues
996 (CS- items; Figure 3A). This was done separately during acquisition and extinction.
997 The CS+ vs CS- comparisons were performed using paired t-tests after Fisher Z-
998 transforming the rho values obtained in each trial and condition, in each subject
999 independently. Note that the computation of item-stability implies that
1000 representational patterns across repeated presentations of the same items cannot
1001 be averaged as is often done in RSA.

1002 In addition, we investigated the presence of context-specific representations
1003 by comparing the similarity of same-contexts versus different-contexts in each of our
1004 ROIs. We assessed similarities for same-context and different-context item pairs and
1005 averaged across all combinations of items in the same condition in each subject
1006 independently. The same-context and different-context correlations were then
1007 statistically compared at the group level using t-tests (rho values were Fisher z-
1008 transformed before contrasting them). This metric was also computed during
1009 acquisition and extinction. Similar to the item stability metric previously described, we
1010 did not average the representational patterns corresponding to the repeated
1011 presentation of the same context video but performed the similarity comparisons
1012 across all instances in which a particular video was presented. In some of our
1013 context-specific analysis (presented in Figure 3), we locked the data to the
1014 presentation of the cue to facilitate the comparison with item stability results. Note
1015 that both cues and videos are presented simultaneously during the cue presentation
1016 (the cue is overlayed on top of the video), but the videos are additionally presented
1017 separately for two seconds before the cues.

1018 Representational patterns were defined by specifying a 500ms time window,
1019 in which averaged time courses of frequency-specific power values (all frequencies
1020 in the 1-100Hz range) were included across all contacts in the respective ROI. A
1021 representational pattern was thus composed of activity of N electrodes x 44
1022 frequencies in each 500ms window. Note that the number of channels included
1023 varied depending on the number of electrodes available for a particular subject/ROI.
1024 These two-dimensional arrays were concatenated into 1D vectors for similarity
1025 comparisons. In total, the number of electrodes included in each of our ROIs was the
1026 following (Mean ± STD across subjects): TMP: 7.06 ± 4.23; Amygdala: 2.56 ± 1.39;
1027 Hippocampus: 3.38 ± 1.82; OFC: 2.91 ± 0.9; IPFC: 3.21 ± 1.83.

1028   Pairwise (Spearman) correlations among representational patterns were
1029 computed in sliding time windows (windows of 500ms computed at every 50ms, i.e.,
1030 with a 90% overlap), resulting in a time series of neural representational similarity
1031 matrices (RSMs) for each of our ROIs. In the main analyses (Figure 3 and 4), we
1032 assessed the correlation of these representational patterns across all matching time-
1033 points, resulting in a time-series of item stability or context specificity values for each
1034 condition. The time-series were averaged in each condition for each subject
1035 independently (after Fisher z-transforming them), and the resulting average time-
1036 series were contrasted via paired *t*-tests across conditions at the group level.
1037 Multiple comparisons correction was performed using cluster-based permutation
1038 statistics (see below)

1039   The number of trials for CS+ and CS- items was unbalanced during
1040 acquisition and extinction (two out of three items were labeled as CS+ during
1041 acquisition and only one as CS-, while the reverse was true during extinction). We
1042 thus computed within-item similarity separately for each of the two items with the
1043 same contingency in a given phase, and then averaged the results before
1044 contrasting the time series with the time-series of the item corresponding to the
1045 opposite CS.

1046   We only included subjects with at least 8 trials in each condition, leading to
1047 the following number of participants for the item stability analysis: AMY: 32; HPC: 34;
1048 OFC: 12; IPFC: 22 for both acquisition and extinction. In the TMP, the number of
1049 subjects for the item stability analysis was 43 during acquisition and 44 during
1050 extinction. The number of participants included in the context specificity analysis was
1051 the same in both experimental phases in all regions: IPFC: 22; AMY: 32; HPC: 34;
1052 OFC: 12; TMP: 46.

1053   In all pattern similarity plots, correlations corresponding to each 500ms
1054 window were assigned to the time point at the center of the respective window (e.g.,
1055 a time bin corresponding to activity from 0 to 500ms was assigned to 250ms).

1056

1057 *Coordination analyses*

1058   We computed a single-trial metric of item-stability by assessing the similarity
1059 of the neural representation of an item presented in a trial with the representation of
1060 the same item in all other trials during a particular experimental phase. In addition,
1061 we computed a single-trial metric of context specificity by subtracting the similarity of

1062 each context video with all repeated presentations of the same video and the
1063 similarity with all different videos in a particular experimental phase. These metrics
1064 were computed in each of our ROIs and then correlated between different ROIs
1065 across trials for each participant. Correlations were calculated in sliding time
1066 windows (500ms windows, overlapping by 95%) across all possible combinations of
1067 time points. This resulted in one correlation value for each subject in each time-bin x
1068 time-bin pair. We assessed potential temporal offsets because the latency and
1069 duration of representations may differ across brain regions, making it more likely that
1070 coordination is temporally shifted rather than simultaneous (see King & Dehaene,
1071 2014; Pacheco Estefan et al., 2019). Notably, the relative timing of coordination
1072 offers insights into the temporal order of representations across brain regions during
1073 fear extinction. Statistical significance was assessed by contrasting these rho values
1074 against zero at the group level, and correction for multiple comparisons was
1075 performed using cluster-based permutation statistics (see below).

1076 To confirm that the trial-level results were not driven by the main condition
1077 differences in the representational feature vectors representing CS+ and CS- items,
1078 we performed all trial-level correlation analyzes after separately Z-scoring the
1079 reinstatement values in CS+ and CS- trials. We obtained equivalent results in all of
1080 reported effects. In additional analyses, we assessed connectivity between ROIs
1081 where we observed significant coordination effects using the weighted phase lag
1082 index (wPLI). Results are presented in Supplementary Note 6.

1083 The number of subjects included in the trial-based RSA analyses depended
1084 on two criteria: 1) the number of trials in each condition, which was set to a minimum
1085 of 8 trials, consistent with the contrast-based RSA analyses (see above). A subject
1086 was excluded from this analysis if the number of trials was below 8 in any of the
1087 conditions tested, and 2) the number of subjects with electrodes in both ROIs, which
1088 was based solely on clinical needs. The number of participants included in the trial-
1089 based analyses comparing item stability between regions was the following: TMP-
1090 AMY: 28; TMP-HPC: 31; TMP-lPFC: 18; TMP-OFC; 11; AMY-HPC: 24; AMY-lPFC:
1091 14; AMY-OFC: 8. The number of participants included in the trial-based comparisons
1092 of lPFC context specificity with context specificity in the other areas was the
1093 following: lPFC-AMY: 14; lPFC-TMP: 18; lPFC-HPC: 15; lPFC-OFC: 11.

1094

1095 *Reinstatement analysis*

1096 To investigate whether lPFC context specificity during extinction predicted
1097 reinstatement in the TMP during the test phase, we conducted a subject-level
1098 correlation analysis. First, we calculated a metric of context specificity in the lPFC by
1099 averaging the difference between same-context and different-context correlations
1100 within the time period where significant context effects were observed in this region
1101 during extinction (from 0.8s-1.15s; Figure 4C, middle). These correlations were
1102 averaged both across trials and over time within this time window, resulting in a
1103 single value of lPFC context specificity for each subject. Second, we calculated
1104 reinstatement across experimental phases, comparing activity patterns during the
1105 test with those observed either during acquisition or during extinction (i.e.,
1106 acquisition-test reinstatement and extinction-test reinstatement). We specifically
1107 focused on the TMP, where the main condition differences were observed (i.e.,
1108 higher item stability for CS+ as compared to CS- trials). We correlated each trial
1109 during acquisition and extinction with each trial during test in each subject
1110 independently. We then averaged these correlations across trials and across time
1111 (focusing on the same time window of significant lPFC context specificity) and
1112 subtracted acquisition-test reinstatement and extinction-test reinstatement, resulting
1113 in a single value of TMP reinstatement for each subject. Finally, we correlated
1114 context specificity and reinstatement across subjects using Spearman correlations
1115 (Figure 6B).

1116 To further investigate whether the observed correlation between the lPFC and
1117 TMP was unique to the period of context specificity in the lPFC, we conducted a
1118 temporally resolved analysis. Employing sliding time windows of 500ms with a 95%
1119 overlap, we again correlated same and different items during both acquisition and
1120 extinction and quantified the time course of item-specific reinstatement during the
1121 test phase. In the lPFC, we selected context-specific representations in the same
1122 cluster used in the primary analysis (described above), and correlated these values
1123 across subjects at every time point of the TMP reinstatement time course using
1124 Spearman correlations. This approach yielded a time course of Rho and p-values,
1125 which we subsequently corrected for multiple comparisons using cluster-based
1126 permutation statistics (see below).

1127

1128 *Functional relevance of extinction reinstatement*

1129 To assess the functional significance of reinstatement during the test phase, we
1130 computed a trial-level metric of reinstatement values. Specifically, we measured the
1131 similarity of each trial in the test phase to the same and different items from the
1132 extinction phase and subtracted these two values. This was conducted at matching
1133 time points within each trial, generating a time course of item reinstatement for every
1134 trial. We focused on the TMP, and specifically on the time period when significant
1135 differences in item stability between CS+ and CS- trials were detected. We averaged
1136 reinstatement values in this window, resulting in a trial-level metric of item-specific
1137 reinstatement for each subject. These reinstatement values were then correlated
1138 with the participants' subjective ratings using Spearman correlations (after Fisher Z-
1139 transforming them). The resulting values were tested against zero at the group level
1140 to determine statistical significance (Figure 7).

1141 In our experiment, the average number of trials in which participants failed to
1142 provide a response was considerable (12.65 ± 15.47 per subject), primarily because
1143 responding was not a prerequisite for progressing to the next trial in any
1144 experimental phase. Consequently, we applied a more liberal threshold for the
1145 minimum number of trials included in this analysis, setting it at 4 instead of 8 trials.
1146 We repeated the analysis with the stricter threshold of 8 trials and found qualitatively
1147 similar results (CS++: $t(14)= 2.17$; $p = 0.048$; CS+-: $t(14)= 2.29$; $p = 0.037$; CS--:
1148 $t(13)= -0.67$; $p = 0.51$).

1149 To account for potential group-level biases introduced by outlier effects, we
1150 repeated our analysis after excluding trials with TMP reinstatement values exceeding
1151 2 standard deviations from the mean for each subject. On average, 7.2 ± 4.9 trials
1152 were excluded per subject, resulting in a final sample of 17 subjects. The results
1153 remained largely consistent (CS++: $t(17) = 0.849$, $p = 0.408$; CS+-: $t(16) = 2.51$, $p =$
1154 0.0231; CS--: $t(18) = -0.0126$, $p = 0.99$).

1155

1156 *Multiple comparisons corrections*
1157 We performed cluster-based permutation statistics to correct for multiple
1158 comparisons in the oscillatory power analyses (Figure 2), the contrast-based pattern
1159 similarity analyses (Figures 3B, 3D, 4B and 4C), the trial-based pattern similarity
1160 analysis (Figures 3F and 4D), and the subject-level correlation analysis between
1161 lPFC and TMP (Figure 6).

1162   In the oscillatory power analyses, we contrasted CS+ and CS- time-frequency

1163   maps using paired t-tests, after shuffling the trial labels 1,000 times in each subject

1164   independently. We considered significant a time-point if the difference between these

1165   surrogate conditions was significant at $p < 0.05$ (two-tailed tests were employed). At

1166   every permutation, we computed clusters of significant values defined as contiguous

1167   regions in time-frequency space where significant correlations were observed. We

1168   took the largest cluster at each permutation to build the null distribution. We only

1169   considered significant those contiguous time pairs in the empirical (non-shuffled)

1170   data whose summed t-values exceeded the summed t-value of 95% of the surrogate

1171   clusters (corresponding to a corrected P < 0.05; see Maris & Oostenveld, 2007).

1172   Since the contrast-based pattern similarity analyses were based on matching

1173   time points, we contrasted time-series of different conditions (CS+ vs CS- items or

1174   same vs different contexts) at different time-points using t-tests, after shuffling the

1175   trial labels of each condition 1,000 times in each subject independently. Adjacent

1176   regions of significant differences were formed along one dimension only (time),

1177   resulting in a distribution of surrogate t-values under the assumption of the null

1178   hypothesis. Similar to the procedure employed in the oscillatory power analysis, we

1179   only considered significant those contiguous time pairs in the empirical data whose

1180   summed t-values were above the summed t-value of 95% of the distribution of

1181   surrogate clusters (Maris & Oostenveld, 2007).

1182   In the trial-based pattern similarity analysis, a temporal generalization

1183   approach was employed, and therefore clusters of significant regions were formed in

1184   2 dimensions (time x time). Item stability or context specificity was correlated across

1185   trials between brain regions at the group level using paired t-tests as in the original

1186   data. T-values of adjacent regions where significant differences were observed were

1187   summed. At every permutation, we shuffled the trial labels in one of the regions and

1188   performed the analysis again, storing the cluster with the highest absolute t-value.

1189   This resulted in a distribution of surrogate t-values under the null hypothesis of no

1190   condition differences. Similar to the procedure we employed in the contrast-based

1191   analyses, we then computed p-values by ranking the observed t-values of the

1192   biggest cluster in the empirical data with this surrogate distribution. Time periods with

1193   contiguous significant regions exceeding 95% of surrogate clusters were considered

1194   significant (Maris & Oostenveld, 2007).

1195   In the temporally-resolved subject-level correlation analysis (Figure 6D), we

1196 identified clusters of contiguous subject-level correlations after shuffling the lPFC

1197 and TMP subject IDs 1,000 times. This approach yielded a total of 1,000 time

1198 courses of Rho and p-values. Rho values were summed in the biggest cluster of

1199 contiguous significant correlations at every permutation, resulting in a distribution of

1200 Rho values expected under the null hypothesis. We then ranked the summed rho

1201 values within this cluster in the empirical data with respect to the surrogate

1202 distribution to determine statistical significance.

1203   In addition to cluster-based permutations, we also corrected our results for

1204 multiple comparisons using the Bonferroni method. Given that we included 5 ROIs,

1205 the alpha level for significance was set to $p = 0.05/5 = 0.01$.

1206

Data Availability

1212 Anonymized intracranial EEG data supporting the findings of this study will be made

1213 available upon publication.

Code Availability

1215 Custom-written Matlab code supporting the findings of this study will be made

1216 available upon publication.

Contributions

1218 Conceptualization, M-C.F., D.P.-E. and N.A.; methodology, D.P.-E., A.B., G.J., and

1219 N.A.; data collection, M.-C.F., K.L., V.L., V.F. and J.Y.; data analysis: D.P.-E. and

1220 N.A.; writing—original draft, D.P.-E.; writing—review and editing, D.P.-E. and N.A.;

1221 funding acquisition, N.A., O.G.; resources, V.N., O.G., L.S., B.H., Q.C.; supervision
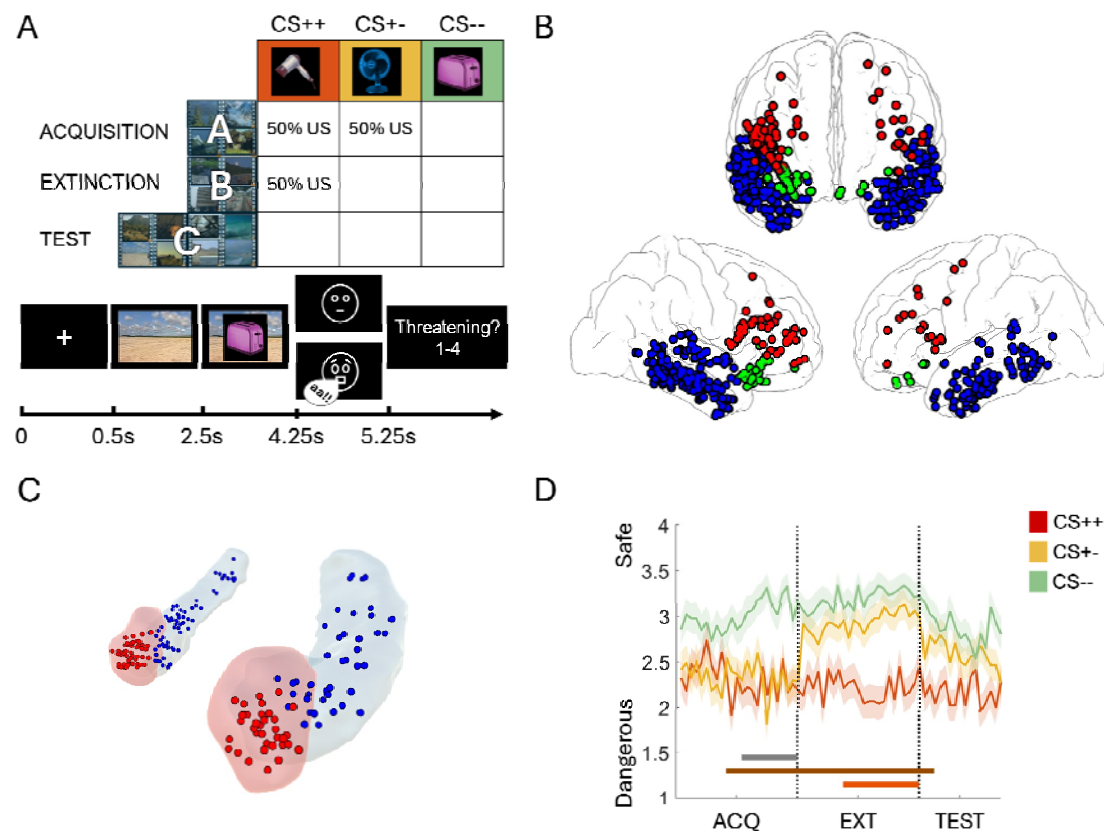
1222 N.A.

Figure 1. Experimental paradigm, electrode implantation and behavioral results

A) Experimental design. Top: Contingencies of consistently threatening CS++ cues (red), consistently safe CS-- cues (green) and cues with changing contingencies (yellow) across the tree experimental phases (acquisition, extinction, and test). Cues were presented together with a set of context videos that were unrelated to the contingency in a given trial and that changed between phases (ABC paradigm). Bottom: Trial structure. Contexts were presented first, but remained on screen during CS presentation. The US consisted of either a neutral face (CS- trials) or a fearful face paired with a loud scream (CS+ trials). The cartoon faces shown are for illustrative purposes only and do not depict the actual stimuli used in the experiment. Participants rated the perceived cue contingencies on each trial. B) Electrode implantation. Selected contacts across our group of participants are overlaid on an average brain surface in Montreal Neurological Institute (MNI) coordinates. Neocortical regions of interest selected for analysis included electrodes implanted bilaterally in temporal (shown in blue), prefrontal (shown in red) and orbitofrontal (shown in green) cortices. C) Electrode locations in the amygdala (shown in red) and the hippocampus (shown in blue). D) Behavioral performance. Wilcoxon signed rank tests were conducted at every trial position to investigate the progressive learning of task contingencies across experimental phases. Regions where significant effects were observed after correction for multiple comparisons are indicated at the bottom of the figure (cluster-based permutation statistics with 1,000 permutations; grey: CS+- vs. CS--; brown: CS++ vs. CS--; orange: CS+- vs. CS++). Shaded lines depict group average ratings at each trial position ± S.E.M.
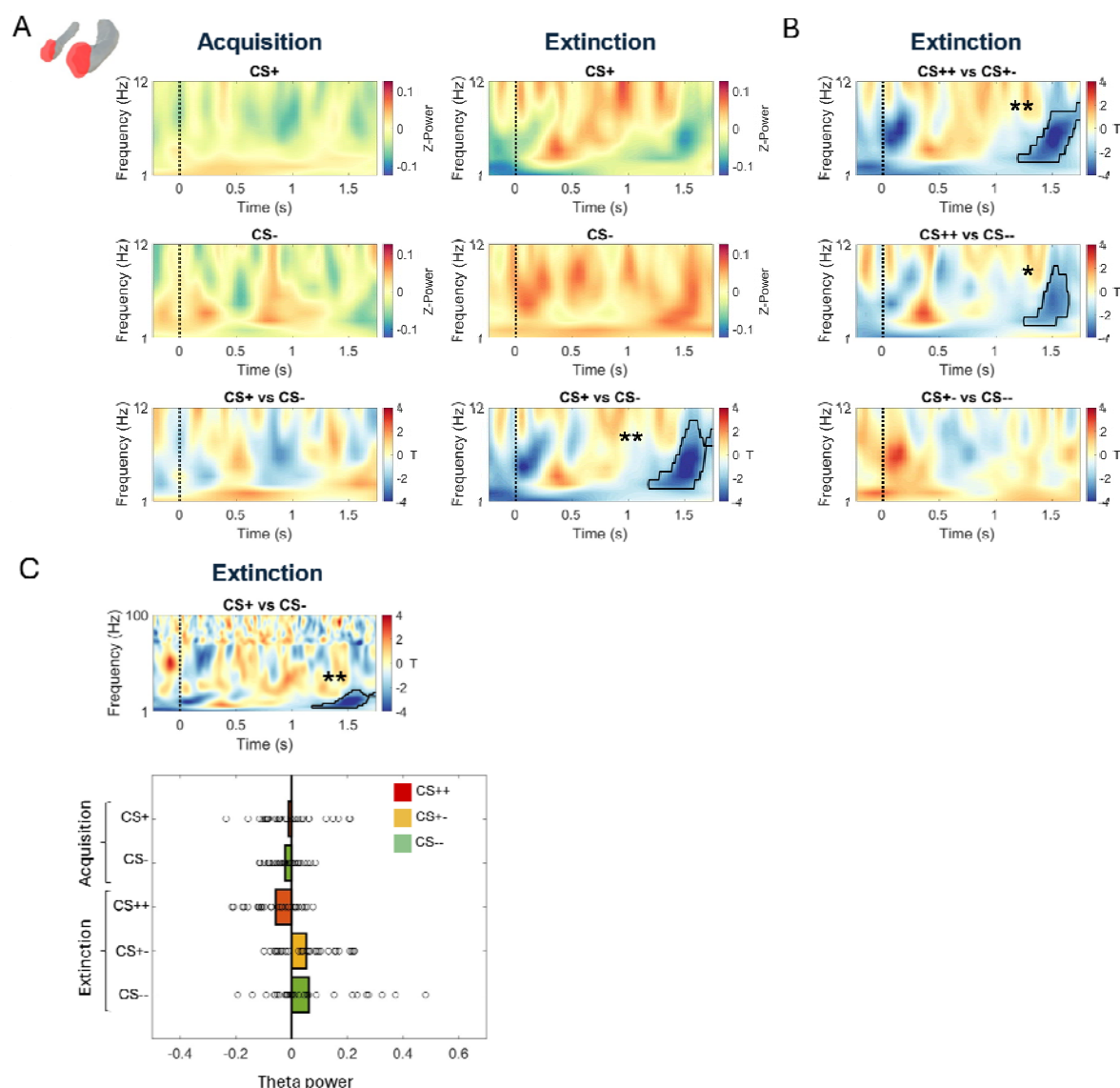
Figure 2. Higher theta power in the amygdala for CS- as compared to CS+ trials during extinction

A) Amygdala 1-12Hz power. During acquisition (left), no significant differences in low frequency power were observed between CS+ and CS- trials. Bottom row shows the result of the contrast between the two conditions, while top and middle rows reflect levels of Z-scored power for each condition. During extinction (right), oscillatory power was significantly higher for CS- trials in a late time period of cue acquisition, i.e., from 1.18 to 1.75s in the 3-10Hz frequency range. Negative T-values depicted in blue reflect higher theta power for CS- items. B) T-maps of the CS++ vs C+- (top), CS++ vs CS-- (middle) and CS+- vs CS-- (bottom) contrasts. C) Top: CS+ vs CS- contrast in the frequency spectrum between 1-100Hz. Bottom: average theta power in the cluster observed in panel A, bottom right, and C, top, for all CS types during the phases of acquisition and extinction. In panels A, B and C, significant regions surviving correction for multiple comparisons using cluster-based permutation statistics are outlined in black in the time-frequency maps. *: p < 0.05; **: p < 0.01.

1257
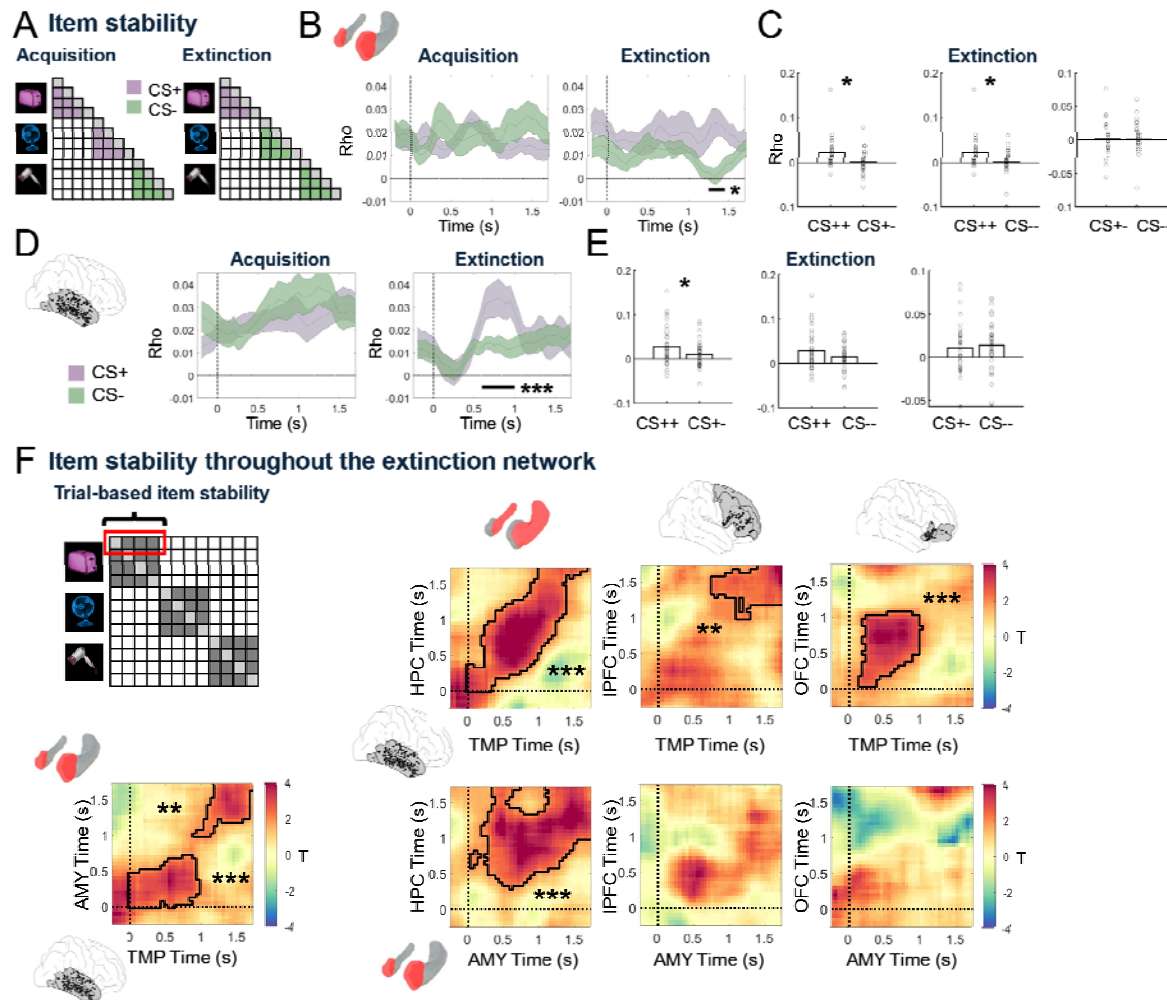
1258
1259

Figure 3. Item stability in lateral temporal cortex, amygdala, and across the extinction network

A) Item stability, defined as the average similarity of neural patterns representing individual items across repeated presentations, was assessed separately for CS+ and CS- trials during acquisition and extinction. B) Analysis of item stability in the AMY revealed no significant differences between CS+ and CS- trials during acquisition (left). During extinction (right), item stability was significantly higher for CS+ as compared to CS- items in a late time period during cue presentation (1.2-1.4s after cue onset). C) Item stability during extinction in the significant time period shown in panel B, right, for CS++ and CS+- (left panel), CS++ and CS-- (middle panel), and CS+- and CS-- trials (right panel). D) Analysis of item stability in the TMP revealed no significant differences between CS+ and CS- trials during acquisition (left). During extinction (right), item stability was significantly higher for CS+ as compared to CS- items in a time period from 0.65s-1s after cue onset. E) Item stability during extinction in the time period shown in panel D, right, for CS++ and CS+- (left), CS++ and CS-- (middle) and CS+- and CS-- (right) trials. F) Top left: A single-trial metric of item stability was computed by assessing the similarity of each trial across repeated presentations (average of the comparisons highlighted with a red rectangle in the RSA matrix). Bottom left: Fluctuations in item

stability were coordinated across trials between the TMP and the AMY during the first second of cue presentation and towards the end of the cue presentation period. Right: Item stability was computed in each ROI and correlated across trials with the values observed in TMP (upper row) and AMY (lower row). In panels B and D, black horizontal lines depict significant time periods after correction for multiple comparisons using cluster-based permutation statistics. Time zero indicates the onset of the CS, and shaded lines depict group average Rho values ± S.E.M. Two-sided paired t-tests were applied at every time point (panels B and D) or every time-by-time point (panel F). In panel F, significant regions surviving multiple comparisons corrections using cluster-based permutation statistics are outlined in black. Time zero indicates the onset of the CS in both time axis. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. AMY: Amygdala; HPC: Hippocampus; lPFC: Lateral Prefrontal Cortex; OFC: Orbitofrontal Cortex.

1315

1316
1317
Figure 4. Context specificity in HPC, lPFC and throughout the extinction network

A) Context specificity analysis. Representational patterns corresponding to the presentation of individual context videos were compared across same contexts (blue) and different contexts (pink) during acquisition and extinction. Periods of context specificity were defined as those where same context correlations were significantly higher than different context correlations. B) Context specificity during video presentation in HPC (top) and lPFC (bottom). During acquisition (left), context-specific representations were observed in the HPC locked to the onset of the video. During extinction (right), a similar effect was observed in the lPFC. C) Context specificity during cue presentation. During acquisition (left), no significant differences were observed between same and different contexts. During extinction (middle), context-specific representations were observed in a time period from 0.8 to 1.15s. Right: Context specificity during extinction (brown) was significantly higher than during acquisition (green). D) Left: Trial-specific values of context specificity, defined as average same minus different context correlations in each individual trial, were computed in each of our ROIs. Right: Focusing on the lPFC, we correlated context specificity values between regions and across trials. Significant regions surviving multiple comparisons corrections using cluster-based permutation

1333   statistics are outlined in black. In panels B and C, black horizontal lines at the bottom of each panel

1334   depict significant time periods after correction for multiple comparisons using cluster-based

1335   permutation statistics. Time zero indicates the onset of the CS, and shaded lines depict group

1336   average Rho values ± S.E.M. Two-sided paired t-tests were applied at every time point (panels B and

1337   C) or every time-by-time point (panel D). AMY: Amygdala; HPC: Hippocampus; lPFC: Lateral

1338   Prefrontal Cortex; OFC: Orbitofrontal Cortex.* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

1339

1340

1341

1342

1343

1344

1345

1346

1347

1348

1349

1350

1351

1352

1353

1354

1355

1356

1357

1358

1359

1360

1361

1362

1363

1364

1365

1366

1367

1368

1369

1370

1371

1372

1373

1374

1375

1376
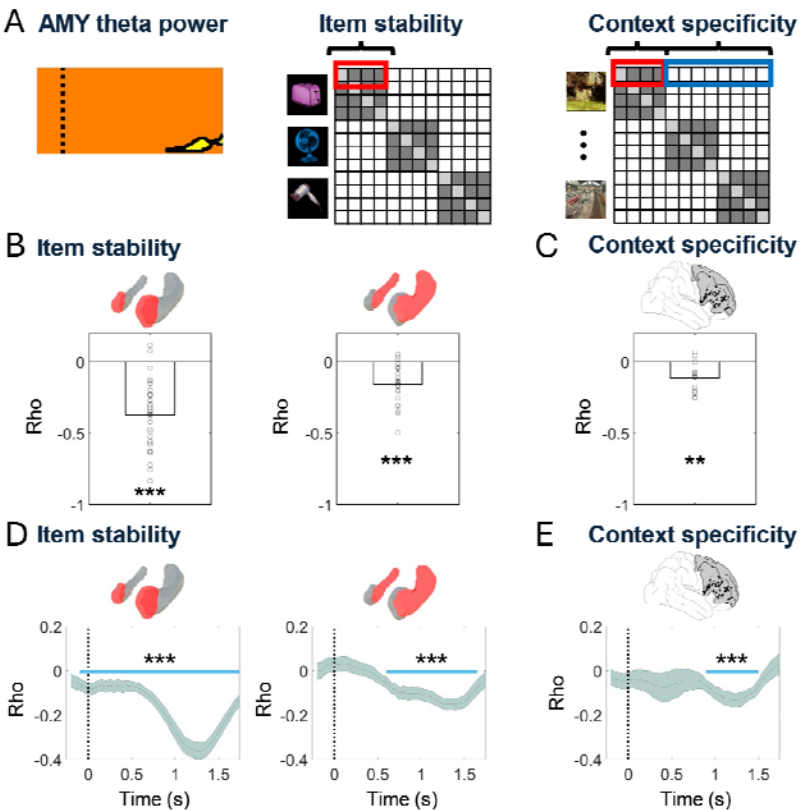
Figure 5. Amygdala theta power correlates with item stability and context specificity across the extinction network

A) Top left: Amygdala theta power was averaged in the time frequency cluster where differences between CS+ and CS- items were observed in the power analysis during extinction. Left: schematic depiction of the selected time frequency period (highlighted in yellow). The average theta power in this cluster was correlated across trials with two metrics: item stability (middle), and context specificity (right). Group level Rho values were contrasted against zero using paired t-tests. B) AMY theta power was correlated across trials with item stability in both AMY (left) and HPC (right). C) AMY theta power also correlated with lPFC context specificity across trials. D) Time-resolved analysis. Left: AMY theta power and item stability correlated throughout the entire cue presentation period. Right: Correlations between AMY theta power and HPC item stability were observed during a later time window (600ms to 1.65s) of cue presentation. E) AMY theta power correlated with lPFC context specificity during a time period that overlapped with the significant condition differences in theta power in AMY (900ms-1.5s). In panels D and E, shaded lines depict group average Rho values ± S.E.M, and time zero marks the onset of cue presentation during the extinction phase. ** p < 0.01; *** p < 0.001
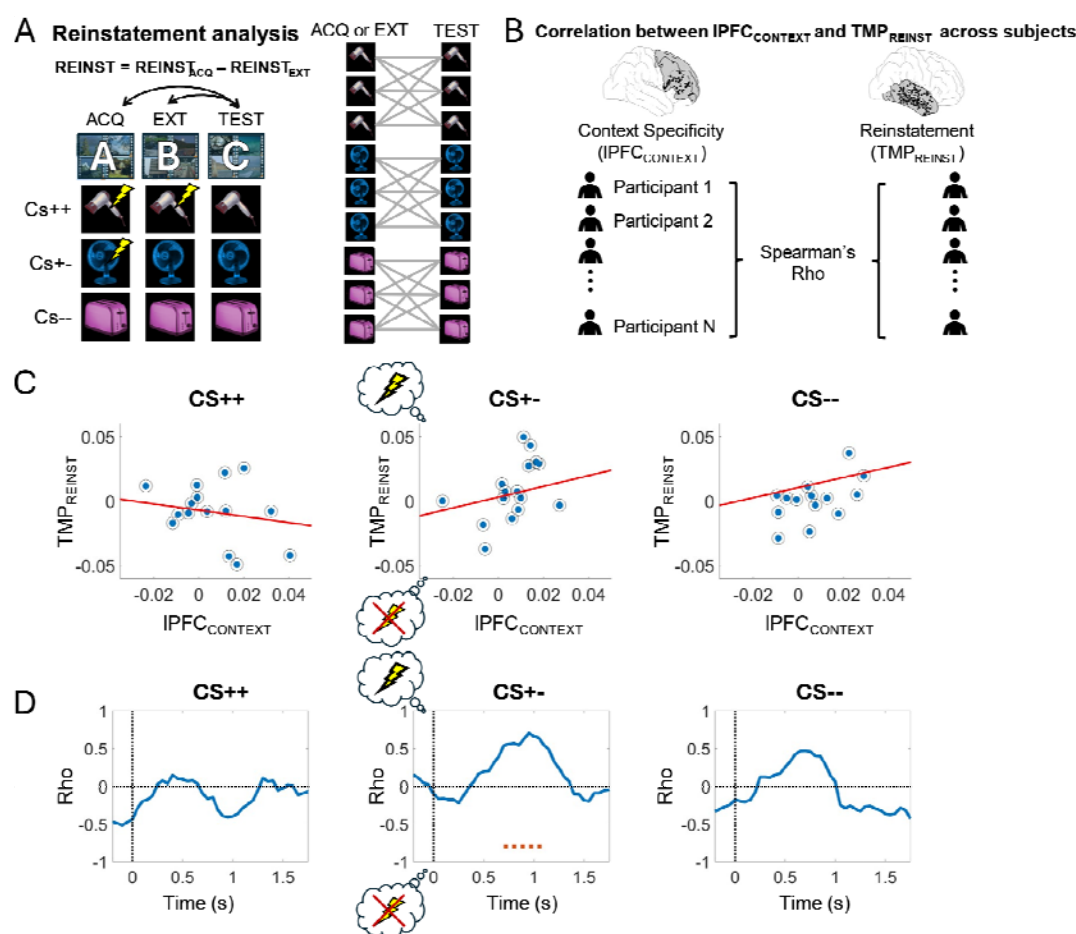
1392

1393

1394

Figure 6. Context specificity shifts the balance in reinstatement of fear vs. extinction memory traces

A) Reinstatement analysis. Reinstatement of acquisition and extinction activity patterns during the test was calculated separately for CS++, CS+- and CS-- items (left), by computing the similarity across repeated exposures to the same item across experimental phases (right). Acquisition and extinction reinstatement values were subtracted and averaged across trials, resulting in a single value of differential reinstatement in each subject and condition (REINST = REINST$_{ACQ}$ − REINST$_{EXT}$). B) Context specificity in lPFC and reinstatement in TMP were correlated across our group of participants using Spearman correlations. C) LPFC context specificity and TMP reinstatement were correlated during the time period of significant context-specific effects in lPFC. Note that for CS+- trials, positive values of TMP reinstatement indicate predominant reinstatement of fear memory traces, while negative values indicate predominant reinstatement of extinction memory traces. D) Context specificity in lPFC was correlated with TMP reinstatement using overlapping 500ms windows with a 95% overlap (i.e., shifted by 50ms). A cluster of significant correlations was identified specifically CS+- trials (middle panel), which coincided with the time period of lPFC context-specific effects. Orange dashed line in panel D, middle, indicates significant regions surviving multiple comparisons correction at $p_{corr} = 0.052$.
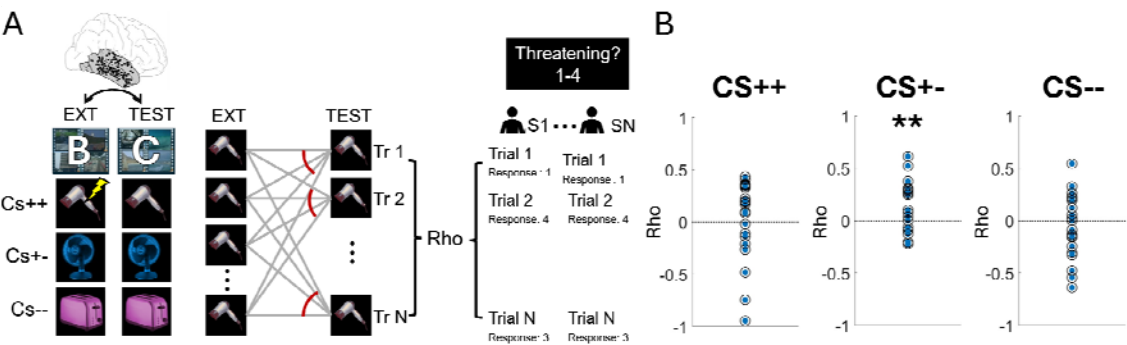
1414

1415

1416

1417 **Figure 7. Reinstatement of extinction memory traces in TMP predicts subjective safety ratings**

1418 A) Extinction-to-test reinstatement analysis: For each trial type (CS++, CS+-, and CS--), the similarity

1419 between extinction and test trials was computed in TMP (left). This was done by averaging the

1420 similarity of one trial during the test across repeated presentations of this same item during the

1421 extinction phase (middle). Spearman correlations were then computed between extinction-to-test

1422 similarity and subjective ratings during the test across trials, in each subject independently (right). B)

1423 Fisher Z-transformed correlations were analyzed at the group level to determine whether they

1424 significantly differed from zero, separately for CS++ (left), CS+- (middle), and CS-- (right) trials. ** p <

1425 0.01.

1426

1427

1428

1429

1430

1431

1432

1433

1434

1435

1436

1437
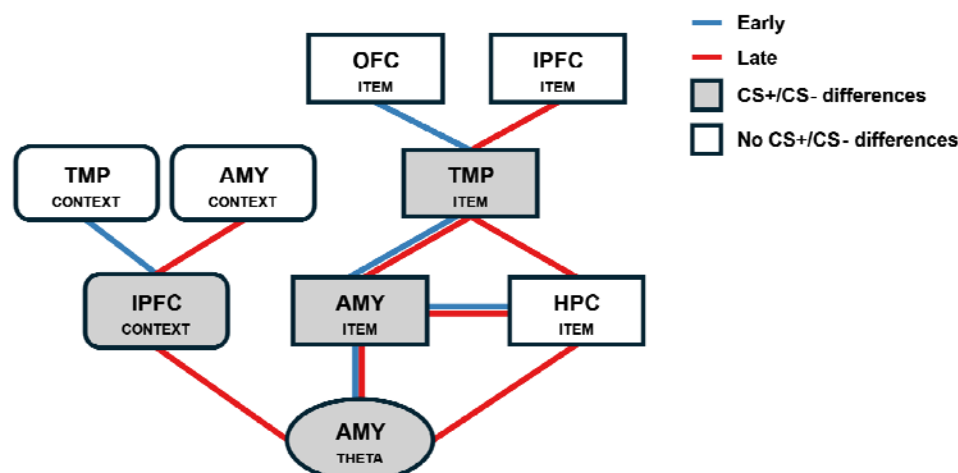
1438

1439

1440

1441

1442

1443

Figure 8: Summary of the main results during extinction learning

The figure summarizes the main results obtained in the power analysis, the item stability analysis and the context specificity analysis during the extinction period across our five ROIs. During the time period of theta power condition differences in the amygdala (CS+ vs CS-; labeled 'AMY$_{THETA}$' in the figure), we also observed lPFC context-specific effects ('lPFC$_{CONTEXT}$') and item stability effects in AMY and lPFC ('AMY$_{ITEM}$'; 'lPFC$_{ITEM}$'). Note that AMY$_{THETA}$ effects occurred during a late time period (indicated with red lines in the figure), but AMY$_{THETA}$ and AMY$_{ITEM}$ also correlated earlier in the trial (indicated with blue lines in the figure). In turn, AMY$_{ITEM}$, TMP$_{ITEM}$ and HPC$_{ITEM}$ effects were observed during early and late time periods. AMY$_{CONTEXT}$ correlated with lPFC$_{CONTEXT}$ during a late time period only, but lPFC$_{CONTEXT}$ correlated with TMP$_{CONTEXT}$ during an earlier time period. Metrics showing main condition differences between CS+ and CS- items during extinction are highlighted in grey. AMY: Amygdala; HPC: Hippocampus; lPFC: Lateral Prefrontal Cortex; OFC: Orbitofrontal Cortex.

References

Axmacher, N. (2023). *Intracranial EEG: A Guide for Cognitive Neuroscientists*. Springer

International Publishing.

Balooch, S. B., Neumann, D. L., & Boschen, M. J. (2012). Extinction treatment in multiple

contexts attenuates ABC renewal in humans. *Behaviour Research and Therapy*,

50(10), 604–609. https://doi.org/10.1016/j.brat.2012.06.003

Bandarian Balooch, S., & Neumann, D. L. (2011). Effects of multiple contexts and context

similarity on the renewal of extinguished conditioned behaviour in an ABA design

with humans. *Learning and Motivation*, *42*(1), 53–63.

https://doi.org/10.1016/j.lmot.2010.08.008

Beckers, T., Hermans, D., Lange, I., Luyten, L., Scheveneels, S., & Vervliet, B. (2023).

Understanding clinical fear and anxiety through the lens of human fear conditioning.

Nature Reviews Psychology, *2*(4), 233–245. https://doi.org/10.1038/s44159-023-

00156-1

Bierwirth, P., Antov, M. I., & Stockhorst, U. (2023). Oscillatory and non-oscillatory brain

activity reflects fear expression in an immediate and delayed fear extinction task.

Psychophysiology, *60*(8), e14283. https://doi.org/10.1111/psyp.14283

Bierwirth, P., Sperl, M. F. J., Antov, M. I., & Stockhorst, U. (2021). Prefrontal Theta

Oscillations Are Modulated by Estradiol Status During Fear Recall and Extinction

Recall. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(11),

1071–1080. https://doi.org/10.1016/j.bpsc.2021.02.011

Bocchio, M., Nabavi, S., & Capogna, M. (2017). Synaptic Plasticity, Engrams, and Network

Oscillations in Amygdala Circuits for Storage and Retrieval of Emotional Memories.

Neuron, *94*(4), 731–743. https://doi.org/10.1016/j.neuron.2017.03.022

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013).

Rating data are underrated: Validity of US expectancy in human fear conditioning.

1486   *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(2), 201–206.

1487   https://doi.org/10.1016/j.jbtep.2012.08.003

1488   Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory*

1489   *(Cold Spring Harbor, N.Y.)*, *11*(5), 485–494. https://doi.org/10.1101/lm.78804

1490   Bouton, M. E., Maren, S., & McNally, G. P. (2021). BEHAVIORAL AND

1491   NEUROBIOLOGICAL MECHANISMS OF PAVLOVIAN AND INSTRUMENTAL

1492   EXTINCTION LEARNING. *Physiological Reviews*, *101*(2), 611–681.

1493   https://doi.org/10.1152/physrev.00016.2020

1494   Bouton, M. E., & Swartzentruber, D. (1991). Sources of relapse after extinction in Pavlovian

1495   and instrumental learning. *Clinical Psychology Review*, *11*(2), 123–140.

1496   Burke, J. F., Long, N. M., Zaghloul, K. A., Sharan, A. D., Sperling, M. R., & Kahana, M. J.

1497   (2014). Human intracranial high-frequency activity maps episodic memory formation

1498   in space and time. *New Horizons for Neural Oscillations*, *85*, 834–843.

1499   https://doi.org/10.1016/j.neuroimage.2013.06.067

1500   Cattani, A., Arnold, D. B., McCarthy, M., & Kopell, N. (2024). Basolateral amygdala

1501   oscillations enable fear learning in a biophysical model. *eLife*, *12*, RP89519.

1502   https://doi.org/10.7554/eLife.89519

1503   Chen, S., Tan, Z., Xia, W., Gomes, C. A., Zhang, X., Zhou, W., Liang, S., Axmacher, N., &

1504   Wang, L. (2021). Theta oscillations synchronize human medial prefrontal cortex and

1505   amygdala during fear learning. *Science Advances*, *7*(34), eabf4198.

1506   https://doi.org/10.1126/sciadv.abf4198

1507   Constantinou, E., Purves, K. L., McGregor, T., Lester, K. J., Barry, T. J., Treanor, M.,

1508   Craske, M. G., & Eley, T. C. (2021). Measuring fear: Association among different

1509   measures of fear learning. *Journal of Behavior Therapy and Experimental Psychiatry*,

1510   *70*, 101618. https://doi.org/10.1016/j.jbtep.2020.101618

1511   Corcoran, K. A., & Maren, S. (2004). Factors regulating the effects of hippocampal

1512   inactivation on renewal of conditional fear after extinction. *Learning & Memory*,

1513   *11*(5), 598–603.

1514   Corneille, O., & Gawronski, B. (2024). Self-reports are better measurement instruments than

1515   implicit measures. *Nature Reviews Psychology*, *3*(12), 835–846.

1516   https://doi.org/10.1038/s44159-024-00376-z

1517   Costa, M., Lozano-Soldevilla, D., Gil-Nagel, A., Toledano, R., Oehrn, C. R., Kunz, L.,

1518   Yebra, M., Mendez-Bertolo, C., Stieglitz, L., Sarnthein, J., Axmacher, N., Moratti, S.,

1519   & Strange, B. A. (2022). Aversive memory formation in humans involves an

1520   amygdala-hippocampus phase code. *Nature Communications*, *13*(1), 6403.

1521   https://doi.org/10.1038/s41467-022-33828-2

1522   Costa, M., Pacheco, D., Gil-Nagel, A., Toledano, R., Imbach, L., Sarnthein, J., & Strange, B.

1523   A. (2024). Retrieval of human aversive memories involves reactivation of gamma

1524   activity patterns in the hippocampus that originate in the amygdala during encoding.

1525   *bioRxiv*, 2024.01.18.576178. https://doi.org/10.1101/2024.01.18.576178

1526   Courtin, J., Chaudun, F., Rozeske, R. R., Karalis, N., Gonzalez-Campo, C., Wurtz, H., Abdi,

1527   A., Baufreton, J., Bienvenu, T. C. M., & Herry, C. (2014). Prefrontal parvalbumin

1528   interneurons shape neuronal activity to drive fear expression. *Nature*, *505*(7481), 92–

1529   96. https://doi.org/10.1038/nature12755

1530   Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing

1531   exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*,

1532   *58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

1533   Darna, M., Stolz, C., Jauch, H.-S., Strumpf, H., Hopf, J.-M., Seidenbecher, C. I., Schott, B.

1534   H., & Richter, A. (2024). Frontal Theta Oscillations and Cognitive Flexibility: Age-

1535   Related Modulations in EEG Activity. *bioRxiv*, 2024.07.05.602082.

1536   https://doi.org/10.1101/2024.07.05.602082

1537   Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-

1538   trial EEG dynamics including independent component analysis. *Journal of*

1539   *Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

1540   Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner,

1541   R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J.

1542   (2006). An automated labeling system for subdividing the human cerebral cortex on

1543   MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980.

1544   https://doi.org/10.1016/j.neuroimage.2006.01.021

1545   Dunsmoor, J. E., Ahs, F., Zielinski, D. J., & LaBar, K. S. (2014). Extinction in multiple

1546   virtual reality contexts diminishes fear reinstatement in humans. *Extinction*, *113*, 157–

1547   164. https://doi.org/10.1016/j.nlm.2014.02.010

1548   Eichenbaum, H. (2017a). Memory: Organization and Control. *Annual Review of Psychology*,

1549   *68*, 19–45. https://doi.org/10.1146/annurev-psych-010416-044131

1550   Eichenbaum, H. (2017b). Prefrontal–hippocampal interactions in episodic memory. *Nature*

1551   *Reviews Neuroscience*, *18*(9), 547–558. https://doi.org/10.1038/nrn.2017.74

1552   Engen, H. G., & Anderson, M. C. (2018). Memory Control: A Fundamental Mechanism of

1553   Emotion Regulation. *Trends in Cognitive Sciences*, *22*(11), 982–995.

1554   https://doi.org/10.1016/j.tics.2018.07.015

1555   Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes.

1556   *Nature Reviews. Neuroscience*, *12*(2), 105–118. https://doi.org/10.1038/nrn2979

1557   Fell, J., Ludowig, E., Staresina, B. P., Wagner, T., Kranz, T., Elger, C. E., & Axmacher, N.

1558   (2011). Medial Temporal Theta/Alpha Power Enhancement Precedes Successful

1559   Memory Encoding: Evidence Based on Intracranial EEG. *The Journal of*

1560   *Neuroscience*, *31*(14), 5392. https://doi.org/10.1523/JNEUROSCI.3668-10.2011

1561 Fellner, M. C., Waldhauser, G. T., & Axmacher, N. (2020). Tracking Selective Rehearsal and

1562   Active Inhibition of Memory Traces in Directed Forgetting. *Current Biology: CB*,

1563   *30*(13), 2638-2644.e4. https://doi.org/10.1016/j.cub.2020.04.091

1564 Fenton, G. E., Halliday, D. M., Mason, R., & Stevenson, C. W. (2014). Medial prefrontal

1565   cortex circuit function during retrieval and extinction of associative learning under

1566   anesthesia. *Neuroscience*, *265*, 204–216.

1567   https://doi.org/10.1016/j.neuroscience.2014.01.028

1568 Foster, B. L., Dastjerdi, M., & Parvizi, J. (2012). Neural populations in human posteromedial

1569   cortex display opposing responses during memory and numerical processing.

1570   *Proceedings of the National Academy of Sciences*, *109*(38), 15514–15519.

1571   https://doi.org/10.1073/pnas.1206580109

1572 Foster, B. L., Rangarajan, V., Shirer, W. R., & Parvizi, J. (2015). Intrinsic and Task-

1573   Dependent Coupling of Neuronal Population Activity in Human Parietal Cortex.

1574   *Neuron*, *86*(2), 578–590. https://doi.org/10.1016/j.neuron.2015.03.018

1575 Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O.,

1576   Radua, J., & Harrison, B. J. (2018). Fear extinction in the human brain: A meta-

1577   analysis of fMRI studies in healthy participants. *Neuroscience & Biobehavioral*

1578   *Reviews*, *88*, 16–25. https://doi.org/10.1016/j.neubiorev.2018.03.002

1579 Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A.,

1580   & Radua, J. (2016). Neural signatures of human fear conditioning: An updated and

1581   extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21*(4), 500–508.

1582   https://doi.org/10.1038/mp.2015.88

1583   Garfinkel, S. N., Abelson, J. L., King, A. P., Sripada, R. K., Wang, X., Gaines, L. M., &

1584   Liberzon, I. (2014). Impaired Contextual Modulation of Memories in PTSD: An fMRI

1585   and Psychophysiological Study of Extinction Retention and Fear Renewal. *The*

1586   *Journal of Neuroscience*, *34*(40), 13435. https://doi.org/10.1523/JNEUROSCI.4287-

1587   13.2014

1588   Gattas, S., Larson, M. S., Mnatsakanyan, L., Sen-Gupta, I., Vadera, S., Swindlehurst, A. L.,

1589   Rapp, P. E., Lin, J. J., & Yassa, M. A. (2023). Theta mediated dynamics of human

1590   hippocampal-neocortical learning systems in memory formation and retrieval. *Nature*

1591   *Communications*, *14*(1), 8505. https://doi.org/10.1038/s41467-023-44011-6

1592   Gilmartin, M. R., Balderston, N. L., & Helmstetter, F. J. (2014). Prefrontal cortical regulation

1593   of fear learning. *Trends in Neurosciences*, *37*(8), 455–464.

1594   https://doi.org/10.1016/j.tins.2014.05.004

1595   Griffiths, B. J., Parish, G., Roux, F., Michelmann, S., van der Plas, M., Kolibius, L. D.,

1596   Chelvarajah, R., Rollings, D. T., Sawlani, V., Hamer, H., Gollwitzer, S.,

1597   Kreiselmeyer, G., Staresina, B., Wimber, M., & Hanslmayr, S. (2019). Directional

1598   coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations

1599   in human episodic memory. *Proceedings of the National Academy of Sciences*,

1600   *116*(43), 21834–21842. https://doi.org/10.1073/pnas.1914180116

1601   Hennings, A. C., McClay, M., Drew, M. R., Lewis-Peacock, J. A., & Dunsmoor, J. E. (2022).

1602   Neural reinstatement reveals divided organization of fear and extinction memories in

1603   the human brain. *Current Biology*, *32*(2), 304-314.e5.

1604   https://doi.org/10.1016/j.cub.2021.11.004

1605   Herweg, N. A., Solomon, E. A., & Kahana, M. J. (2020). Theta Oscillations in Human

1606   Memory. *Trends in Cognitive Sciences*, *24*(3), 208–227.

1607   https://doi.org/10.1016/j.tics.2019.12.006

1608   Holland, P. C., & Bouton, M. E. (1999). Hippocampus and context in classical conditioning.

1609   *Current Opinion in Neurobiology*, *9*(2), 195–202.

1610   Jacobs, J. (2014). Hippocampal theta oscillations are slower in humans than in rodents:

1611   Implications for models of spatial navigation and memory. *Philosophical*

1612   *Transactions of the Royal Society B: Biological Sciences*, *369*(1635), 20130304.

1613   https://doi.org/10.1098/rstb.2013.0304

1614   Karalis, N., Dejean, C., Chaudun, F., Khoder, S., Rozeske, R. R., Wurtz, H., Bagur, S.,

1615   Benchenane, K., Sirota, A., Courtin, J., & Herry, C. (2016). 4-Hz oscillations

1616   synchronize prefrontal–amygdala circuits during fear behavior. *Nature Neuroscience*,

1617   *19*(4), 605–612. https://doi.org/10.1038/nn.4251

1618   King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations:

1619   The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210.

1620   Kinner, V. L., Merz, C. J., Lissek, S., & Wolf, O. T. (2016). Cortisol disrupts the neural

1621   correlates of extinction recall. *NeuroImage*, *133*, 233–243.

1622   https://doi.org/10.1016/j.neuroimage.2016.03.005

1623   Knight, D. C., Nguyen, H. T., & Bandettini, P. A. (2003). Expression of conditional fear with

1624   and without awareness. *Proceedings of the National Academy of Sciences*, *100*(25),

1625   15280–15283. https://doi.org/10.1073/pnas.2535780100

1626   Knight, D. C., Waters, N. S., & Bandettini, P. A. (2009). Neural substrates of explicit and

1627   implicit fear memory. *NeuroImage*, *45*(1), 208–214.

1628   https://doi.org/10.1016/j.neuroimage.2008.11.015

1629   Kobelt, M., Waldhauser, G., Rupietta, A., Heinen, R., Rau, E., Kessler, H., & Axmacher, N.

1630   (2024). The memory trace of an intrusive trauma-analog episode. *Current Biology*,

1631   *34*(8), 1657–1669.

1632   Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the Onion of Brain Representations.

1633   *Annual Review of Neuroscience*, *42*, 407–432. https://doi.org/10.1146/annurev-neuro-

1634   080317-061906

1635   Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-

1636   connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*,

1637   4.

1638   Lebois, L. A., Seligowski, A. V., Wolff, J. D., Hill, S. B., & Ressler, K. J. (2019).

1639   Augmentation of extinction and inhibitory learning in anxiety and trauma-related

1640   disorders. *Annual Review of Clinical Psychology*, *15*(1), 257–284.

1641   Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the

1642   formation of episodic memories. *Hippocampus*, *22*(4), 748–761.

1643   https://doi.org/10.1002/hipo.20937

1644   Lehongre, K., Lambrecq, V., Whitmarsh, S., Frazzini, V., Cousyn, L., Soleil, D., Fernandez-

1645   Vidal, S., Mathon, B., Houot, M., Lemaréchal, J.-D., Clemenceau, S., Hasboun, D.,

1646   Adam, C., & Navarro, V. (2022). Long-term deep intracerebral microelectrode

1647   recordings in patients with drug-resistant epilepsy: Proposed guidelines based on 10-

1648   year experience. *NeuroImage*, *254*, 119116.

1649   https://doi.org/10.1016/j.neuroimage.2022.119116

1650   Lesting, J., Daldrup, T., Narayanan, V., Himpe, C., Seidenbecher, T., & Pape, H.-C. (2013).

1651   Directional Theta Coherence in Prefrontal Cortical to Amygdalo-Hippocampal

1652   Pathways Signals Fear Extinction. *PLOS ONE*, *8*(10), e77707.

1653   https://doi.org/10.1371/journal.pone.0077707

1654   Lesting, J., Narayanan, R. T., Kluge, C., Sangha, S., Seidenbecher, T., & Pape, H.-C. (2011).

1655   Patterns of Coupled Theta Activity in Amygdala-Hippocampal-Prefrontal Cortical

1656   Circuits during Fear Extinction. *PLOS ONE*, *6*(6), e21714.

1657   https://doi.org/10.1371/journal.pone.0021714

1658   Likhtik, E., Stujenske, J. M., A Topiwala, M., Harris, A. Z., & Gordon, J. A. (2014).

1659   Prefrontal entrainment of amygdala activity signals safety in learned fear and innate

1660   anxiety. *Nature Neuroscience*, *17*(1), 106–113. https://doi.org/10.1038/nn.3582

1661   Lissek, S., Biggs, A. L., Rabin, S. J., Cornwell, B. R., Alvarez, R. P., Pine, D. S., & Grillon,

1662   C. (2008). Generalization of conditioned fear-potentiated startle in humans:

1663   Experimental validation and clinical relevance. *Behaviour Research and Therapy*,

1664   *46*(5), 678–687. https://doi.org/10.1016/j.brat.2008.02.005

1665   Lissek, S., & Tegenthoff, M. (2024). Dissimilarities of neural representations of extinction

1666   trials are associated with extinction learning performance and renewal level. *Frontiers*

1667   *in Behavioral Neuroscience*, *18*, 1307825.

1668   Liu, J., Zhang, H., Yu, T., Ren, L., Ni, D., Yang, Q., Lu, B., Zhang, L., Axmacher, N., &

1669   Xue, G. (2021). Transformative neural representations support long-term episodic

1670   memory. *Science Advances*, *7*(41), eabg9715.

1671   Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., &

1672   Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear

1673   memory recall. *Nature*, *484*(7394), 381–385. https://doi.org/10.1038/nature11028

1674   Liu, Y., Ye, S., Li, X.-N., & Li, W.-G. (2024). Memory Trace for Fear Extinction: Fragile yet

1675   Reinforceable. *Neuroscience Bulletin*, *40*(6), 777–794.

1676   https://doi.org/10.1007/s12264-023-01129-3

1677   Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J.,

1678   Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees,

1679   F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., … Merz, C.

1680   J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design

1681   and analysis of studies on human fear acquisition, extinction, and return of fear.

1682   *Neuroscience & Biobehavioral Reviews*, *77*, 247–285.

1683   https://doi.org/10.1016/j.neubiorev.2017.02.026

1684   Lu, Y., Wang, C., Chen, C., & Xue, G. (2015). Spatiotemporal Neural Pattern Similarity

1685   Supports Episodic Memory. *Current Biology*, *25*(6), 780–785.

1686   https://doi.org/10.1016/j.cub.2015.01.055

1687   Ma, Y., Kyuchukova, D., Jiao, F., Batsikadze, G., Nitsche, M. A., & Yavari, F. (2025). The

1688   impact of temporal distribution on fear extinction learning. *International Journal of*

1689   *Clinical and Health Psychology*, *25*(1), 100536.

1690   Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: A manifesto for change. *Journal*

1691   *of Experimental Psychology: General*, *142*(4), 1180.

1692   Manning, J. R. (2023). How Can I Identify Stimulus-Driven Neural Activity Patterns in

1693   Multi-Patient ECoG Data? In N. Axmacher (Ed.), *Intracranial EEG: A Guide for*

1694   *Cognitive Neuroscientists* (pp. 803–836). Springer International Publishing.

1695   https://doi.org/10.1007/978-3-031-20910-9_48

1696   Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: Implications for fear

1697   conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*, *14*(6),

1698   417–428. https://doi.org/10.1038/nrn3492

1699   Maren, S., & Quirk, G. J. (2004). Neuronal signalling of fear memory. *Nature Reviews*

1700   *Neuroscience*, *5*(11), 844–852. https://doi.org/10.1038/nrn1535

1701   Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.

1702   *Journal of Neuroscience Methods*, *164*(1), 177–190.

1703   https://doi.org/10.1016/j.jneumeth.2007.03.024

1704   McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are

1705   complementary learning systems in the hippocampus and neocortex: Insights from the

1706   successes and failures of connectionist models of learning and memory.

1707   *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-

1708   295X.102.3.419

1709   McNally, R. J. (2007). Mechanisms of exposure therapy: How neuroscience can improve

1710   psychological treatments for anxiety disorders. *New Approaches to the Study of*

1711   *Change in Cognitive Behavioral Therapies*, *27*(6), 750–759.

1712   https://doi.org/10.1016/j.cpr.2007.01.003

1713   Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah,

1714   Y. H., Vuilleumier, P., Gil-Nagel, A., & Strange, B. A. (2016). A fast pathway for

1715   fear in human amygdala. *Nature Neuroscience*, *19*(8), 1041–1049.

1716   Michelmann, S., Bowman, H., & Hanslmayr, S. (2016). The Temporal Signature of

1717   Memories: Identification of a General Mechanism for Dynamic Memory Replay in

1718   Humans. *PLOS Biology*, *14*(8), e1002528.

1719   https://doi.org/10.1371/journal.pbio.1002528

1720   Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational

1721   neuroscience: Ten years of progress. *Annual Review of Psychology*, *63*(1), 129–151.

1722   Miller, J. F., Neufang, M., Solway, A., Brandt, A., Trippel, M., Mader, I., Hefft, S., Merkow,

1723   M., Polyn, S. M., Jacobs, J., Kahana, M. J., & Schulze-Bonhage, A. (2013). Neural

1724   Activity in Human Hippocampal Formation Reveals the Spatial Context of Retrieved

1725   Memories. *Science*, *342*(6162), 1111–1114. https://doi.org/10.1126/science.1244056

1726   Mueller, E. M., Panitz, C., Hermann, C., & Pizzagalli, D. A. (2014). Prefrontal Oscillations

1727   during Recall of Conditioned and Extinguished Fear in Humans. *The Journal of*

1728   *Neuroscience*, *34*(21), 7059. https://doi.org/10.1523/JNEUROSCI.3427-13.2014

1729   Neumann, D. L., & Kitlertsirivatana, E. (2010). Exposure to a novel context after extinction

1730   causes a renewal of extinguished conditioned responses: Implications for the

1731   treatment of fear. *Behaviour Research and Therapy*, *48*(6), 565–570.

1732   https://doi.org/10.1016/j.brat.2010.03.002

1733 Neumann, D. L., Lipp, O. V., & Cory, S. E. (2007). Conducting extinction in multiple

1734   contexts does not necessarily attenuate the renewal of shock expectancy in a fear-

1735   conditioning procedure with humans. *Behaviour Research and Therapy*, *45*(2), 385–

1736   394. https://doi.org/10.1016/j.brat.2006.02.001

1737 Nguyen, R., Koukoutselos, K., Forro, T., & Ciocchi, S. (2023). Fear extinction relies on

1738   ventral hippocampal safety codes shaped by the amygdala. *Science Advances*, *9*(22),

1739   eadg4881. https://doi.org/10.1126/sciadv.adg4881

1740 Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source

1741   software for advanced analysis of MEG, EEG, and invasive electrophysiological data.

1742   *Computational Intelligence and Neuroscience*, *2011*, 156869.

1743   https://doi.org/10.1155/2011/156869

1744 Orsini, C. A., Kim, J. H., Knapska, E., & Maren, S. (2011). Hippocampal and prefrontal

1745   projections to the basal amygdala mediate contextual regulation of fear after

1746   extinction. *Journal of Neuroscience*, *31*(47), 17269–17277.

1747 Orsini, C. A., & Maren, S. (2012). Neural and cellular mechanisms of fear and extinction

1748   memory formation. *Memory Formation*, *36*(7), 1773–1802.

1749   https://doi.org/10.1016/j.neubiorev.2011.12.014

1750 Pacheco Estefan, D. (2023). How Can We Track Cognitive Representations with Deep

1751   Neural Networks and Intracranial EEG? In N. Axmacher (Ed.), *Intracranial EEG: A*

1752   *Guide for Cognitive Neuroscientists* (pp. 849–862). Springer International Publishing.

1753   https://doi.org/10.1007/978-3-031-20910-9_50

1754 Pacheco Estefan, D., Sánchez-Fibla, M., Duff, A., Principe, A., Rocamora, R., Zhang, H.,

1755   Axmacher, N., & Verschure, P. F. M. J. (2019). Coordinated representational

1756   reinstatement in the human hippocampus and lateral temporal cortex during episodic

1757   memory retrieval. *Nature Communications*, *10*(1), 1–13.

1758 Pacheco Estefan, D., Zucca, R., Arsiwalla, X., Principe, A., Zhang, H., Rocamora, R.,

1759   Axmacher, N., & Verschure, P. F. M. J. (2021). Volitional learning promotes theta

1760   phase coding in the human hippocampus. *Proceedings of the National Academy of*

1761   *Sciences*, *118*(10).

1762 Pacheco-Estefan, D., Fellner, M.-C., Kunz, L., Zhang, H., Reinacher, P., Roy, C., Brandt, A.,

1763   Schulze-Bonhage, A., Yang, L., Wang, S., Liu, J., Xue, G., & Axmacher, N. (2024).

1764   Maintenance and transformation of representational formats during working memory

1765   prioritization. *Nature Communications*, *15*(1), 8234. https://doi.org/10.1038/s41467-

1766   024-52541-w

1767 Paré, D., & Collins, D. R. (2000). Neuronal Correlates of Fear in the Lateral Amygdala:

1768   Multiple Extracellular Recordings in Conscious Cats. *The Journal of Neuroscience*,

1769   *20*(7), 2701. https://doi.org/10.1523/JNEUROSCI.20-07-02701.2000

1770 Parvizi, J., & Kastner, S. (2018). Promises and limitations of human intracranial

1771   electroencephalography. *Nature Neuroscience*, *21*(4), 474–483.

1772 Pirazzini, G., Starita, F., Ricci, G., Garofalo, S., di Pellegrino, G., Magosso, E., & Ursino, M.

1773   (2023). Changes in brain rhythms and connectivity tracking fear acquisition and

1774   reversal. *Brain Structure and Function*, *228*(5), 1259–1281.

1775   https://doi.org/10.1007/s00429-023-02646-7

1776 Popescu, A. T., Popa, D., & Paré, D. (2009). Coherent gamma oscillations couple the

1777   amygdala and striatum during learning. *Nature Neuroscience*, *12*(6), 801–807.

1778   https://doi.org/10.1038/nn.2305

Qasim, S. E., Mohan, U. R., Stein, J. M., & Jacobs, J. (2023). Neuronal activity in the human amygdala and hippocampus enhances emotional memory encoding. *Nature Human Behaviour*, *7*(5), 754–764. https://doi.org/10.1038/s41562-022-01502-8

Quinn, J. J., Ma, Q. D., Tinsley, M. R., Koch, C., & Fanselow, M. S. (2008). Inverse temporal contributions of the dorsal hippocampus and medial prefrontal cortex to the expression of long-term fear memories. *Learning & Memory*, *15*(5), 368–372.

Quirk, G. J., & Mueller, D. (2008). Neural Mechanisms of Extinction Learning and Retrieval. *Neuropsychopharmacology*, *33*(1), 56–72. https://doi.org/10.1038/sj.npp.1301555

Rahman, M. M., Shukla, A., & Chattarji, S. (2018). Extinction recall of fear memories formed before stress is not affected despite higher theta activity in the amygdala. *eLife*, *7*, e35450. https://doi.org/10.7554/eLife.35450

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. *Classical Conditioning, Current Research and Theory*, *2*, 64–69.

Rivière, D., Geffroy, D., Denghien, I., Souedet, N., & Cointepas, Y. (2011). *Anatomist: A python framework for interactive 3D visualization of neuroimaging data*. 3–4.

Rolls, E. T. (1999). Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, *9*(4), 467–480.

Saint Amour di Chanaz, L., Pérez-Bellido, A., Wu, X., Lozano-Soldevilla, D., Pacheco-Estefan, D., Lehongre, K., Conde-Blanco, E., Roldan, P., Adam, C., Lambrecq, V., Frazzini, V., Donaire, A., Carreño, M., Navarro, V., Valero-Cabré, A., & Fuentemilla, L. (2023). Gamma amplitude is coupled to opposed hippocampal theta-phase states during the encoding and retrieval of episodic memories in humans. *Current Biology*, *33*(9), 1836-1843.e6. https://doi.org/10.1016/j.cub.2023.03.073

Santini, E., Quirk, G. J., & Porter, J. T. (2008). Fear Conditioning and Extinction Differentially Modify the Intrinsic Excitability of Infralimbic Neurons. *The Journal of Neuroscience*, *28*(15), 4028. https://doi.org/10.1523/JNEUROSCI.2623-07.2008

Seger, S. E., Kriegel, J. L. S., Lega, B. C., & Ekstrom, A. D. (2023). Memory-related processing is the primary driver of human hippocampal theta oscillations. *Neuron*, *111*(19), 3119-3130.e4. https://doi.org/10.1016/j.neuron.2023.06.015

Seidenbecher, T., Laxmi, T. R., Stork, O., & Pape, H.-C. (2003). Amygdalar and Hippocampal Theta Rhythm Synchronization During Fear Memory Retrieval. *Science*, *301*(5634), 846–850. https://doi.org/10.1126/science.1085818

Shiban, Y., Pauli, P., & Mühlberger, A. (2013). Effect of multiple context exposure on renewal in spider phobia. *Behaviour Research and Therapy*, *51*(2), 68–74. https://doi.org/10.1016/j.brat.2012.10.007

Sperl, M. F. J., Panitz, C., Rosso, I. M., Dillon, D. G., Kumar, P., Hermann, A., Whitton, A. E., Hermann, C., Pizzagalli, D. A., & Mueller, E. M. (2019). Fear Extinction Recall Modulates Human Frontomedial Theta and Amygdala Activity. *Cerebral Cortex*, *29*(2), 701–715. https://doi.org/10.1093/cercor/bhx353

Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., Elger, C. E., Axmacher, N., & Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience*, *18*(11), 1679–1686. https://doi.org/10.1038/nn.4119

Staresina, B. P., Michelmann, S., Bonnefond, M., Jensen, O., Axmacher, N., & Fell, J. (2016). Hippocampal pattern completion is linked to gamma power increases and alpha power decreases during recollection. *eLife*, *5*, e17397. https://doi.org/10.7554/eLife.17397

1827   Stujenske, J. M., Likhtik, E., Topiwala, M. A., & Gordon, J. A. (2014). Fear and Safety

1828   Engage Competing Patterns of Theta-Gamma Coupling in the Basolateral Amygdala.

1829   *Neuron*, *83*(4), 919–933. https://doi.org/10.1016/j.neuron.2014.07.026

1830   Szeska, C., Richter, J., Wendt, J., Weymar, M., & Hamm, A. O. (2020). Promoting long-term

1831   inhibition of human fear responses by non-invasive transcutaneous vagus nerve

1832   stimulation during extinction training. *Scientific Reports*, *10*(1), 1529.

1833   https://doi.org/10.1038/s41598-020-58412-w

1834   Taschereau-Dumouchel, V., Kawato, M., & Lau, H. (2020). Multivoxel pattern analysis

1835   reveals dissociations between subjective fear and its physiological correlates.

1836   *Molecular Psychiatry*, *25*(10), 2342–2354. https://doi.org/10.1038/s41380-019-0520-

1837   3

1838   Taub, A. H., Perets, R., Kahana, E., & Paz, R. (2018). Oscillations synchronize amygdala-to-

1839   prefrontal primate circuits during aversive learning. *Neuron*, *97*(2), 291–298.

1840   Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*,

1841   *53*(1), 1–25.

1842   Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity

1843   predicts long-term fear memory. *Nature Neuroscience*, *16*(4), 388–390.

1844   https://doi.org/10.1038/nn.3345

1845   Wang, Y., Olsson, S., Lipp, O. V., & Ney, L. J. (2024). Renewal in human fear conditioning:

1846   A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *159*,

1847   105606. https://doi.org/10.1016/j.neubiorev.2024.105606

1848   Weber, J., Iwama, G., Solbakk, A.-K., Blenkmann, A. O., Larsson, P. G., Ivanovic, J.,

1849   Knight, R. T., Endestad, T., & Helfrich, R. (2023). Subspace partitioning in the

1850   human prefrontal cortex resolves cognitive interference. *Proceedings of the National*

1851   *Academy of Sciences*, *120*(28), e2220523120.

1852   https://doi.org/10.1073/pnas.2220523120

1853   Wu, X., Packard, P. A., García-Arch, J., Bunzeck, N., & Fuentemilla, L. (2023). Contextual

1854   incongruency triggers memory reinstatement and the disruption of neural stability.

1855   *NeuroImage*, *273*, 120114. https://doi.org/10.1016/j.neuroimage.2023.120114

1856   Xie, W., Wittig, J. H., & Zaghloul, K. A. (2023). What Do I Need to Consider for

1857   Multivariate Analysis of iEEG Data? In N. Axmacher (Ed.), *Intracranial EEG: A*

1858   *Guide for Cognitive Neuroscientists* (pp. 557–566). Springer International Publishing.

1859   https://doi.org/10.1007/978-3-031-20910-9_34

1860   Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater

1861   Neural Pattern Similarity Across Repetitions Is Associated with Better Memory.

1862   *Science*, *330*(6000), 97–101. https://doi.org/10.1126/science.1193125

1863   Yaffe, R. B., Kerr, M. S. D., Damera, S., Sarma, S. V., Inati, S. K., & Zaghloul, K. A. (2014).

1864   Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal

1865   dynamics during successful memory retrieval. *Proceedings of the National Academy*

1866   *of Sciences*, *111*(52), 18727–18732. https://doi.org/10.1073/pnas.1417017112

1867   Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual

1868   binding theory of episodic memory: Systems consolidation reconsidered. *Nature*

1869   *Reviews Neuroscience*, *20*(6), 364–375. https://doi.org/10.1038/s41583-019-0150-4

1870   Zhang, H., Skelin, I., Ma, S., Paff, M., Mnatsakanyan, L., Yassa, M. A., Knight, R. T., & Lin,

1871   J. J. (2024). Awake ripples enhance emotional memory encoding in the human brain.

1872   *Nature Communications*, *15*(1), 215. https://doi.org/10.1038/s41467-023-44295-8

1873   Zhang, X., Kim, J., & Tonegawa, S. (2020). Amygdala Reward Neurons Form and Store Fear

1874   Extinction Memory. *Neuron*, *105*(6), 1077-1093.e7.

1875   https://doi.org/10.1016/j.neuron.2019.12.025

1876   Zheng, J., Anderson, K. L., Leal, S. L., Shestyuk, A., Gulsen, G., Mnatsakanyan, L., Vadera,

1877   S., Hsu, F. P. K., Yassa, M. A., Knight, R. T., & Lin, J. J. (2017). Amygdala-

1878   hippocampal dynamics during salient information processing. *Nature*

1879   *Communications*, *8*(1), 14413. https://doi.org/10.1038/ncomms14413

1880
1881
1882