1   # Learning and language in the unconscious human hippocampus

2

3   Kalman A. Katlowitz[1], Shraddha Shah[1], Melissa C. Franch[1], Joshua Adkinson[1], James L.
4   Belanger[1,2], Raissa K. Mathura[1], Domokos Meszéna[3,4,5], Elizabeth A. Mickiewicz[1], Matthew
5   McGinley[6], William Muñoz[7], Garrett P. Banks[1], Sydney S. Cash[7], Chih-Wei Hsu[8], Angelique C.
6   Paulk[7], Nicole R. Provenza[1,9], Andrew Watrous[1], Ziv Williams[7], Sarah R. Heilbronner[1], Robert
7   Kim[10], Nuttida Rungratsameetaweemana[11], Benjamin Y. Hayden*[1,9], Sameer A. Sheth*[1,9]

8

9   *these authors contributed equally.*

10

11   [1] Department of Neurosurgery, Baylor College of Medicine, Houston, TX, USA
12   *and* Neuroengineering Initiative, Rice University, Houston, TX, USA
13   [2] Department of Linguistics, Rice University, Houston, TX, USA
14   [3] Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston
15   MA USA
16   [4] HUN-REN Research Centre for Natural Sciences, Budapest, Hungary
17   [5] PPCU Faculty of Information Technology and Bionics, Budapest, Hungary
18   [6] Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA
19   [7] Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School,
20   Boston MA USA
21   [8] Department of Integrative Physiology, Baylor College of Medicine, Houston, TX, USA
22   [9] Department of Electrical & Computer Engineering, Rice University, Houston, TX, USA
23   [10] Department of Neurology, Cedars-Sinai Medical Center, Los Angeles CA, USA
24   [11] Department of Biomedical Engineering, Columbia University, New York, NY USA
25

26 **ABSTRACT**

27   Consciousness is a fundamental component of cognition,[1] but the degree to which higher-

28 order perception relies on it remains disputed.[2,3] Here we demonstrate the persistence of learning,

29 semantic processing, and online prediction in individuals under general anesthesia-induced loss

30 of consciousness.[4,5] Using high-density Neuropixels microelectrodes[6] to record neural activity in

31 the human hippocampus while playing a series of tones to anesthetized patients, we found that

32 hippocampal neurons could reliably detect oddball tones. This effect size grew over the course of

33 the experiment (~10 minutes), consistent with learning effects. A biologically plausible recurrent

34 neural network model showed that learning and oddball representation are an emergent property

35 of flexible tone discrimination. Last, when we played language stimuli, single units and

36 ensembles carried information about the semantic and grammatical features of natural speech,

37 even predicting semantic information about upcoming words. Together these results indicate that

38 in the hippocampus, which is anatomically and functionally distant from primary sensory

39 cortices,[7] complex processing of sensory stimuli occurs even in the unconscious state.

40 **MAIN TEXT**

41

42 **Neuropixels recordings in the human hippocampus**

43   We performed intraoperative hippocampal recordings with Neuropixels probes[6] in five

44 patients (**Supplemental Table 1**) undergoing anterior temporal lobectomies for drug resistant

45 epilepsy. One patient also had recordings at two separate depths in the middle temporal gyrus

46 prior to the neocortical resection. Across these seven recordings, we isolated 555 units (295

47 single units, 260 multi-units; mean: 79.3 units per recording; range: 22-172). Hippocampal

48 recordings were conducted in the anterior body after resection of the lateral temporal cortex and

49 prior to resection of the mesial temporal structures such as parahippocampal gyrus and amygdala

50 (**Figure 1A**). In the hippocampus, we isolated 405 units (157 single units, 248 multi-units, mean:

51 81 units per recording; range: 22-172). Based on coregistration between anatomical maps and

52 preoperative imaging, postoperative high-resolution CT, and electrophysiological properties, we

53 expect our units to be drawn from the dentate gyrus, CA4, and CA1 (**Figure 1B-D**).[8]

54   Average firing rates were lower for hippocampus (1.6 +/- 1.2 Hz) than for temporal

55 cortex (2.5 +/- 1.7 Hz, p<0.0001, Student's t-test).[9] Motion artifacts, a major challenge for

56 human cortical Neuropixels recordings,[10] were markedly less conspicuous in hippocampal

57 recordings than in the cortical recordings (**Figure 1E**). This increased stability may be due to the

58 central location of the hippocampus within the brain, and because it is anchored by the dura of

59 the middle fossa via the parahippocampal gyrus. Consistent with this hypothesis, the reduction in

60 motion was especially clear when we compare the respiratory and heartbeat frequency bands

61 (p=0.001, t-test on power between 0.1 to 3 Hz of motion trajectories between the hippocampal

62 and cortical recordings). After a brief baseline recording, we conducted recordings during

63   presentation of auditory stimuli composed of pure tones (3 patients) or a continuous podcast (2

64   patients, **Figure 1F**).

65

66   **Auditory environment monitoring in the anesthetized patient**

67   The ability to recognize patterns and detect violations of those patterns is a hallmark of

68   cognition. In the classic oddball task,[11,12] participants are presented with a series of stimuli that

69   constitute a pattern (e.g., a series of tones of the same frequency) interspersed with deviant

70   "oddballs". In three of the hippocampal recordings (P4, P5, and P6), we played a series of 100

71   ms pure tones; 20% were oddballs (higher or lower frequency than standard, **Figure 2A,**

72   **Methods**). Most units (n=122/172, 70.9%, signed-rank test, α=0.05) showed tone-evoked

73   responses (**Figure 2B**), consistent with established auditory responses within hippocampus.[13]

74   Neural responses to tones often showed a biphasic temporal firing rate profile (**Figure 2B**).

75   Across all units, response latencies showed a clear bimodal temporal dynamic (Gaussian Mixture

76   Model fit via Expectation Maximization, **Figure 2C**). Hippocampal units encoded tone

77   frequency (n=39/172, 22.7% of units, rank sum test, α=0.05, **Figure 2D**).

78   Having established auditory responses despite the anesthetic state, we next examined the

79   representation of stimulus features. For two patients, we balanced tone frequency and oddball

80   status (**Figure 2A,** n=150 units). At the single unit (**Figure 2E**) and population (**Figure 2F**)

81   levels, neuronal responses differentiated standards from oddballs. This divergence was most

82   notable within the first 300 ms, with 24.7% (n=37/150) of units signaling oddballs. Thus, further

83   analyses focused on this first time segment. Local field potentials (LFPs) also showed oddball-

84   evoked responses, observed as a negative deflection in the evoked response potential (ERP,

85   **Figure 2G**) and an increase in gamma amplitude (**Figure 2H**).

86   Next, z-scored sensory responses for all units were modelled as a function of tone

87   frequency, context (standard vs. oddball), and their interaction using linear regression. We

88   observed comparable encoding for all terms: 29.3% of units showed tone encoding; 24.7%

89   showed oddball encoding; 22.7% showed an interaction. The absolute values of the beta weights

90   for the oddball term were greater than the corresponding tone and mixed selectivity terms (paired

91   t-test on absolute values, p<0.0001 for both, **Figure 2I**). We found similar proportions of units

92   with a significant oddball effect (n=43) in P5 (37/127, 29.1%) and P6 (6/23, 26.1%) (p=0.8, χ^2

93   test). Mean broadband LFP power and gamma band amplitude also demonstrated tone, oddball,

94   and mixed selectivity at similar rates across channels (broadband LFP: 40.9%, 47.2%, and

95   46.0%; gamma: 20.1%, 17.6%, and 18.7%, respectively).

96   Leveraging the power of large-scale recordings, we used a 10-fold cross-validated

97   support vector machine (SVM) to decode stimulus features on a trial-by-trial level across the

98   neuronal population. Tone identity was robustly represented in both patients across units,

99   broadband LFP, and gamma power, with mean accuracy ranging between 0.6 and 0.76 (p<0.001

100   for all, t-test. **Figure 2J**). Oddball identity could also be decoded above chance for both patients

101   (p<0.05 for all except for ERP and unit decoding in P6), albeit at lower levels, ranging from 0.52

102   to 0.56 (accuracy rates on shuffled data ranged from 0.496 to 0.503).

103

Hippocampal signatures of learning in the unconscious state

105   While the oddball task by definition relies on a working memory of the statistical

106   distribution of recent tones,[14] this does not prove that the unconscious hippocampus was learning

107   the task structure. We thus examined the temporal evolution of the oddball identity

108   representation. In oddball-responsive units (n=43), we found that the oddball response grew in

109   magnitude over the course of the experiment (~10 minutes, example unit, **Figure 3A)**. Splitting

110   our task into halves, we found a significant increase in oddball encoding for both patients (P5:

111   p=0.01, P6: p<0.001, t-test, **Figure 3C**). Surprisingly, we also observed a concomitant decrease

112   in frequency encoding, raising the possibility of compensatory mechanisms (p<0.001, t-test for

113   both) (**Figure 3B**). Using a sliding window of subsets of 50 trials, we found a continuous

114   increase in oddball decoding accuracy across the approximately 10-minute duration of the

115   experiment (p<0.001, Pearson's correlation, **Figure 3D, purple**). Again, this increase in oddball

116   performance was accompanied by an initial decrease in tone encoding (p<0.0001, **Figure 3D,**

117   **green**), [15] demonstrating the neural population was sacrificing its tone responses for the sake of

118   oddball representations over the course of the experiment, suggesting that the hippocampal

119   responses were shifting to represent the salient features of the stimulus.[16]

120   We created neural vectors of the average standard tone response as well as each

121   individual oddball trial (43-dimensional vectors composed of the mean response of the oddball

122   units). We found a gradual divergence in Euclidean distance between standard and oddball

123   vectors over the course of the session (r=0.34, p=0.007, Pearson's correlation; **Figure 3E, left**).

124   Discriminability was even stronger when considering cosine angle, indicating the effect is not

125   merely a consequence of a response gain in oddball cells (r=0.5, p=0.0002; **Figure 3E, right**).

126   These effects were mostly consistent for individual patients (P5 distance: r=0.25; p=0.056, angle:

127   r=0.43 p=0.002; P6 distance: r=0.32, p=0.012, angle: r=0.48; p=0.0002). These results indicate

128   that the hippocampus does not simply improve encoding using gain modulation;[17] instead,

129   oddball responses reflect a rotation of the neural population vector in a high dimensional space,

130   meaning that oddball learning alters the warping of the neural response manifold.[18] Thus,

131   complex reshaping of responses can occur even under general anesthesia.[19,20]

132   To gain further mechanistic insight at the level of individual units, we turned to a

133   continuous-rate recurrent neural network (RNN) trained to perform a signal-detection task

134   similar to the task used for the human Neuropixels data (**Figure 3F**).[21] The network model

135   underwent three stages of training, simulating the different contexts used in the experimental

136   data, with the prevalence of specific tones varied at each stage (**Figure 3G, H, Methods**). Tone

137   A was presented to the network in 80% of the trials in the first stage, followed by a washout

138   period, and then a third stage with probabilities reversed relative to the first. By the end of

139   training (range of 1400 to 2600 trials) the model was able to reliably differentiate tone identities

140   (**Figure 3H**). Notably, despite being only explicitly trained on tone frequency discrimination, the

141   model was able to perform not only frequency discrimination (tone frequency, $p<0.005$ signed

142   Wilcoxon test vs. shuffled data) but also context (oddball identity, $p<0.005$, signed Wilcoxon test

143   vs. shuffled data, **Figure 3I**). The model also recreated the pattern observed in the Euclidean and

144   vector angle distance between standard and oddball representations (**Figure 3J**), suggesting that

145   the divergence of representations can be due to local computations rather than inherited from

146   other networks.

147

148   **Unconscious encoding of semantics and grammar in the hippocampus**

149   We next tested whether the unconscious hippocampus could perform even higher order

150   functions associated with parsing semantic and syntactic features of natural speech. In two

151   participants (P6 and P8), we recorded neural activity while playing 10-20 minutes of podcast

152   episodes (see **Methods**[22]). We aligned neural activity to word onset and offset (n=3024 words

153   for P6 and n=1565 words for P8), and computed word-evoked neural responses (example unit,

154   average response to all words presented, **Figure 4A**). Given the oddball effects described above,

155 we first hypothesized that the brain would respond differentially based on word lexical

156 frequency, which we defined using a standard database.[23] All 195 units had a robust correlation

157 between lexical frequency and neuronal firing rate (**Figure 4B,** mean r=0.55+/- 0.08, Spearman's

158 correlation, α<0.05). To address possible confounds between word duration and frequency, we

159 reran the analysis with subsets of words within a limited duration range, i.e. 0-200 ms, 200-400

160 ms, 400-600 ms, and consistently observed a positive correlation (p<0.001 for combined units).

161 Additionally, a linear model that incorporated both logarithmic word duration and logarithmic

162 word frequency still found significance in word frequency as a predictor of firing rate (p<0.001,

163 t-test on coefficients). This correlation could not be solely explained by difficulty in lexical

164 access, as there was also a consistent relationship of the neural responses with the relative

165 surprise of each word (r=0.11 +/- 0.03, 187/195 units significant at α<0.05), a metric that

166 evaluates the relative probability of each word as a function of the prior words.[24]

167 These results suggest that the unconscious hippocampus has access to the semantic

168 information conveyed by each word. To explicitly test this possibility, we regressed the firing

169 rates of each neuron against the semantic embeddings of each word that demonstrated a response

170 (see **Methods**).[22,25,26] In semantic embedding space, similar words (e.g. 'dog' and 'cat') are

171 closer (Euclidean distance, d=1.8) than dissimilar words (e.g. 'dog' and 'pen', d=2.5). Using 10-

172 fold cross-validation, we found that the RMSE of a linear model outperformed shuffled data in

173 all units (α=0.05, one tailed t-test on real versus shuffled RMSE **Figure 4C**), with an average

174 correlation between true and predicted firing rates of 0.46 +/- 0.07 (n=195 units). However,

175 given that conversational English has many words that are repeated these results could be

176 confounded by the fact that cells had consistent responses to words, perhaps even matching

177 acoustic features. To show that units generalize across word embeddings, we re-ran the analysis

178   using only unique words (n=746 and 573 unique words for P5 and P6, respectively). We found a

179   significant result in 84.1% of the recorded units (159/189 units with at least 50 words that had a

180   non-zero response), with an average correlation of r=0.17 +/-0.08, **Figure 4D**). In other words, it

181   is possible to predict the firing rate of units to a given word based on responses to other words by

182   leveraging their similarities in semantic space,[27] demonstrating that the unconscious

183   hippocampus has access to abstract semantic relationships between words.

184   We then analyzed the representation of word features. We semantically categorized each

185   word into one of 12 possible groups (**Figure 4E**).[22] Nearly all units (193/195) showed some form

186   of semantic category selectivity (α=0.05, Kruskal Wallis test for any difference between

187   semantic categories). Rank-sum tests for each category versus all others showed that units were

188   selective for multiple semantic categories, consistent with our previously reported findings in

189   awake patients (corrected for multiple comparisons, α<0.05).[22] Specifically, 165/195 (84.6%)

190   units were able to discriminate between at least two of the twelve semantic categories and 76/195

191   (39.0%) were able to discriminate across at least four (**Figure 4G**), with a median of three

192   categories per neuron. We also investigated encoding of grammatical features. We classified

193   each word into a part of speech using the Stanford CoreNLP toolkit[28] (**Figure 4H**, n=2906 words

194   for P5 and 1497 for P8). We found that 191/195 units carried information about part of speech

195   (α=0.05, Kruskal Wallis test). Again, there was broad representation of different categories

196   (**Figure 4I**). Interestingly, nearly all units (P6: 82.6%; P8: 94.8%) distinguished nouns from non-

197   nouns, but only a few (P6: 4.3%; P8: 5.8%) distinguished verbs from non-verbs, consistent with

198   the greater role of the hippocampus in object over action representations.[29] Overall, the median

199   number of categories represented was four (out of 11 possible, **Figure 4J**), with 178/195 (91.3%)

200   units discriminating at least two categories and 100/195 (51.30%) discriminating across at least

201   four. Interestingly, we found a strong correlation between the number of semantic categories and

202   the number of part of speech categories represented across neurons (r=0.38, p<0.001 Spearman's

203   correlation) suggesting that language responsive neurons can represent multiple features, with no

204   evidence of separation of the two tasks.

205   Relying on statistical differences in distributions, however, would not be sufficiently

206   accurate for online processing of speech. To study the ability of the hippocampal network to

207   provide real time information about language we examined its decoding ability on a word-by-

208   word basis. We used a SVM to compare each category against all others. We found that all

209   categories in both semantics and part of speech were decodable at the level of individual words

210   (p<0.001 vs. shuffled data performance at chance rates of 0.5, **Figure 4K and L**). Semantic

211   categories had a higher average classification accuracy (60.5 +/- 4.0%) than part of speech

212   categories (56.5 +/- 5.3%, p=0.03, t-test). These results indicate that both semantic and syntactic

213   information (independent of the acoustic features) is encoded in real time within the unconscious

214   hippocampus.

215   We next asked whether the unconscious hippocampus could represent recent or

216   upcoming words, a fundamental aspect of speech comprehension.[30] We reran our linear

217   regression analysis (**Figure 4C, D**) but instead of predicting neural data using the word being

218   played, we tested previous and upcoming words. Here we found that neural responses

219   corresponded to not only semantic features of prior words (**Figure 4M**, negative indices), which

220   could be due to short term memory[31] or even hysteresis back to baseline, but also to the

221   semantics of future words[32] (**Figure 4M**, positive indices). Future words were able to be decoded

222   nearly as well as past words, though with a 21.3% larger tail for past words at τ_{past}=0.97 versus

223   future words at τ_{future}=0.81. These findings demonstrate that not only is recent speech being

224   actively tracked, it is also being used to predict future words, a high level cognitive function

225   crucial to speech comprehension that depends on engagement of the language network.[33]

226   Notably, this analysis also precludes the possibility that the responses are solely due to the

227   underlying acoustics, as it disconnects the speech features from the ongoing auditory

228   information.

229

230   **DISCUSSION**

231   Our study identified neural signatures of learning, semantic processing, and prediction in

232   the unconscious human hippocampus, core cognitive functions often assumed to be absent in the

233   unconscious state.[34] These analyses do not have obvious explanations based solely on low-level

234   sensory responses. The long and slow increase in oddball detection over the course of 10 minutes

235   is unlikely to reflect adaptation or repetition suppression,[35] and can even emerge from local

236   circuit properties based on our modelling results. Additionally, the representation of semantic

237   features of adjacent words requires more than just the ongoing acoustic information. We

238   therefore show that within anesthetic induced unconsciousness it is not sensory integration that is

239   completely blocked[36] but rather its ability to consolidate into explicit memories.[37,38] These results

240   provide the foundation for previous reports of post-anesthesia implicit recall,[39,40] which would

241   depend on sensory processing and memory despite the absence of consciousness.

242   These results also complement a growing body of work showing the central importance

243   of the hippocampus in language processing.[22,41,42] While the hippocampus is not considered part

244   of the classic cortical language network,[7] its ability to flexibly associate different features[43,44] and

245   perform online prediction,[45] as well as its established importance for pattern separation and

246   completion,[46] make it a likely site for semantics and composition. In this study we not only

247 provide further evidence for semantic and grammar representations within the hippocampus, we

248 even demonstrate continuous prediction of upcoming words. Our results therefore extend these

249 language models of hippocampal computations by showing they are sufficiently robust that they

250 do not even require conscious awareness.

251

252 The datasets generated during and/or analysed during the current study are available from the

253 corresponding author on reasonable request.

254 Supplementary Information is available for this paper.

255

256   1. van Gaal, S., De Lange, F. P. & Cohen, M. X. The role of consciousness in cognitive control

257   and decision making. *Front. Hum. Neurosci.* **6**, (2012).

258   2. Dehaene, S., Lau, H. & Kouider, S. What is consciousness, and could machines have it?

259   *Science* **358**, 486–492 (2017).

260   3. Sikkens, T., Bosman, C. A. & Olcese, U. The Role of Top-Down Modulation in Shaping

261   Sensory Processing Across Brain States: Implications for Consciousness. *Front. Syst.*

262   *Neurosci.* **13**, (2019).

263   4. Alkire, M. T., Hudetz, A. G. & Tononi, G. Consciousness and Anesthesia. *Science* **322**,

264   876–880 (2008).

265   5. Brown, E. N., Lydic, R. & Schiff, N. D. General Anesthesia, Sleep, and Coma. *N. Engl. J.*

266   *Med.* **363**, 2638–2650 (2010).

267   6. Coughlin, B. *et al.* Modified Neuropixels probes for recording human neurophysiology in the

268   operating room. *Nat Protoc* **18**, 2927–2953 (2023).

269   7. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev.*

270   *Neurosci.* **8**, 393–402 (2007).

271   8. DeKraker, J., Köhler, S. & Khan, A. R. Surface-based hippocampal subfield segmentation.

272   *Trends Neurosci.* **44**, 856–863 (2021).

273   9. Buzsaki, G. *Rhythms of the Brain*. (Oxford University Press, New York, NY, 2006).

274   10. Chung, J. E. *et al.* High-density single-unit human cortical recordings using the Neuropixels

275   probe. *Neuron* **110**, 2409-2421.e3 (2022).

276   11. Tzovara, A. *et al.* Predictable and unpredictable deviance detection in the human

277   hippocampus and amygdala. *Cereb Cortex* **34**, bhad532 (2024).

278   12. García-Larrea, L., Lukaszewicz, A. C. & Mauguière, F. Revisiting the oddball paradigm.

279   Non-target vs neutral stimuli and the evaluation of ERP attentional effects.

280   *Neuropsychologia* **30**, 723–741 (1992).

281   13. Billig, A. J., Lad, M., Sedley, W. & Griffiths, T. D. The hearing hippocampus. *Prog.*

282   *Neurobiol.* **218**, 102326 (2022).

283   14. Näätänen, R., Paavilainen, P. & Reinikainen, K. Do event-related potentials to infrequent

284   decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neurosci.*

285   *Lett.* **107**, 347–352 (1989).

286   15. Kelemen, E. & Fenton, A. A. Dynamic Grouping of Hippocampal Neural Activity During

287   Cognitive Control of Two Spatial Frames. *PLoS Biol.* **8**, e1000403 (2010).

288   16. Fontanini, A. & Katz, D. B. Behavioral States, Network States, and Sensory Response

289   Variability. *J. Neurophysiol.* **100**, 1160–1168 (2008).

290   17. Treue, S. & Maunsell, J. H. Effects of attention on the processing of motion in macaque

291   middle temporal and medial superior temporal visual cortical areas. *J Neurosci* **19**, 7591–

292   7602 (1999).

293   18. Ebitz, R. B. & Hayden, B. Y. The population doctrine in cognitive neuroscience. *Neuron* **109**,

294   3055–3068 (2021).

295   19. Mirabella, G. *et al.* Neurons in Area V4 of the Macaque Translate Attended Visual Features

296   into Behaviorally Relevant Categories. *Neuron* **54**, 303–318 (2007).

297   20. David, S. V., Hayden, B. Y., Mazer, J. A. & Gallant, J. L. Attention to Stimulus Features

298   Shifts Spectral Tuning of V4 Neurons during Natural Vision. *Neuron* **59**, 509–521 (2008).

299   21. Rungratsameetaweemana, N., Kim, R., Chotibut, T. & Sejnowski, T. J. Random noise

300   promotes slow heterogeneous synaptic dynamics important for robust working memory

301   computation. *Proc Natl Acad Sci U A* **122**, e2316745122 (2025).

302   22. Franch, M. *et al.* A vectorial code for semantics in human hippocampus. *bioRxiv* 2025–02

303   (2025).

304   23. Brysbaert, M. & New, B. Moving beyond Kučera and Francis: A critical evaluation of current

305   word frequency norms and the introduction of a new and improved word frequency measure

306   for American English. *Behav. Res. Methods* **41**, 977–990 (2009).

307   24. Radford, A. *et al.* Language Models are Unsupervised Multitask Learners. in (2019).

308   25. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in

309   vector space. *ArXiv CsCL* (2013).

310   26. Jamali, M. *et al.* Semantic encoding during language comprehension at single-cell

311   resolution. *Nature* **631**, 610–616 (2024).

312   27. Khanna, A. R. *et al.* Single-neuronal elements of speech production in humans. *Nature* **626**,

313   603–610 (2024).

314   28. Manning, C. D. *et al.* The Stanford CoreNLP Natural Language Processing Toolkit. in

315   *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics:*

316   *System Demonstrations* 55–60 (2014).

317   29. Mestres-Missé, A., Rodriguez-Fornells, A. & Münte, T. F. Neural differences in the mapping

318   of verb and noun concepts onto novel words. *Neuroimage* **49**, 2826–2835 (2010).

319   30. Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P. & de Lange, F. P. A hierarchy of

320   linguistic predictions during natural language comprehension. *Proc Natl Acad Sci U A* **119**,

321   e2201968119 (2022).

322   31. Jang, A. I., Wittig, J. H., Jr, Inati, S. K. & Zaghloul, K. A. Human cortical neurons in the

323   anterior temporal lobe reinstate spiking activity during verbal memory retrieval. *Curr Biol* **27**,

324   1700-1705.e5 (2017).

325   32. Pickering, M. J. & Gambi, C. Predicting while comprehending language: A theory and

326   review. *Psychol Bull* **144**, 1002–1044 (2018).

327   33. Ryskin, R. & Nieuwland, M. S. Prediction during language comprehension: what is next?

328   *Trends Cogn. Sci.* **27**, 1032–1052 (2023).

329   34. Franks, N. P. & Lieb, W. R. Mechanisms of general anesthesia. *Env. Health Perspect* **87**,

330   199–205 (1990).

331   35. Strange, B. A. & Dolan, R. J. Adaptive anterior hippocampal responses to oddball stimuli.

332   *Hippocampus* **11**, 690–698 (2001).

333   36. Krom, A. J. *et al.* Anesthesia-induced loss of consciousness disrupts auditory responses

334   beyond primary cortex. *Proc Natl Acad Sci U A* **117**, 11770–11780 (2020).

335   37. Moon, D. U. *et al.* Propofol modulates early memory consolidation in humans. *eNeuro* **7**,

336   ENEURO.0537-19.2020 (2020).

337   38. Iggena, D. *et al.* Post-encoding modulation of spatial memory consolidation by propofol.

338   *Cortex* **156**, 1–12 (2022).

339   39. Kihlstrom, J. F., Schacter, D. L., Cork, R. C., Hurt, C. A. & Behr, S. E. Implicit and Explicit

340   Memory following Surgical Anesthesia. *Psychol Sci* **1**, 303–306 (1990).

341   40. Linassi, F. *et al.* Implicit memory and anesthesia: A systematic review and meta-analysis.

342   *Life Basel* **11**, 850 (2021).

343   41. Covington, N. V. & Duff, M. C. Expanding the Language Network: Direct Contributions from

344   the Hippocampus. *Trends Cogn. Sci.* **20**, 869–870 (2016).

345   42. Dijksterhuis, D. E. *et al.* Pronouns reactivate conceptual representations in human

346   hippocampal neurons. *Science* **385**, 1478–1484 (2024).

347   43. Duff, M. C. & Brown-Schmidt, S. The hippocampus and the flexible use and processing of

348   language. *Front. Hum. Neurosci.* **6**, 69 (2012).

349   44. Olsen, R. K., Moses, S. N., Riggs, L. & Ryan, J. D. The hippocampus supports multiple

350   cognitive processes through relational binding and comparison. *Front. Hum. Neurosci.* **6**,

351   (2012).

352   45. Bonhage, C. E., Mueller, J. L., Friederici, A. D. & Fiebach, C. J. Combined eye tracking and

353   fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex J.*

354   *Devoted Study Nerv. Syst. Behav.* **68**, 33–47 (2015).

355   46. Yassa, M. A. & Stark, C. E. L. Pattern separation in the hippocampus. *Trends Neurosci.* **34**,
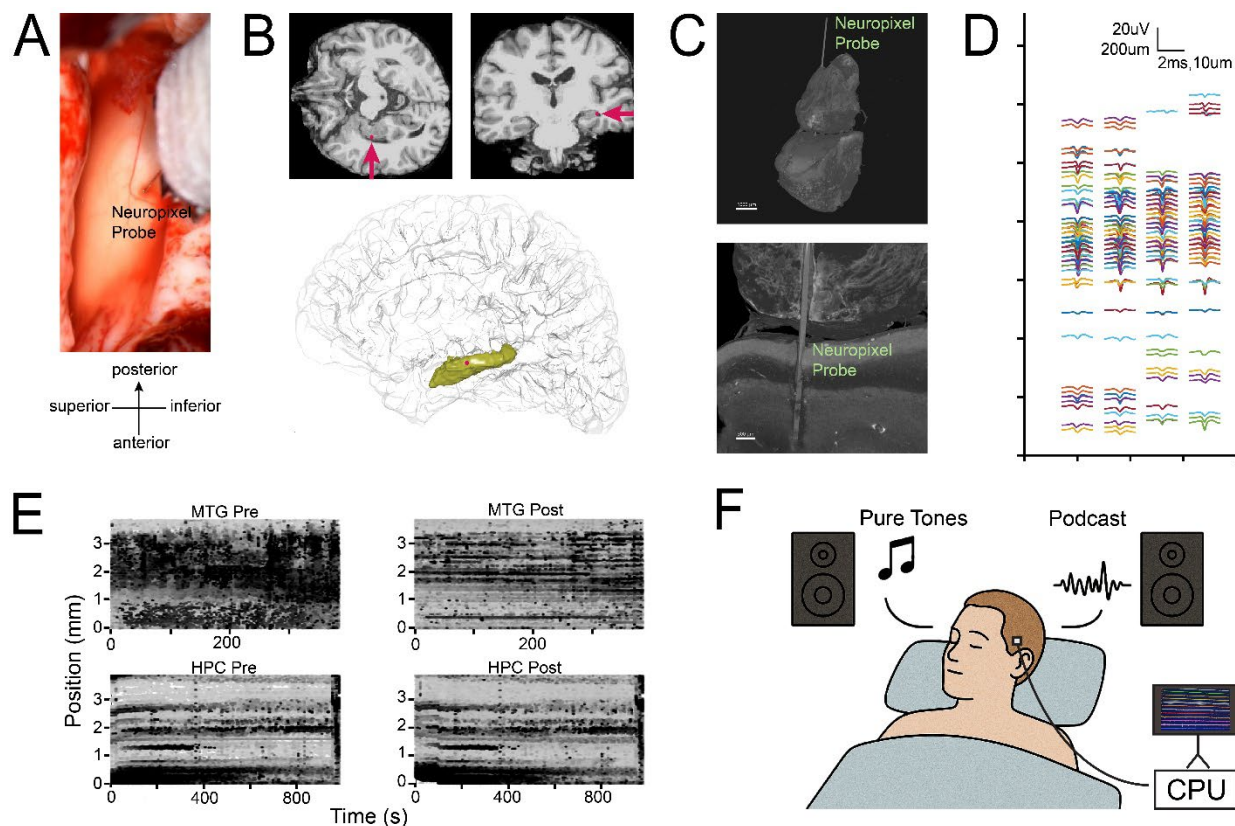
356   515–525 (2011).

357

358

Figure 1. Intraoperative Neuropixels recordings in the human hippocampus.
A. Photograph of the hippocampal brain tissue with the inserted Neuropixels probe during intraoperative recording (middle right), with the anatomical orientation indicated below. **B.** Top: axial (left) and coronal (right) sections of a T1 MRI for P8. Crimson dot indicates probe entry site, and arrows demonstrate trajectory of probe. Bottom: Probe entry site for P8 warped onto canonical brain, illustrated with a crimson dot within the hippocampus shown in yellow. **C.** 3D reconstruction of microCT identifying the probe within resected hippocampal tissue (top) with coronal slice identifying the probe penetrating the hippocampus (bottom). Superior globule is fibrin glue adhering to the ependymal lining. **D.** Example waveforms from all units (n=127) within a single hippocampal recording (P5). Each unit is represented by the average waveforms at the three maximal electrodes. **E.** Spiking activity wherein points represent the amplitude and location of individual spikes along the probe as a function of time and depth pre (left) and post (right) motion correction in a cortical (top) and hippocampal (bottom) recording. 0 indicates the bottom of the probe and the most lateral contact. MTG: Middle Temporal Gyrus, HPC: Hippocampus **F.** Task schematic. Patients listen to either pure tones (P4, P5, P6) or podcasts (P6, P8) during high density neural recordings.
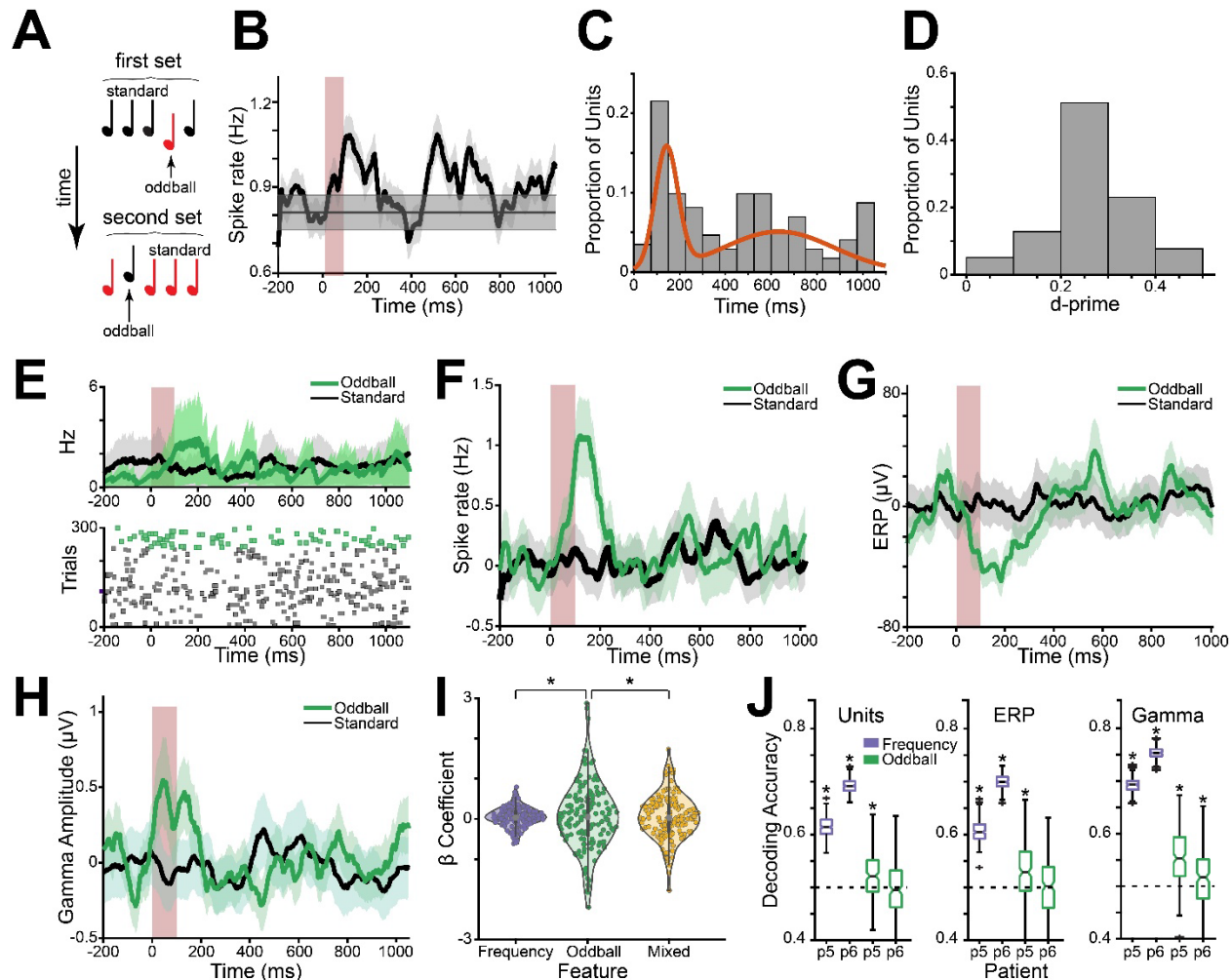.

Figure 2. Oddball responses in the anesthetized human hippocampus.
A. Schematic of the auditory oddball task (n=3 patients). Each trial consisted of pure tones with predetermined tone frequencies played at different probabilities. For two of the patients, the oddball and standard tone identities were interchanged halfway through i.e. the tone that was oddball for the first set (top) is the standard for the second set (below), and vice versa. **B.** Mean response to tone onset, averaged across all units (n=172 units, 3 patients). Vertical red bar indicates tone presentation (100ms), horizontal grey bar indicates baseline +/- standard error of the mean (SEM) and the shading represents +/- SEM. **C.** Distribution of the tone response onset latencies across all units. A mixed Gaussian model fit to the distribution is shown in orange, with peaks at 142.5 ms and 633.2 ms, trough at 291 ms. **D.** Distribution of d-prime values for units selective for tone frequency. **E.** Example unit that is selective for oddball trials. Top: Average response (spike rate, Hz) to oddball in green and standard in black, shading represents SEM, red bar is the tone presentation. Bottom: trial-wise spike raster plot, color-coded as oddball trials (green) and standard trials (black). **F.** Average neuronal response across all units (n=150 units, P5 and P6 to oddball (green) vs. standard (black) tones. Shading represents +/- SEM, red bar is tone presentation. **G.** Average ERP (µV) across 10 channels from each patient. Shading represents +/- SEM, red bar is tone presentation. **H.** Average gamma amplitude (µV) across 10 channels from each patient. Shading represents +/- SEM, red bar is tone presentation. **I.** Violin plot showing the distribution of β coefficients obtained from a linear regression model run per

397   unit, to determine neuronal response modulation as a function of tone frequency (tone frequency
398   β, purple, left) oddball identity (oddball β, green, middle), and an interaction/mixed term (mixed
399   β, yellow, right). Asterisks denote statistical significance of difference in absolute amplitude. **J.**
400   Box and whisker plot of encoding accuracy for tone frequency (purple) and oddball identity
401   (green), obtained using a Support Vector Machine (SVM) decoder for P5 and P6, across Units
402   (left), ERP (middle), and Gamma power (right) after tone presentation. Pluses are outliers and
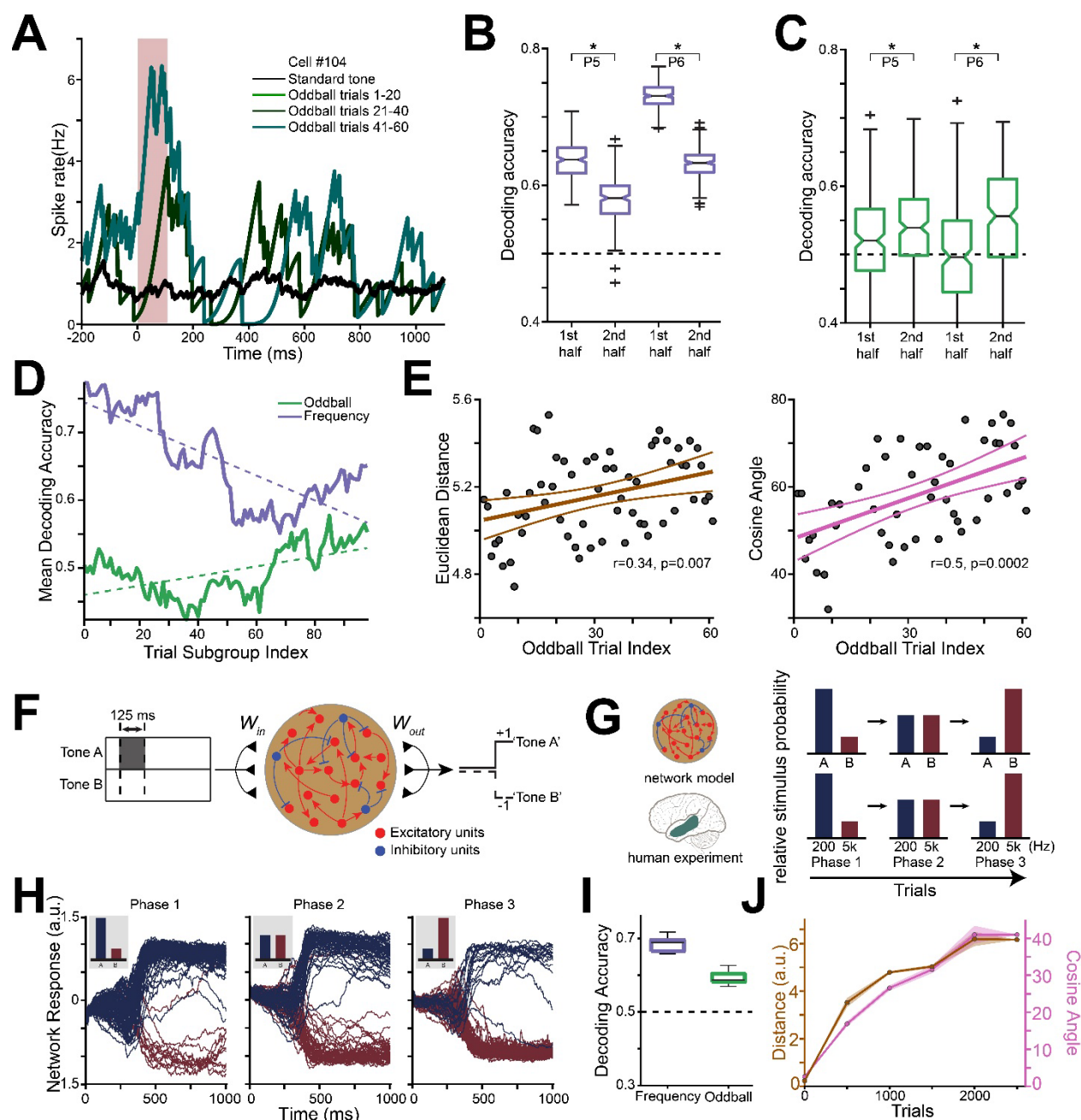403   asterisks denote statistical significance relative to change.

Figure 3: Evolution of the oddball representation across the neuronal population in experimental data and an RNN model.
A. Responses to tones as a function of oddball identity and index in an example unit. Red bar indicates tone presentation. **B.** Accuracy of tone frequency identity decoding across the neuronal population for patients P5 and P6, for the first half trials (left) or second half trials (right), combined across both blocks. Statistically significant differences indicated with an asterisk **C.** Similar to **B** but for oddball identity. **D.** Decoding accuracy as a function of trial position across both patients (n=43 oddball-responsive units). Each point represents SVM accuracy within a set of 50 trials starting at the index location. Decoding accuracy for tone frequency shown in purple, and for oddball identity shown in green, along with line fits shown as dashed lines in purple and green respectively. **E.** Euclidean distance (left) and Cosine angle (right) between standard and

416   oddball neuronal population response vectors, computed for each oddball trial. Each datapoint
417   (in grey) indicates the value of the Euclidean distance (left) and cosine angle (right) per trial,
418   with the lines showing a linear fit with 95% confidence intervals **F.** Schematic of the RNN
419   model trained to differentiate between two different tone frequencies, indicated as Tone A and
420   Tone B. **G.** Training paradigm for the RNN as compared to the human experiment. **H.** Network
421   response across trials for a single example RNN model unit for the three training phases. **I.**
422   Decoding accuracy of RNN for tone frequency (purple, left) and oddball identity (green, right).
423   **J.** Evolution of Euclidean distance (brown) and cosine angle (pink) between oddball trials and
424   the average standard trial across the RNN population. Shading represents SEM across 10 runs.
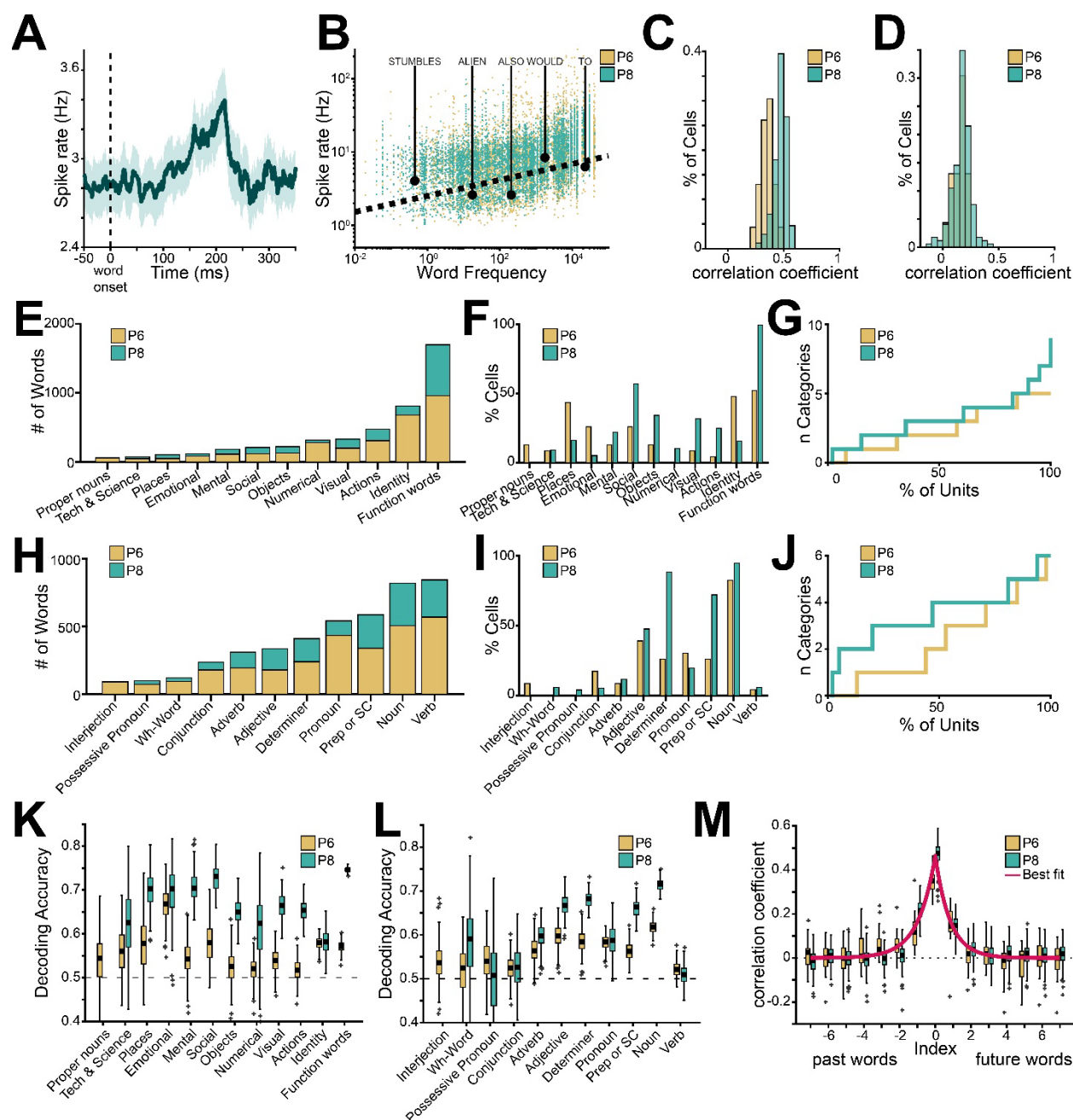425

Figure 4. Neuronal responses to natural language in the anesthetized human hippocampus. **A.** Average neuronal response (spike rate, Hz) of an example unit across all words presented (n=3019 words), shown aligned to word onset, indicated with a vertical dashed line. Shaded area represents +/- SEM. Activity more than 50 ms past word offset was removed to avoid contamination from adjacent words. **B.** Neuronal responses (spike rate, Hz) as a function of word frequency (n=3019 words) shown for each neuron, each data point corresponds to an individual unit and word, color-coded according to patient (ochre, P6; turquoise, P8). Example words indicated with black dots, and the dashed line corresponds to a linear fit across patients. **C-D.** Distribution of Pearson's correlation coefficient for predicted vs. actual firing rates for all words (**C**) and unique words (**D**), data shown per patient. **E.** Distribution of words within semantic categories sorted by frequency found within the podcast episodes, shown per patient. **F.**

438   Percentage of units selective for each semantic category, compared to all other semantic
439   categories, shown per patient. **G.** Number of categories decoded by individual units, shown per
440   patient. **H-J.** Similar to E-G but for part of speech categories. **K-L.** Box plots for decoding
441   accuracy of a SVM across units for semantic (**K**) and part of speech (**L**) categories, shown per
442   patient. Dashed lines represent chance and pluses indicate outliers. **M.** Correlation coefficients
443   for predicted vs. actual firing rates as a function of word index, index=0, current word, positive
444   indices correspond to future words (right) and negative indices to past words (left), shown per
445   patient.
446

Methods

Patient recruitment

Experiments were conducted according to protocol guidelines approved by the Institutional Review Board for Baylor College of Medicine and Affiliated Hospitals, Houston TX (H-50885). All recruited patients were diagnosed with drug resistant temporal lobe epilepsy and were scheduled to undergo an anteromesial temporal lobectomy for seizure control. All patients provided written informed consent to participate in the study and were aware that participation was voluntary and would not affect their clinical course. Included patients' age ranged from 31-54 years old (average 43.6 +/- 8.4), with three females and two male patients. Two resections were on the left side, and three were on the right. None of the patients reported explicit memory of intraoperative events after the case when discussed in the post-operative care unit or while recovering in the hospital the next day.

Neuropixels Data Acquisition Setup and Intraoperative Recordings

Neuropixels 1.0-S probes (IMEC) with 384 recording channels (total recording contacts = 960, usable recording contacts = 384) were used for recordings (dimensions: 70μm width, 100μm thickness, 10mm length). The Neuropixels probe, consisting of both the recording shank and the headstage, were individually sterilized with ethylene oxide (Bioseal, CA).[1] Our intraoperative data acquisition system included a custom-built rig including a PXI chassis affixed with an IMEC/Neuropixels PXIe Acquisition module (PXIe-1071) and National Instruments DAQ (PXI-6133) for acquiring neuronal signals and any other task-relevant analog/digital signals respectively. Our recording rig was certified by the Biomedical Engineering at Baylor St. Luke's Medical Center, where the intraoperative recording experiments were conducted. A high-

469   performance computer (10-core processor) was used for neural data acquisition using open-source

470   software such as SpikeGLX 3.0 and OpenEphys version 0.6x for data acquisition (AP band

471   (spiking data), band-pass filtered from 0.3kHz to 10kHz was acquired at 30kHz sampling rate;

472   LFP band, band-pass filtered from 0.5Hz to 500Hz, was acquired at 2500Hz sampling rate). We

473   used a "short-map" probe channel configuration for recording, selecting the 384 contacts located

474   along the bottom 1/3 of the recording shank.

475   Audio was played via a separate computer using pre-generated wav files and captured at 30kHz or

476   1,000kHz on the NIDAQ via a coaxial cable splitter that sent the same signal to speakers adjacent

477   to the patient. MATLAB (MathWorks, Inc.; Natick, MA) in conjunction with a LabJack (LabJack

478   U6; Lakewood, CO) was used to generate a continuous TTL pulse whose width was modulated by

479   the current timestamp and recorded on both the neural and audio datafiles. Online synchronization

480   of the AP and LFP files was performed by the OpenEphys recording software. Offline

481   synchronization of the neural and audio data was performed by calculating a scale and offset factor

482   via a linear regression between the time stamps of the reconstructed TTL pulses and confirmed

483   with visual inspection of the aligned traces.

484   Acute intraoperative recordings were conducted in brain tissue designated for resection based on

485   purely clinical considerations. The probe was positioned using a ROSA ONE Brain (Zimmer

486   Biomet) robotic arm and lowered into the brain 5-6mm from the ependymal surface using an

487   AlphaOmega microdrive. The penetration was monitored via online visualization of the neuronal

488   data and through direct visualization with the operating microscope (Kinevo 900). Reference and

489   ground signals on the Neuropixels probe were acquired separately by connecting to a sterile

490   microneedle placed in the scalp (separate needles inserted at distinct scalp locations for ground

491   and reference respectively).

492   For all patients (n=5), we conducted neuronal recordings under general anesthesia for at most 30

493   minutes as per the experimental protocol. All patients were under total intravenous anesthesia

494   (TIVA), with propofol as the main anesthetic per experimental protocol. Inhaled anesthetics were

495   only used for induction and stopped at least an hour prior to recordings. The anesthesiologist

496   titrated the anesthetic drug infusion rates so that the BIS monitor (Medtronic; Minneapolis, MN)

497   value was between 45 and 60 for the duration of the surgical case.[2] Of note, BIS values range

498   between 0 (completely comatose) and 100 (fully awake), with standard intraoperative values to be

499   between 40 and 60. In the first patient (P3), we first conducted neural recordings in the temporal

500   cortex (middle temporal gyrus). We then carried out hippocampal recordings in the same patient

501   after resection of the lateral temporal lobe but prior to any resection of the hippocampus. For the

502   remaining patients (P4, P5, P6, and P8), only hippocampal recordings were performed.

503   For P4, P5, and P6, we recorded neuronal activity during passive auditory stimuli presentation.

504   For P4, a sequence of auditory stimuli (pure tones; f1=1kHz, f2=3kHz) were presented with 80-

505   20 probability distribution, with the less frequent auditory stimulus serving as an "auditory oddball

506   stimulus" (n=300 trials). For P5 and P6, a sequence of auditory stimuli (pure tones; f1=200Hz,

507   f2=5kHz) were presented with 80-20 probability distribution, while counterbalancing the tone

508   frequency designated as the auditory oddball stimulus (first half, n=150 trials, f2 is auditory

509   oddball; second half, n=150 trials, f1 is auditory oddball). We interleaved a washout period (30

510   trials) using the same auditory stimuli but presented at 50-50 probability distribution in between

511   the two counterbalanced tasks. The auditory pure tone stimuli were presented for a 100 ms

512   duration, and the intertrial interval for the auditory oddball task was randomly drawn from between

513   1-3s. The different frequency waveforms were amplitude-matched.

514   For P6 and P8 we also recorded neuronal activity during podcast episodes. P6 listened to three

515   stories, each approximately 7 minutes long, taken from The Moth Radio Hour

516   (https://themoth.org/podcast). The stories were "Wild Women and Dancing Queens", "My

517   Father's Hands" and "Juggling and Jesus". Each episode consists of a single speaker narrating an

518   autobiographical story. P8 listened to "Why We Should NOT Look for Aliens - The Dark

519   Forest", an educational video created by the Kurzgesagt group (Kurzgesagt GmbH; Munich,

520   Germany) (https://www.youtube.com/watch?v=xAUJYP8tnRE). The selected stories were

521   chosen to be varied, engaging, and linguistically rich.

522   **Micro CT**

523   Since recordings were only performed in tissue planned for resection, we first removed a small

524   cube of tissue around the probe and then proceeded with the remainder of the resection. The cube

525   specimens were processed following previously described methods.[3] In brief, resected specimens

526   were fixed in 4% PFA for 16 hours at 4°C. They were then stabilized using a modified Stability

527   buffer (mStability), containing 4% acrylamide (BIO-RAD, cat. no. 1610140), 0.25% w/v VA044

528   (Wako Chemical, cat. no. 017-19362), 0.05% w/v saponin (MilliporeSigma, cat. no. 84510), and

529   0.1% sodium azide (MilliporeSigma, cat. no. S2002). Samples were equilibrated in the hydrogel

530   solution for 16 hours at 4°C before undergoing thermo-induced crosslinking at -90kPa and 37°C

531   for 3 hours. Following crosslinking, excess hydrogel solution was removed, and specimens were

532   washed four times with 1X PBS. Next, samples were immersed in 0.1N iodine and incubated with

533   gentle agitation for 24 hours at room temperature before being embedded in agarose and imaged

534   using a Zeiss Xradia Context micro-CT at 3μm/voxel resolution. The acquired back-projection

535   images were reconstructed using Scout-and-Scan Reconstructor (Carl Zeiss, Ver. 16.8) and

536   converted to NRRD format via Harwell Automated Recon Processor (HARP, Ver. 2.4.1),[4] an

537   open-source, cross-platform application developed in Python. The 3D volumes were analyzed, and

538   optical sections were captured using 3D Slicer.[5]

539   **Neuronal Data Processing**

540   *Motion-correction*

541   We utilized previously developed and validated motion estimation and interpolation algorithms

542   to correct for the motion artifacts from brain movement.[6] Motion was estimated via the DREDge

543   software package (Decentralized Registration of Electrophysiology Data software,

544   https://github.com/evarol/DREDge) using either a combination of motion traces obtained using

545   raw LFP and/or AP band data, fine-tuned for individual recordings. Motion-correction was then

546   implemented using interpolation methods

547   (https://github.com/williamunoz/InterpolationAfterDREDge). Both the AP and LFP band data

548   are motion-corrected and utilized for further pre-processing and analysis steps. If the estimated

549   motion led to no improvement in the spike locations then spike sorting proceeded with the

550   motion correction package built into Kilosort 4 without performing interpolation.

551   *Unit extraction and classification*

552   Automated spike detection and clustering were performed by Kilosort 2.0 if motion correction

553   was already applied using the DREDge algorithm or KiloSort 4.0[7] if motion correction was not

554   applied separately. Manually curation of spike clustered was performed using the open-source

555   software Phy.[8] Unit quality metrics were calculated using SpikeInterface[9] and were considered

556   single units if they had a d-prime (d') greater than 1 and fewer than 3% of spikes were violations

557   of a 2ms inter-spike interval refractory period.

558   *Local Field Potential data*

559   LFP data was bandpass-filtered between 0.1-20Hz and aligned to task events to extract local ERPs.

560   Gamma band amplitude was calculated in the "high gamma" range, first bandpass filtering it

561   between 70-150Hz and then calculating the absolute value of the Hilbert-transformed complex

562   signal. Given the high correlation between adjacent channels, only 10 channels equally spanning

563   the length of the probe were used to calculate statistics.

564

565   **Neuronal Data Analysis:**

566   All analyses were performed using custom MATLAB code.

567   *Motion Analysis*

568   The motion-corrected location estimates were obtained at a 250Hz sampling frequency using the

569   DREDge algorithm. This signal was downsampled to 10Hz. The power spectrum of the calculated

570   motion was then estimated using Welch's overlapped segment averaging estimator for frequencies

571   between 0.1 and 3Hz. The amount of motion was defined as the root mean square error of the

572   location trace of the probes center relative to its average location.

573   *Tone Responses*

574   Both single units and multi-units were used for all analyses. A tone responsive neuron was defined

575   as having a statistically significant increase in the average firing rate in the first second after tone

576   onset (shifted by 50ms to account for auditory delay) relative to the preceding 200ms baseline

577   ($\propto$<0.05, Wilcoxon signed-rank test). Visual demonstrations of the peri-stimulus average firing

578   rate were smoothed via a causal Gaussian filter with a standard deviation of 150ms for

579   visualization, however, all statistical analyses were performed on the raw spike count. Response

580   onset latency was computed as the time taken to the peak response. A Mixed Gaussian Model with

581   two components was then fit to the distribution of latencies. Given the trough between the two

582   peaks at 291ms and evidence of average oddball response occurring in the first segment, a window

583   of 0-300ms was used for analysis characterizing tone and oddball selectivity.

584   *Neural Tuning*

585   To determine response tuning properties, we modeled trial responses in the peristimulus period

586   using general linear regression models. Neural data in the analysis time window of 0-300ms was

587   used for tuning analyses. Unit response was defined as the average firing rate, LFP power was

588   defined as the root mean square (RMS) value of the bandpass-filtered LFP, and gamma power was

589   defined as the average gamma band amplitude. All vectors were z-scored to allow for comparison

590   of the neural response modulation across units/channels. The independent variables were effects-

591   coded for tone type (frequency 1 vs. frequency 2), trial type (standard vs. oddball), and an

592   interaction term (conjunctive coding). We set the α level at 0.05 to determine if the beta coefficient

593   for the independent variables were statistically significant.

594   *Neuronal Population Coding*

595   To determine the information content present in the population, a Support Vector Machine with a

596   linear kernel was trained using 10-fold cross validation for 200 iterations. Accuracy for each

597   iteration was defined as the average accuracy across the 10 folds. Significant coding was defined

598   as the distribution of 200 iterations being statistically different from 0.5 (chance). Algorithm

599   validation was performed by shuffling the dataset and demonstrating that it always performed at

600   chance level. Subsampling was performed to avoid performance bias from an unbalanced dataset

601   (i.e. more standard trials than oddball trials). To investigate the neuronal population response

602   dynamics for tone and oddball encoding as a function of time, we used sets of sequential trials (50

603   trials) from each of the two counterbalanced blocks (total of 100 trials). For example, the first set

604   was using trials 1:50 and 181:230, whereas the last set was using trials 101:150 and 281:330.

605   Decoding analyses were also run separately for early vs. late trials (first 75 vs. last 75 trials within

606   a 150-trial block) for tone and oddball encoding respectively.

607   *Neuronal response learning dynamics*

608   Next, to determine the neural mechanism underlying statistical learning required for oddball

609   detection, we evaluated single-trial response dynamics across the neuronal population. For each

610   trial, we generated a neuronal response population vector. We then computed the Euclidean

611   distance ($||u - v||$) and cosine angle ($invcos(u \cdot v/||u|| * ||v||)$) between the mean vector across

612   all standard trials and each individual oddball unit vector, evaluating each as a function of the

613   oddball index.

614   **Continuous-rate RNN model.** We implemented a continuous-rate recurrent neural network

615   (RNN) and trained it to perform an oddball detection task, closely mirroring the one used for the

616   experimental dataset. The network contains 200 recurrently connected units (80% of which are

617   excitatory and 20% of which are inhibitory units). The network is governed by the following

618   equation:

619
$$\tau_i \frac{dx_i}{dt} = -x_i(t) + W \cdot r(t) + u(t)$$

620
$$r_i(t) = \frac{1}{1 + e^{-x_i(t)}}$$

621
$$o(t) = W_{out} \cdot r(t)$$

622 where i represents the synaptic decay time constant for unit i, xi(t) indicates the synaptic current

623 variable for neuron i at time point t, W is the recurrent connectivity matrix (N-by-N; i.e. 200-by-

624 200), and u(t) is the input data given to the network at time point t. u is a 2-by-200 matrix where

625 the first dimension refers to the number of input channels and the second dimension is the total

626 number of time points. A firing rate of a unit was estimated by passing the synaptic current variable

627 (x) through a standard logistic sigmoid function. The output (o) of the network was computed as a

628 linear weighted sum of the entire population of units.

629 In each trial, the network model receives input data mimicking auditory signals. The input consists

630 of two signal streams, each representing a distinct auditory tone (i.e. tone A vs. tone B; [**Figure**

631 **3F,G**]). Only one tone is presented to the network per trial. The model was trained to produce an

632 output signal approaching +1 when Tone A was presented and an output signal approaching -1

633 when Tone B was presented. To closely replicate the experimental task design, we employed three

634 different sequential contexts during network training. In the first stage, Tone A was presented

635 predominantly (80% of trials), followed by an equal distribution of Tone A and Tone B (50/50) in

636 the second stage. In the third stage, Tone B was predominant (80%).

637 We optimized the network parameters, including recurrent connectivity, readout weights, and

638 synaptic decay time constants, using gradient descent via backpropagation through time (BPTT).

639 The network was required to achieve over 95% task accuracy in the current context before a new

640 context was introduced.

641 **Neuronal Data Analysis: Natural Language stimuli**

642   *Natural language stimuli*

643   All patients were native English speakers. The podcast played during the task was automatically

644   transcribed using Assembly AI (https://www.assemblyai.com/). The transcribed words and

645   corresponding timestamp outputs from Assembly AI were converted to a TextGrid and then loaded

646   into Praat.[10] The original wav file was also loaded into Praat and the spectrograms and labels and

647   timestamps were manually checked and corrected to ensure the word onset and offset times were

648   accurate. This process was then repeated by a second reviewer to ensure the validity of the time

649   stamps. The TextGrid output of corrected words and timestamps from Praat was converted to a xls

650   and loaded into Matlab and Python for further analysis.

651   *Natural Language statistics*

652   Word frequency was defined based on a corpus of movie subtitles spanning a total of 51 million

653   words.[11] Words that did not elicit a response during the duration of the word were excluded from

654   this analysis. To compare the relative contributions to firing rate, a linear model was trained to

655   estimate the logarithmic firing rate from the logarithmic duration and corpus frequency of each

656   word. Word surprisal values were calculated using the GPT-2 large model[12] from the Hugging

657   Face Transformers library,[13] computing the negative log probability of each word conditioned on

658   the preceding context. Specifically, surprisal was defined by the equation: $surprisal =$

659    $-log\ P(w_i|w_{i-1}, w_{i-2}, \ldots, w_2, w_1)$ where $P(w_i)$ refers to the probability of word i given the

660   proceding words.

661   We utilized the pre-trained fastText Word2Vec model in MATLAB to extract word embeddings

662   for all words in our dataset.[14,15] This pre-trained model provides 300-dimensional word

663   embedding vectors, trained on 16 billion tokens from Wikipedia, UMBC webbase corpus, and

664 statmt.org, to capture semantic relationships between words. Notably, Word2Vec is a non

665 contextual embedder, so all instances of the same word are represented the same. Some surname

666 words, such as "Harwood" or proper nouns like "Applebee's" did not have word embeddings

667 and were discarded from the analysis. A simple linear model was trained to predict the firing rate

668 of individual neurons from the semantic matrices using 10-fold cross-validation. Accuracy was

669 defined as the correlation between true and predicted firing rates. Words with 0Hz or above

670 25Hz were removed from this analysis. To prevent overfitting, Principal Component Analysis

671 (PCA) was used to reduce the dimensionality to account for 30% of the variance prior to

672 modeling. This threshold was defined as the minimum of the RMSE of the model that balanced

673 under and overfitting. To predict future or previous words the alignment between words was

674 shifted forwards or backwards, respectively. This relation was then fit with a piecewise

675 exponential decay

676

677
$$r(i) = \beta_0 * \begin{cases} e^{1/\beta_1} \text{ for } i >= 0 \\ e^{-i/\beta_2} \text{ for } i < 0 \end{cases}$$

678 Wherein β_0 is the amplitude of the correlation at 0 lag, and β_1 and β_2 are equivalent to the time

679 constant of the decay for positive and negative lags, respectively.

680 *Word embedding, Semantic clustering, and Part of Speech classification*

681 To identify the natural semantic categories present in our word data, all unique words heard by the

682 participants were clustered into groups using a word embedding approach.[14–16] We used the same

683 300-dimensional embedding from the prior GLM analysis. To compute and visualize semantic

684 clusters, we first used a t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm on word

685   embedding values to reduce the dimensionality of each unique word based on their cosine distance

686   to all other words, thus reflecting their semantic similarity. Words with similar meanings now have

687   similar 2D coordinates. We then applied the k-means clustering algorithm to these 2D word

688   representations and visualized clustered words on a 2D word map (12 clusters).[17] We then

689   manually inspected and assigned a distinct label to each semantic cluster and adjusted clusters for

690   accuracy. For example, words bordering the edges of clusters would sometimes get mis-grouped

691   and were manually corrected. The final 12 semantic categories of the words are body parts, places,

692   emotional words, mental words, social words, objects, visual words, numerical words, actions,

693   identity words, function words, and proper nouns. Correction for multiple comparisons was

694   performed using the Benjamini Hochberg approach.[18] A SVM was trained for each semantic

695   category (versus all other categories) using a radial basis function ('RBF') kernel. Model training

696   and accuracy metrics were weighted to the relative frequency of each group. We used 10-fold cross

697   validation and 200 iterations.

698   To extract part-of-speech (POS) for each word in the dataset, we utilized an automated pipeline

699   through Stanford CoreNLP, a natural language processing toolkit.[19] We initialized a

700   CoreNLPParser with the 'pos' tagtype, which specializes in part-of-speech tagging. The transcript

701   was first segmented into sentences based on punctuation. Each sentence was then tokenized and

702   passed through the CoreNLPParser's tagging function. This process leveraged CoreNLP's

703   advanced linguistic models to analyze the context and structure of each sentence, assigning

704   appropriate POS tags to individual words. The 15 POS types were: 'Noun', 'Adjective', 'Numeral',

705   'Determiner', 'Conjunction', 'Preposition or Subordinating Conjunction', 'Auxiliary', 'Possessive

706   Pronoun', 'Pronoun', 'Adverb', 'Particle', 'Interjection', 'Verb', 'Wh-Word', and 'Existential'. POS

707   types with fewer than 45 words were removed from analysis. A similar SVM was used for POS.

708   Bibliography

709   1. Coughlin, B. *et al.* Modified Neuropixels probes for recording human neurophysiology in the

710   operating room. *Nat Protoc* **18**, 2927–2953 (2023).

711   2. Singh, H. Bispectral index (BIS) monitoring during propofol-induced sedation and

712   anaesthesia. *Eur J Anaesthesiol* **16**, 31–36 (1999).

713   3. Hsu, C.-W. *et al.* High resolution imaging of mouse embryos and neonates with X-ray micro-

714   computed tomography. *Curr Protoc Mouse Biol* **9**, e63 (2019).

715   4. Brown, J. M. *et al.* A bioimage informatics platform for high-throughput embryo phenotyping.

716   *Brief Bioinform* **19**, 41–51 (2018).

717   5. Fedorov, A. *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging

718   Network. *Magn Reson Imaging* **30**, 1323–1341 (2012).

719   6. Windolf, C. *et al.* DREDge: robust motion correction for high-density extracellular recordings

720   across species. *Nat Methods* 1–13 (2025).

721   7. Pachitariu, M., Sridhar, S., Pennington, J. & Stringer, C. Spike sorting with Kilosort4. *Nat*

722   *Methods* **21**, 914–921 (2024).

723   8. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat Neurosci* **19**, 634–641

724   (2016).

725   9. Buccino, A. P. *et al.* SpikeInterface, a unified framework for spike sorting. *Elife* **9**, (2020).

726   10. Boersma, P. Praat: doing phonetics by computer [Computer program]. *Httpwww Praat Org*

727   (2011).

728   11. Brysbaert, M. & New, B. Moving beyond Kučera and Francis: A critical evaluation of current

729   word frequency norms and the introduction of a new and improved word frequency measure

730   for American English. *Behav. Res. Methods* **41**, 977–990 (2009).

731   12. Radford, A. *et al.* Language Models are Unsupervised Multitask Learners. in (2019).

732   13. Wolf, T. *et al.* Transformers: State-of-the-Art Natural Language Processing. in *Proceedings*

733   *of the 2020 Conference on Empirical Methods in Natural Language Processing: System*

734   *Demonstrations* (eds. Liu, Q. & Schlangen, D.) 38–45 (Association for Computational

735   Linguistics, Online, 2020). doi:10.18653/v1/2020.emnlp-demos.6.

736 14. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in

737   vector space. *ArXiv CsCL* (2013).

738 15. Joulin, A., Grave, É., Bojanowski, P. & Mikolov, T. Bag of Tricks for Efficient Text

739   Classification. in *Proceedings of the 15th Conference of the European Chapter of the*

740   *Association for Computational Linguistics: Volume 2, Short Papers* 427–431 (2017).

741 16. Henry, S., Cuffy, C. & McInnes, B. T. Vector representations of multi-word terms for

742   semantic relatedness. *J Biomed Inf.* **77**, 111–119 (2018).

743 17. Franch, M. *et al.* A vectorial code for semantics in human hippocampus. *bioRxiv* 2025–02

744   (2025).

745 18. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and

746   Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300

747   (1995).

748 19. Manning, C. D. *et al.* The Stanford CoreNLP Natural Language Processing Toolkit. in

749   *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics:*

750   *System Demonstrations* 55–60 (2014).

751

752

Author Contributions

KAK, SS, JA, GPB, NRP, SSC, ACP, AW, ZW, BYH, and SAS contributed to the design of the work and the acquisition of data

KAK, SS, MCF, JLB, DM, EM, MM, WM, CWH, ACP, SRH, RK, NR, BYH, and SAS contributed to the analysis and interpretation of data

KAK drafted the manuscript and SS, BYH and SAS substantively revised it.

All authors approved the submitted version

Correspondence and requests for materials should be addressed to Sameer A. Sheth

sameer.sheth@bcm.edu

Reprints and permissions information is available at www.nature.com/reprints