Learning and language in the unconscious human hippocampus

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

* *these authors contributed equally*.

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

ABSTRACT

Consciousness is a fundamental component of cognition,[1] but the degree to which higher-order perception relies on it remains disputed.[2,3] Here we demonstrate the persistence of learning, semantic processing, and online prediction in individuals under general anesthesia-induced loss of consciousness.[4,5] Using high-density Neuropixels microelectrodes[6] to record neural activity in the human hippocampus while playing a series of tones to anesthetized patients, we found that hippocampal neurons could reliably detect oddball tones. This effect size grew over the course of the experiment (~10 minutes), consistent with learning effects. A biologically plausible recurrent neural network model showed that learning and oddball representation are an emergent property of flexible tone discrimination. Last, when we played language stimuli, single units and ensembles carried information about the semantic and grammatical features of natural speech, even predicting semantic information about upcoming words. Together these results indicate that in the hippocampus, which is anatomically and functionally distant from primary sensory 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

presentation of auditory stimuli composed of pure tones (3 patients) or a continuous podcast (2 patients, **Figure 1F**).

Auditory environment monitoring in the anesthetized patient

The ability to recognize patterns and detect violations of those patterns is a hallmark of cognition. In the classic oddball task,[11,12] participants are presented with a series of stimuli that constitute a pattern (e.g., a series of tones of the same frequency) interspersed with deviant "oddballs". In three of the hippocampal recordings (P4, P5, and P6), we played a series of 100 ms pure tones; 20% were oddballs (higher or lower frequency than standard, **Figure 2A, Methods**). Most units (n=122/172, 70.9%, signed-rank test, α=0.05) showed tone-evoked responses (**Figure 2B**), consistent with established auditory responses within hippocampus.[13] Neural responses to tones often showed a biphasic temporal firing rate profile (**Figure 2B**). Across all units, response latencies showed a clear bimodal temporal dynamic (Gaussian Mixture Model fit via Expectation Maximization, **Figure 2C**). Hippocampal units encoded tone frequency (n=39/172, 22.7% of units, rank sum test, α=0.05, **Figure 2D**).

Having established auditory responses despite the anesthetic state, we next examined the representation of stimulus features. For two patients, we balanced tone frequency and oddball status (**Figure 2A,** n=150 units). At the single unit (**Figure 2E**) and population (**Figure 2F**) levels, neuronal responses differentiated standards from oddballs. This divergence was most notable within the first 300 ms, with 24.7% (n=37/150) of units signaling oddballs. Thus, further analyses focused on this first time segment. Local field potentials (LFPs) also showed oddball-evoked responses, observed as a negative deflection in the evoked response potential (ERP, **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
104 **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

magnitude over the course of the experiment (~10 minutes, example unit, **Figure 3A)**. Splitting our task into halves, we found a significant increase in oddball encoding for both patients (P5: $p=0.01$, P6: $p<0.001$, t-test, **Figure 3C**). Surprisingly, we also observed a concomitant decrease in frequency encoding, raising the possibility of compensatory mechanisms ($p<0.001$, t-test for both) (**Figure 3B**). Using a sliding window of subsets of 50 trials, we found a continuous increase in oddball decoding accuracy across the approximately 10-minute duration of the experiment ($p<0.001$, Pearson's correlation, **Figure 3D, purple**). Again, this increase in oddball performance was accompanied by an initial decrease in tone encoding ($p<0.0001$, **Figure 3D, green**), [15] demonstrating the neural population was sacrificing its tone responses for the sake of oddball representations over the course of the experiment, suggesting that the hippocampal responses were shifting to represent the salient features of the stimulus.[16]

We created neural vectors of the average standard tone response as well as each individual oddball trial (43-dimensional vectors composed of the mean response of the oddball units). We found a gradual divergence in Euclidean distance between standard and oddball vectors over the course of the session ($r=0.34$, $p=0.007$, Pearson's correlation; **Figure 3E, left**). Discriminability was even stronger when considering cosine angle, indicating the effect is not merely a consequence of a response gain in oddball cells ($r=0.5$, $p=0.0002$; **Figure 3E, right**). These effects were mostly consistent for individual patients (P5 distance: $r=0.25$; $p=0.056$, angle: $r=0.43$ $p=0.002$; P6 distance: $r=0.32$, $p=0.012$, angle: $r=0.48$; $p=0.0002$). These results indicate that the hippocampus does not simply improve encoding using gain modulation;[17] instead, oddball responses reflect a rotation of the neural population vector in a high dimensional space, meaning that oddball learning alters the warping of the neural response manifold.[18] Thus, complex reshaping of responses can occur even under general anesthesia.[19,20]

To gain further mechanistic insight at the level of individual units, we turned to a continuous-rate recurrent neural network (RNN) trained to perform a signal-detection task similar to the task used for the human Neuropixels data (**Figure 3F**).[21] The network model underwent three stages of training, simulating the different contexts used in the experimental data, with the prevalence of specific tones varied at each stage (**Figure 3G, H, Methods**). Tone A was presented to the network in 80% of the trials in the first stage, followed by a washout period, and then a third stage with probabilities reversed relative to the first. By the end of training (range of 1400 to 2600 trials) the model was able to reliably differentiate tone identities (**Figure 3H**). Notably, despite being only explicitly trained on tone frequency discrimination, the model was able to perform not only frequency discrimination (tone frequency, $p<0.005$ signed Wilcoxon test vs. shuffled data) but also context (oddball identity, $p<0.005$, signed Wilcoxon test vs. shuffled data, **Figure 3I**). The model also recreated the pattern observed in the Euclidean and vector angle distance between standard and oddball representations (**Figure 3J)**, suggesting that the divergence of representations can be due to local computations rather than inherited from other networks.

Unconscious encoding of semantics and grammar in the hippocampus

We next tested whether the unconscious hippocampus could perform even higher order functions associated with parsing semantic and syntactic features of natural speech. In two participants (P6 and P8), we recorded neural activity while playing 10-20 minutes of podcast episodes (see **Methods**[22]). We aligned neural activity to word onset and offset (n=3024 words for P6 and n=1565 words for P8), and computed word-evoked neural responses (example unit, 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, $\alpha<0.05$). 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 $\alpha<0.05$), 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 ($\alpha=0.05$, 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

using only unique words (n=746 and 573 unique words for P5 and P6, respectively). We found a significant result in 84.1% of the recorded units (159/189 units with at least 50 words that had a non-zero response), with an average correlation of r=0.17 +/-0.08, **Figure 4D**). In other words, it is possible to predict the firing rate of units to a given word based on responses to other words by leveraging their similarities in semantic space,[27] demonstrating that the unconscious hippocampus has access to abstract semantic relationships between words.

We then analyzed the representation of word features. We semantically categorized each word into one of 12 possible groups (**Figure 4E)**.[22] Nearly all units (193/195) showed some form of semantic category selectivity (α=0.05, Kruskal Wallis test for any difference between semantic categories). Rank-sum tests for each category versus all others showed that units were selective for multiple semantic categories, consistent with our previously reported findings in awake patients (corrected for multiple comparisons, $\alpha<0.05$).[22] Specifically, 165/195 (84.6%) units were able to discriminate between at least two of the twelve semantic categories and 76/195 (39.0%) were able to discriminate across at least four (**Figure 4G**), with a median of three categories per neuron. We also investigated encoding of grammatical features. We classified each word into a part of speech using the Stanford CoreNLP toolkit[28] (**Figure 4H**, n=2906 words for P5 and 1497 for P8). We found that 191/195 units carried information about part of speech (α=0.05, Kruskal Wallis test). Again, there was broad representation of different categories (**Figure 4I**). Interestingly, nearly all units (P6: 82.6%; P8: 94.8%) distinguished nouns from non-nouns, but only a few (P6: 4.3%; P8: 5.8%) distinguished verbs from non-verbs, consistent with the greater role of the hippocampus in object over action representations.[29] Overall, the median number of categories represented was four (out of 11 possible, **Figure 4J**), with 178/195 (91.3%) 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 $\tau_{past}=0.97$ versus
223 future words at $\tau_{future}=0.81$. 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

provide further evidence for semantic and grammar representations within the hippocampus, we even demonstrate continuous prediction of upcoming words. Our results therefore extend these language models of hippocampal computations by showing they are sufficiently robust that they do not even require conscious awareness.

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Supplementary Information is available for this paper.

1. van Gaal, S., De Lange, F. P. & Cohen, M. X. The role of consciousness in cognitive control and decision making. *Front. Hum. Neurosci.* **6**, (2012).
2. Dehaene, S., Lau, H. & Kouider, S. What is consciousness, and could machines have it? *Science* **358**, 486–492 (2017).
3. Sikkens, T., Bosman, C. A. & Olcese, U. The Role of Top-Down Modulation in Shaping Sensory Processing Across Brain States: Implications for Consciousness. *Front. Syst. Neurosci.* **13**, (2019).
4. Alkire, M. T., Hudetz, A. G. & Tononi, G. Consciousness and Anesthesia. *Science* **322**, 876–880 (2008).
5. Brown, E. N., Lydic, R. & Schiff, N. D. General Anesthesia, Sleep, and Coma. *N. Engl. J. Med.* **363**, 2638–2650 (2010).
6. Coughlin, B. *et al.* Modified Neuropixels probes for recording human neurophysiology in the operating room. *Nat Protoc* **18**, 2927–2953 (2023).
7. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
8. DeKraker, J., Köhler, S. & Khan, A. R. Surface-based hippocampal subfield segmentation. *Trends Neurosci.* **44**, 856–863 (2021).
9. Buzsaki, G. *Rhythms of the Brain*. (Oxford University Press, New York, NY, 2006).
10. Chung, J. E. *et al.* High-density single-unit human cortical recordings using the Neuropixels probe. *Neuron* **110**, 2409-2421.e3 (2022).
11. Tzovara, A. *et al.* Predictable and unpredictable deviance detection in the human hippocampus and amygdala. *Cereb Cortex* **34**, bhad532 (2024).
12. García-Larrea, L., Lukaszewicz, A. C. & Mauguière, F. Revisiting the oddball paradigm. Non-target vs neutral stimuli and the evaluation of ERP attentional effects. *Neuropsychologia* **30**, 723–741 (1992).

13. Billig, A. J., Lad, M., Sedley, W. & Griffiths, T. D. The hearing hippocampus. *Prog. Neurobiol.* **218**, 102326 (2022).
14. Näätänen, R., Paavilainen, P. & Reinikainen, K. Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neurosci. Lett.* **107**, 347–352 (1989).
15. Kelemen, E. & Fenton, A. A. Dynamic Grouping of Hippocampal Neural Activity During Cognitive Control of Two Spatial Frames. *PLoS Biol.* **8**, e1000403 (2010).
16. Fontanini, A. & Katz, D. B. Behavioral States, Network States, and Sensory Response Variability. *J. Neurophysiol.* **100**, 1160–1168 (2008).
17. Treue, S. & Maunsell, J. H. Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J Neurosci* **19**, 7591–7602 (1999).
18. Ebitz, R. B. & Hayden, B. Y. The population doctrine in cognitive neuroscience. *Neuron* **109**, 3055–3068 (2021).
19. Mirabella, G. *et al.* Neurons in Area V4 of the Macaque Translate Attended Visual Features into Behaviorally Relevant Categories. *Neuron* **54**, 303–318 (2007).
20. David, S. V., Hayden, B. Y., Mazer, J. A. & Gallant, J. L. Attention to Stimulus Features Shifts Spectral Tuning of V4 Neurons during Natural Vision. *Neuron* **59**, 509–521 (2008).
21. Rungratsameetaweemana, N., Kim, R., Chotibut, T. & Sejnowski, T. J. Random noise promotes slow heterogeneous synaptic dynamics important for robust working memory computation. *Proc Natl Acad Sci U A* **122**, e2316745122 (2025).
22. Franch, M. *et al.* A vectorial code for semantics in human hippocampus. *bioRxiv* 2025–02 (2025).
23. Brysbaert, M. & New, B. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* **41**, 977–990 (2009).

24. Radford, A. *et al.* Language Models are Unsupervised Multitask Learners. in (2019).
25. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. *ArXiv CsCL* (2013).
26. Jamali, M. *et al.* Semantic encoding during language comprehension at single-cell resolution. *Nature* **631**, 610–616 (2024).
27. Khanna, A. R. *et al.* Single-neuronal elements of speech production in humans. *Nature* **626**, 603–610 (2024).
28. Manning, C. D. *et al.* The Stanford CoreNLP Natural Language Processing Toolkit. in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* 55–60 (2014).
29. Mestres-Missé, A., Rodriguez-Fornells, A. & Münte, T. F. Neural differences in the mapping of verb and noun concepts onto novel words. *Neuroimage* **49**, 2826–2835 (2010).
30. Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P. & de Lange, F. P. A hierarchy of linguistic predictions during natural language comprehension. *Proc Natl Acad Sci U A* **119**, e2201968119 (2022).
31. Jang, A. I., Wittig, J. H., Jr, Inati, S. K. & Zaghloul, K. A. Human cortical neurons in the anterior temporal lobe reinstate spiking activity during verbal memory retrieval. *Curr Biol* **27**, 1700-1705.e5 (2017).
32. Pickering, M. J. & Gambi, C. Predicting while comprehending language: A theory and review. *Psychol Bull* **144**, 1002–1044 (2018).
33. Ryskin, R. & Nieuwland, M. S. Prediction during language comprehension: what is next? *Trends Cogn. Sci.* **27**, 1032–1052 (2023).
34. Franks, N. P. & Lieb, W. R. Mechanisms of general anesthesia. *Env. Health Perspect* **87**, 199–205 (1990).
35. Strange, B. A. & Dolan, R. J. Adaptive anterior hippocampal responses to oddball stimuli. *Hippocampus* **11**, 690–698 (2001).

36. Krom, A. J. *et al.* Anesthesia-induced loss of consciousness disrupts auditory responses beyond primary cortex. *Proc Natl Acad Sci U A* **117**, 11770–11780 (2020).
37. Moon, D. U. *et al.* Propofol modulates early memory consolidation in humans. *eNeuro* **7**, ENEURO.0537-19.2020 (2020).
38. Iggena, D. *et al.* Post-encoding modulation of spatial memory consolidation by propofol. *Cortex* **156**, 1–12 (2022).
39. Kihlstrom, J. F., Schacter, D. L., Cork, R. C., Hurt, C. A. & Behr, S. E. Implicit and Explicit Memory following Surgical Anesthesia. *Psychol Sci* **1**, 303–306 (1990).
40. Linassi, F. *et al.* Implicit memory and anesthesia: A systematic review and meta-analysis. *Life Basel* **11**, 850 (2021).
41. Covington, N. V. & Duff, M. C. Expanding the Language Network: Direct Contributions from the Hippocampus. *Trends Cogn. Sci.* **20**, 869–870 (2016).
42. Dijksterhuis, D. E. *et al.* Pronouns reactivate conceptual representations in human hippocampal neurons. *Science* **385**, 1478–1484 (2024).
43. Duff, M. C. & Brown-Schmidt, S. The hippocampus and the flexible use and processing of language. *Front. Hum. Neurosci.* **6**, 69 (2012).
44. Olsen, R. K., Moses, S. N., Riggs, L. & Ryan, J. D. The hippocampus supports multiple cognitive processes through relational binding and comparison. *Front. Hum. Neurosci.* **6**, (2012).
45. Bonhage, C. E., Mueller, J. L., Friederici, A. D. & Fiebach, C. J. Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex J. Devoted Study Nerv. Syst. Behav.* **68**, 33–47 (2015).
46. Yassa, M. A. & Stark, C. E. L. Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–525 (2011).

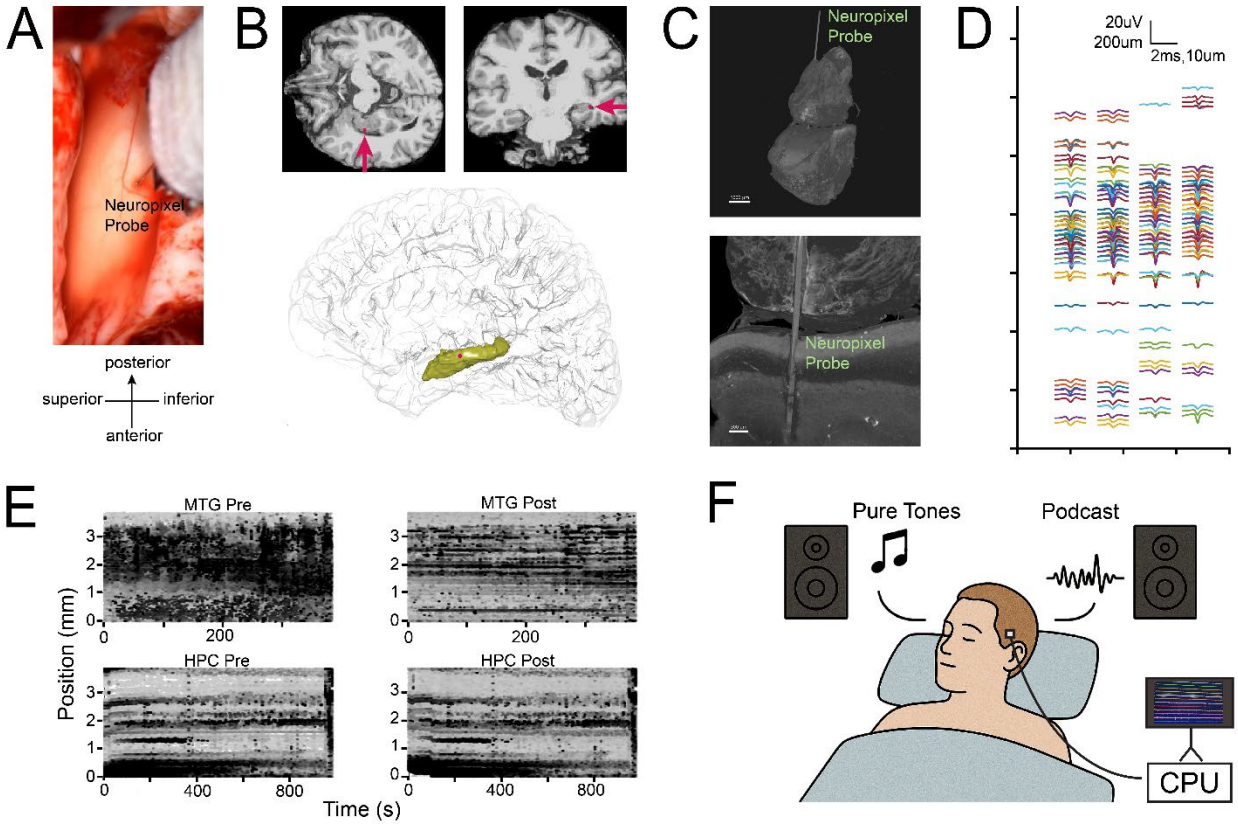

Figure 1. Intraoperative Neuropixels recordings in the human hippocampus. **A.** Photograph of the hippocampal brain tissue with the inserted Neuropixels probe during intraoperative recording (middle right), with the anatomical orientation indicated below. **B.** Top: axial (left) and coronal (right) 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.
.

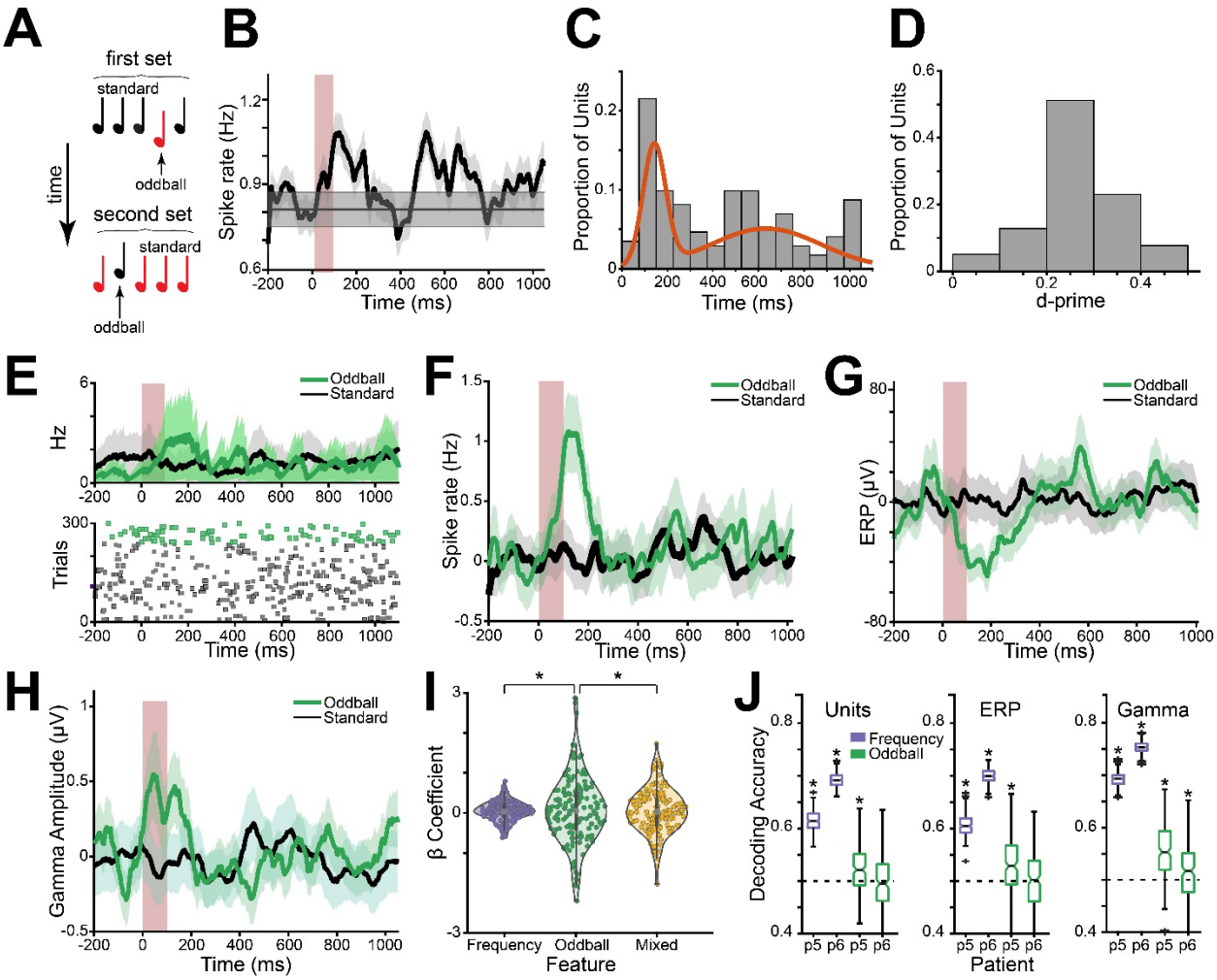

Figure 2. Oddball responses in the anesthetized human hippocampus.
A. Schematic of the auditory oddball task (n=3 patients). Each trial consisted of pure tones with predetermined tone frequencies played at different probabilities. For two of the patients, the oddball and standard tone identities were 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

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

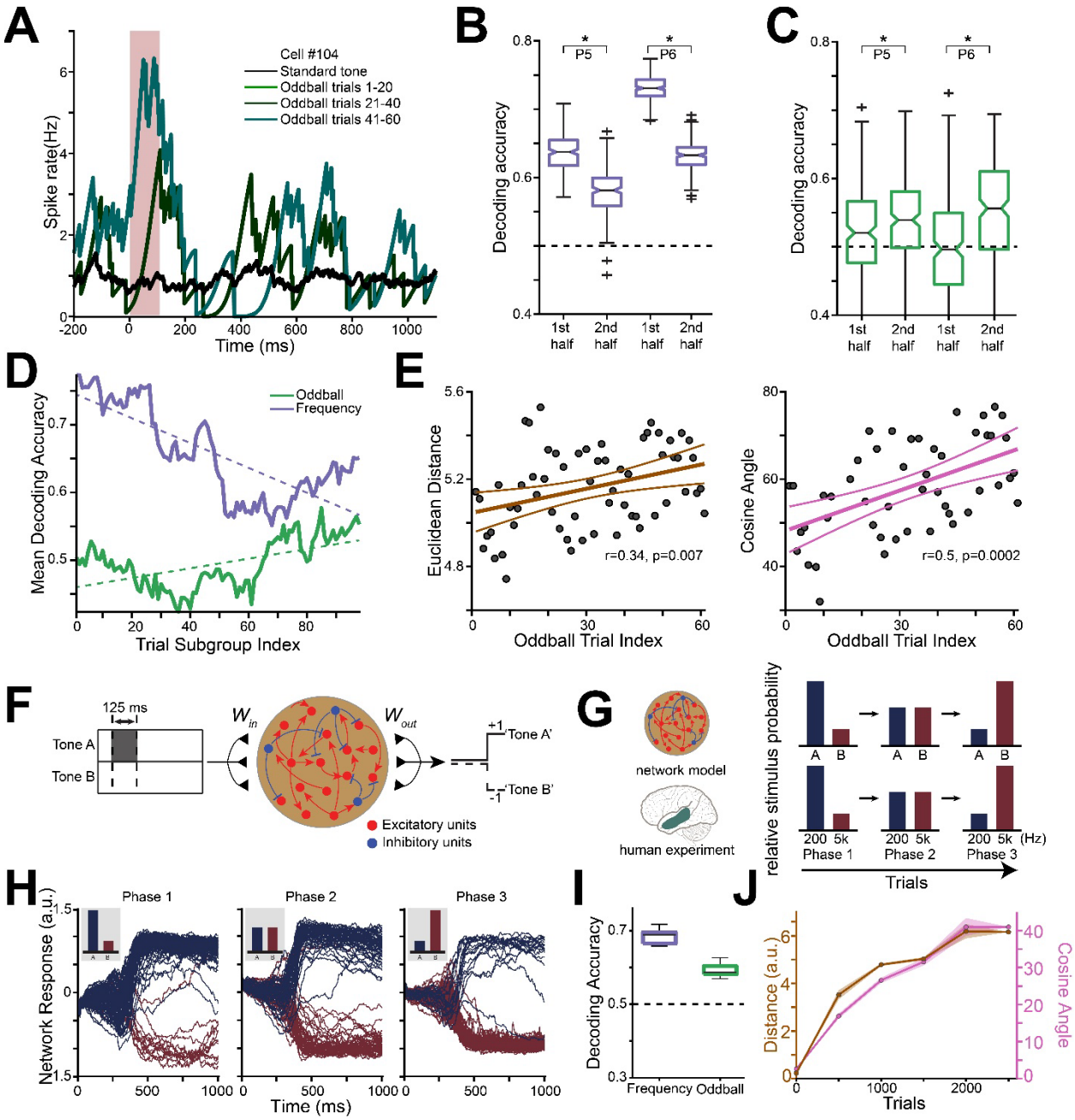

Figure 3: Evolution of the oddball representation across the neuronal population in experimental data and an RNN model.

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

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

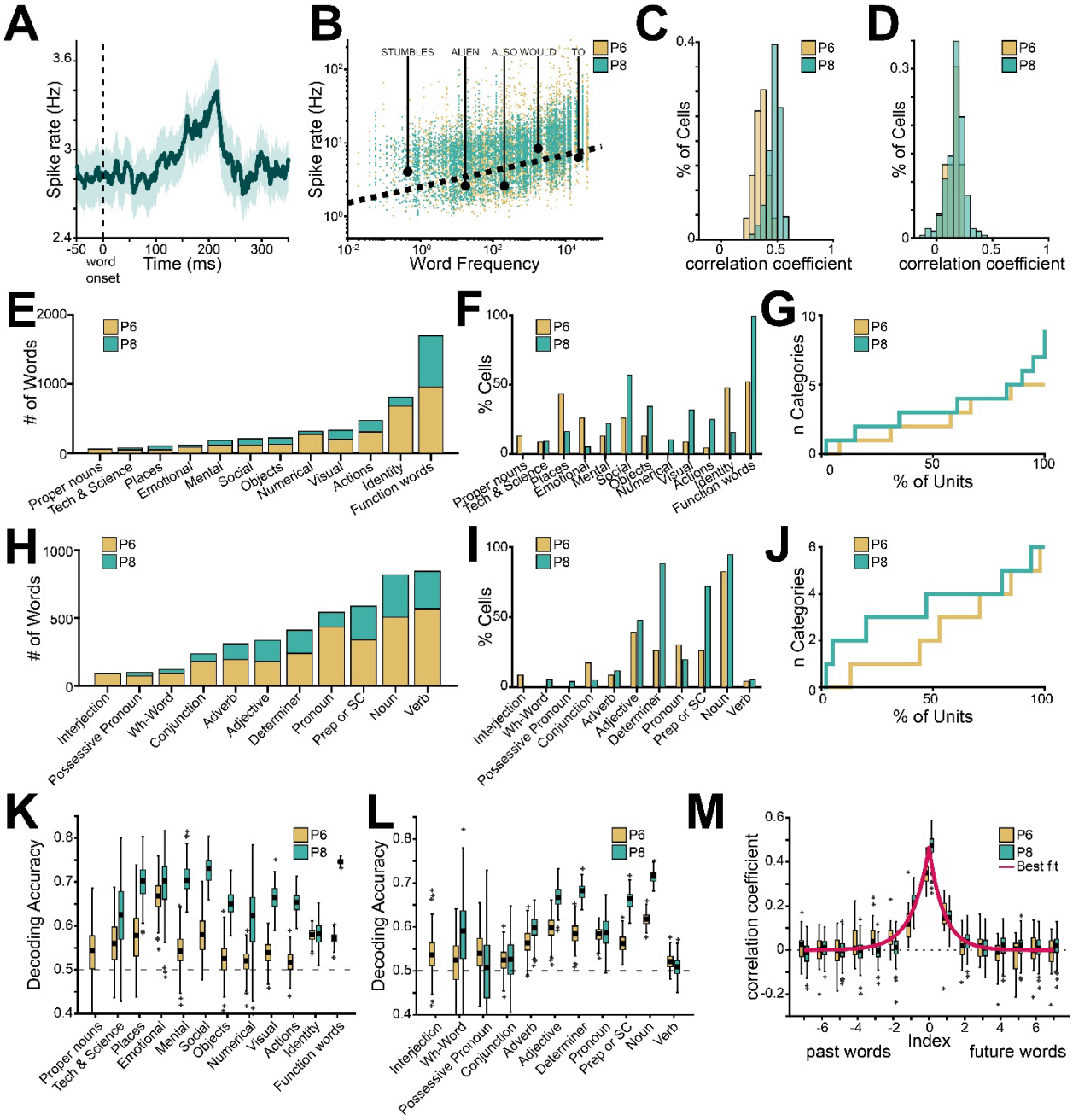

Figure 4. Neuronal responses to natural language in the anesthetized human hippocampus. **A.** Average neuronal response (spike rate, Hz) of an example unit across all words presented (n=3019 words), shown aligned to word onset, indicated with a vertical dashed line. Shaded area represents +/- SEM. Activity 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.**

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

447 # Methods

448 **Patient recruitment**

449 Experiments were conducted according to protocol guidelines approved by the Institutional
450 Review Board for Baylor College of Medicine and Affiliated Hospitals, Houston TX (H-50885).
451 All recruited patients were diagnosed with drug resistant temporal lobe epilepsy and were
452 scheduled to undergo an anteromesial temporal lobectomy for seizure control. All patients
453 provided written informed consent to participate in the study and were aware that participation
454 was voluntary and would not affect their clinical course. Included patients' age ranged from 31-
455 54 years old (average 43.6 +/- 8.4), with three females and two male patients. Two resections were
456 on the left side, and three were on the right. None of the patients reported explicit memory of
457 intraoperative events after the case when discussed in the post-operative care unit or while
458 recovering in the hospital the next day.

459 **Neuropixels Data Acquisition Setup and Intraoperative Recordings**

460 Neuropixels 1.0-S probes (IMEC) with 384 recording channels (total recording contacts = 960,
461 usable recording contacts = 384) were used for recordings (dimensions: 70µm width, 100µm
462 thickness, 10mm length). The Neuropixels probe, consisting of both the recording shank and the
463 headstage, were individually sterilized with ethylene oxide (Bioseal, CA).[1] Our intraoperative data
464 acquisition system included a custom-built rig including a PXI chassis affixed with an
465 IMEC/Neuropixels PXIe Acquisition module (PXIe-1071) and National Instruments DAQ (PXI-
466 6133) for acquiring neuronal signals and any other task-relevant analog/digital signals
467 respectively. Our recording rig was certified by the Biomedical Engineering at Baylor St. Luke's
468 Medical Center, where the intraoperative recording experiments were conducted. A high-

performance computer (10-core processor) was used for neural data acquisition using open-source software such as SpikeGLX 3.0 and OpenEphys version 0.6x for data acquisition (AP band (spiking data), band-pass filtered from 0.3kHz to 10kHz was acquired at 30kHz sampling rate; LFP band, band-pass filtered from 0.5Hz to 500Hz, was acquired at 2500Hz sampling rate). We used a "short-map" probe channel configuration for recording, selecting the 384 contacts located along the bottom 1/3 of the recording shank.

Audio was played via a separate computer using pre-generated wav files and captured at 30kHz or 1,000kHz on the NIDAQ via a coaxial cable splitter that sent the same signal to speakers adjacent to the patient. MATLAB (MathWorks, Inc.; Natick, MA) in conjunction with a LabJack (LabJack U6; Lakewood, CO) was used to generate a continuous TTL pulse whose width was modulated by the current timestamp and recorded on both the neural and audio datafiles. Online synchronization of the AP and LFP files was performed by the OpenEphys recording software. Offline synchronization of the neural and audio data was performed by calculating a scale and offset factor via a linear regression between the time stamps of the reconstructed TTL pulses and confirmed with visual inspection of the aligned traces.

Acute intraoperative recordings were conducted in brain tissue designated for resection based on purely clinical considerations. The probe was positioned using a ROSA ONE Brain (Zimmer Biomet) robotic arm and lowered into the brain 5-6mm from the ependymal surface using an AlphaOmega microdrive. The penetration was monitored via online visualization of the neuronal data and through direct visualization with the operating microscope (Kinevo 900). Reference and ground signals on the Neuropixels probe were acquired separately by connecting to a sterile microneedle placed in the scalp (separate needles inserted at distinct scalp locations for ground 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

open-source, cross-platform application developed in Python. The 3D volumes were analyzed, and optical sections were captured using 3D Slicer.[5]

Neuronal Data Processing

Motion-correction

We utilized previously developed and validated motion estimation and interpolation algorithms to correct for the motion artifacts from brain movement.[6] Motion was estimated via the DREDge software package (Decentralized Registration of Electrophysiology Data software, https://github.com/evarol/DREDge) using either a combination of motion traces obtained using raw LFP and/or AP band data, fine-tuned for individual recordings. Motion-correction was then implemented using interpolation methods (https://github.com/williamunoz/InterpolationAfterDREDge). Both the AP and LFP band data are motion-corrected and utilized for further pre-processing and analysis steps. If the estimated motion led to no improvement in the spike locations then spike sorting proceeded with the motion correction package built into Kilosort 4 without performing interpolation.

Unit extraction and classification

Automated spike detection and clustering were performed by Kilosort 2.0 if motion correction was already applied using the DREDge algorithm or KiloSort 4.0[7] if motion correction was not applied separately. Manually curation of spike clustered was performed using the open-source software Phy.[8] Unit quality metrics were calculated using SpikeInterface[9] and were considered single units if they had a d-prime (d') greater than 1 and fewer than 3% of spikes were violations of a 2ms inter-spike interval refractory period.

Local Field Potential data

LFP data was bandpass-filtered between 0.1-20Hz and aligned to task events to extract local ERPs. Gamma band amplitude was calculated in the "high gamma" range, first bandpass filtering it between 70-150Hz and then calculating the absolute value of the Hilbert-transformed complex signal. Given the high correlation between adjacent channels, only 10 channels equally spanning the length of the probe were used to calculate statistics.

Neuronal Data Analysis:

All analyses were performed using custom MATLAB code.

Motion Analysis

The motion-corrected location estimates were obtained at a 250Hz sampling frequency using the DREDge algorithm. This signal was downsampled to 10Hz. The power spectrum of the calculated motion was then estimated using Welch's overlapped segment averaging estimator for frequencies between 0.1 and 3Hz. The amount of motion was defined as the root mean square error of the location trace of the probes center relative to its average location.

Tone Responses

Both single units and multi-units were used for all analyses. A tone responsive neuron was defined as having a statistically significant increase in the average firing rate in the first second after tone onset (shifted by 50ms to account for auditory delay) relative to the preceding 200ms baseline ($\propto<0.05$, Wilcoxon signed-rank test). Visual demonstrations of the peri-stimulus average firing

rate were smoothed via a causal Gaussian filter with a standard deviation of 150ms for visualization, however, all statistical analyses were performed on the raw spike count. Response onset latency was computed as the time taken to the peak response. A Mixed Gaussian Model with two components was then fit to the distribution of latencies. Given the trough between the two peaks at 291ms and evidence of average oddball response occurring in the first segment, a window of 0-300ms was used for analysis characterizing tone and oddball selectivity.

Neural Tuning

To determine response tuning properties, we modeled trial responses in the peristimulus period using general linear regression models. Neural data in the analysis time window of 0-300ms was used for tuning analyses. Unit response was defined as the average firing rate, LFP power was defined as the root mean square (RMS) value of the bandpass-filtered LFP, and gamma power was defined as the average gamma band amplitude. All vectors were z-scored to allow for comparison of the neural response modulation across units/channels. The independent variables were effects-coded for tone type (frequency 1 vs. frequency 2), trial type (standard vs. oddball), and an interaction term (conjunctive coding). We set the α level at 0.05 to determine if the beta coefficient for the independent variables were statistically significant.

Neuronal Population Coding

To determine the information content present in the population, a Support Vector Machine with a linear kernel was trained using 10-fold cross validation for 200 iterations. Accuracy for each iteration was defined as the average accuracy across the 10 folds. Significant coding was defined as the distribution of 200 iterations being statistically different from 0.5 (chance). Algorithm validation was performed by shuffling the dataset and demonstrating that it always performed at

chance level. Subsampling was performed to avoid performance bias from an unbalanced dataset (i.e. more standard trials than oddball trials). To investigate the neuronal population response dynamics for tone and oddball encoding as a function of time, we used sets of sequential trials (50 trials) from each of the two counterbalanced blocks (total of 100 trials). For example, the first set was using trials 1:50 and 181:230, whereas the last set was using trials 101:150 and 281:330. Decoding analyses were also run separately for early vs. late trials (first 75 vs. last 75 trials within a 150-trial block) for tone and oddball encoding respectively.

Neuronal response learning dynamics

Next, to determine the neural mechanism underlying statistical learning required for oddball detection, we evaluated single-trial response dynamics across the neuronal population. For each trial, we generated a neuronal response population vector. We then computed the Euclidean distance ($||u - v||$) and cosine angle ($invcos(u \cdot v/||u|| * ||v||)$) between the mean vector across all standard trials and each individual oddball unit vector, evaluating each as a function of the oddball index.

Continuous-rate RNN model. We implemented a continuous-rate recurrent neural network (RNN) and trained it to perform an oddball detection task, closely mirroring the one used for the experimental dataset. The network contains 200 recurrently connected units (80% of which are excitatory and 20% of which are inhibitory units). The network is governed by the following equation:

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

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

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

where i represents the synaptic decay time constant for unit i, xi(t) indicates the synaptic current variable for neuron i at time point t, W is the recurrent connectivity matrix (N-by-N; i.e. 200-by-200), and u(t) is the input data given to the network at time point t. u is a 2-by-200 matrix where the first dimension refers to the number of input channels and the second dimension is the total number of time points. A firing rate of a unit was estimated by passing the synaptic current variable (x) through a standard logistic sigmoid function. The output (o) of the network was computed as a linear weighted sum of the entire population of units.

In each trial, the network model receives input data mimicking auditory signals. The input consists of two signal streams, each representing a distinct auditory tone (i.e. tone A vs. tone B; [**Figure 3F,G**]). Only one tone is presented to the network per trial. The model was trained to produce an output signal approaching +1 when Tone A was presented and an output signal approaching -1 when Tone B was presented. To closely replicate the experimental task design, we employed three different sequential contexts during network training. In the first stage, Tone A was presented predominantly (80% of trials), followed by an equal distribution of Tone A and Tone B (50/50) in the second stage. In the third stage, Tone B was predominant (80%).

We optimized the network parameters, including recurrent connectivity, readout weights, and synaptic decay time constants, using gradient descent via backpropagation through time (BPTT). The network was required to achieve over 95% task accuracy in the current context before a new context was introduced.

Neuronal Data Analysis: Natural Language stimuli

Natural language stimuli

All patients were native English speakers. The podcast played during the task was automatically transcribed using Assembly AI (https://www.assemblyai.com/). The transcribed words and corresponding timestamp outputs from Assembly AI were converted to a TextGrid and then loaded into Praat.[10] The original wav file was also loaded into Praat and the spectrograms and labels and timestamps were manually checked and corrected to ensure the word onset and offset times were accurate. This process was then repeated by a second reviewer to ensure the validity of the time stamps. The TextGrid output of corrected words and timestamps from Praat was converted to a xls and loaded into Matlab and Python for further analysis.

Natural Language statistics

Word frequency was defined based on a corpus of movie subtitles spanning a total of 51 million words.[11] Words that did not elicit a response during the duration of the word were excluded from this analysis. To compare the relative contributions to firing rate, a linear model was trained to estimate the logarithmic firing rate from the logarithmic duration and corpus frequency of each word. Word surprisal values were calculated using the GPT-2 large model[12] from the Hugging Face Transformers library,[13] computing the negative log probability of each word conditioned on the preceding context. Specifically, surprisal was defined by the equation: $surprisal = -log\ P(w_i|w_{i-1}, w_{i-2}, \ldots, w_2, w_1)$ where $P(w_i)$ refers to the probability of word i given the proceding words.

We utilized the pre-trained fastText Word2Vec model in MATLAB to extract word embeddings for all words in our dataset.[14,15] This pre-trained model provides 300-dimensional word embedding vectors, trained on 16 billion tokens from Wikipedia, UMBC webbase corpus, and

statmt.org, to capture semantic relationships between words. Notably, Word2Vec is a non contextual embedder, so all instances of the same word are represented the same. Some surname words, such as “Harwood” or proper nouns like “Applebee’s” did not have word embeddings and were discarded from the analysis. A simple linear model was trained to predict the firing rate of individual neurons from the semantic matrices using 10-fold cross-validation. Accuracy was defined as the correlation between true and predicted firing rates. Words with 0Hz or above 25Hz were removed from this analysis. To prevent overfitting, Principal Component Analysis (PCA) was used to reduce the dimensionality to account for 30% of the variance prior to modeling. This threshold was defined as the minimum of the RMSE of the model that balanced under and overfitting. To predict future or previous words the alignment between words was shifted forwards or backwards, respectively. This relation was then fit with a piecewise exponential decay

$$r(i) = \beta_0 * \left\{ \frac{e^{1/\beta_1} for\ i >= 0}{e^{-i/\beta_2} for\ i < 0} \right\}$$

Wherein β_0 is the amplitude of the correlation at 0 lag, and β_1 and β_2 are equivalent to the time constant of the decay for positive and negative lags, respectively.

Word embedding, Semantic clustering, and Part of Speech classification

To identify the natural semantic categories present in our word data, all unique words heard by the participants were clustered into groups using a word embedding approach.[14–16] We used the same 300-dimensional embedding from the prior GLM analysis. To compute and visualize semantic clusters, we first used a t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm on word

embedding values to reduce the dimensionality of each unique word based on their cosine distance to all other words, thus reflecting their semantic similarity. Words with similar meanings now have similar 2D coordinates. We then applied the k-means clustering algorithm to these 2D word representations and visualized clustered words on a 2D word map (12 clusters).[17] We then manually inspected and assigned a distinct label to each semantic cluster and adjusted clusters for accuracy. For example, words bordering the edges of clusters would sometimes get mis-grouped and were manually corrected. The final 12 semantic categories of the words are body parts, places, emotional words, mental words, social words, objects, visual words, numerical words, actions, identity words, function words, and proper nouns. Correction for multiple comparisons was performed using the Benjamini Hochberg approach.[18] A SVM was trained for each semantic category (versus all other categories) using a radial basis function (‘RBF’) kernel. Model training and accuracy metrics were weighted to the relative frequency of each group. We used 10-fold cross validation and 200 iterations.

To extract part-of-speech (POS) for each word in the dataset, we utilized an automated pipeline through Stanford CoreNLP, a natural language processing toolkit.[19] We initialized a CoreNLPParser with the 'pos' tagtype, which specializes in part-of-speech tagging. The transcript was first segmented into sentences based on punctuation. Each sentence was then tokenized and passed through the CoreNLPParser's tagging function. This process leveraged CoreNLP's advanced linguistic models to analyze the context and structure of each sentence, assigning appropriate POS tags to individual words. The 15 POS types were: 'Noun', 'Adjective', 'Numeral', 'Determiner', 'Conjunction', 'Preposition or Subordinating Conjunction', 'Auxiliary', 'Possessive Pronoun', 'Pronoun', 'Adverb', 'Particle', 'Interjection', 'Verb', 'Wh-Word', and 'Existential'. POS types with fewer than 45 words were removed from analysis. A similar SVM was used for POS.

Bibliography

1. Coughlin, B. *et al.* Modified Neuropixels probes for recording human neurophysiology in the operating room. *Nat Protoc* **18**, 2927–2953 (2023).
2. Singh, H. Bispectral index (BIS) monitoring during propofol-induced sedation and anaesthesia. *Eur J Anaesthesiol* **16**, 31–36 (1999).
3. Hsu, C.-W. *et al.* High resolution imaging of mouse embryos and neonates with X-ray micro-computed tomography. *Curr Protoc Mouse Biol* **9**, e63 (2019).
4. Brown, J. M. *et al.* A bioimage informatics platform for high-throughput embryo phenotyping. *Brief Bioinform* **19**, 41–51 (2018).
5. Fedorov, A. *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* **30**, 1323–1341 (2012).
6. Windolf, C. *et al.* DREDge: robust motion correction for high-density extracellular recordings across species. *Nat Methods* 1–13 (2025).
7. Pachitariu, M., Sridhar, S., Pennington, J. & Stringer, C. Spike sorting with Kilosort4. *Nat Methods* **21**, 914–921 (2024).
8. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat Neurosci* **19**, 634–641 (2016).
9. Buccino, A. P. *et al.* SpikeInterface, a unified framework for spike sorting. *Elife* **9**, (2020).
10. Boersma, P. Praat: doing phonetics by computer [Computer program]. *Httpwww Praat Org* (2011).
11. Brysbaert, M. & New, B. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* **41**, 977–990 (2009).
12. Radford, A. *et al.* Language Models are Unsupervised Multitask Learners. in (2019).
13. Wolf, T. *et al.* Transformers: State-of-the-Art Natural Language Processing. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System*

Demonstrations (eds. Liu, Q. & Schlangen, D.) 38–45 (Association for Computational Linguistics, Online, 2020). doi:10.18653/v1/2020.emnlp-demos.6.

14. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. *ArXiv CsCL* (2013).
15. Joulin, A., Grave, É., Bojanowski, P. & Mikolov, T. Bag of Tricks for Efficient Text Classification. in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* 427–431 (2017).
16. Henry, S., Cuffy, C. & McInnes, B. T. Vector representations of multi-word terms for semantic relatedness. *J Biomed Inf.* **77**, 111–119 (2018).
17. Franch, M. *et al.* A vectorial code for semantics in human hippocampus. *bioRxiv* 2025–02 (2025).
18. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
19. Manning, C. D. *et al.* The Stanford CoreNLP Natural Language Processing Toolkit. in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* 55–60 (2014).

Acknowledgements
This project was supported by the Optical Imaging & Vital Microscopy Core at the Baylor College of Medicine and by the McNair Foundation.
This project was funded in part by U01 NS121472, NINDS

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