

# 1 Decoding movie content from neuronal 2 population activity in the human medial 3 temporal lobe

4 Franziska Gerken <sup>1\*</sup>, Alana Darcher <sup>2\*</sup>, Pedro J Gonçalves <sup>3,4,5</sup>, Rachel Rapp <sup>3</sup>, Ismail Elezi <sup>1§</sup>,  
5 Johannes Niediek, <sup>2,8</sup>, Marcel S Kehl <sup>2,9</sup>, Thomas P Reber <sup>2,10</sup>, Stefanie Liebe, <sup>3,6</sup>, Jakob H  
6 Macke <sup>3,7‡</sup>, Florian Mormann <sup>2‡</sup>, Laura Leal-Taixé <sup>1†‡</sup>

7 <sup>1</sup>Dynamic Vision and Learning Group, Technical University of Munich, Munich, Germany; <sup>2</sup>Dept. of  
8 Epileptology, University Medical Center of Bonn, Bonn, Germany; <sup>3</sup>Machine Learning in Science,  
9 Excellence Cluster Machine Learning and Tübingen AI Center, University of Tübingen, Tübingen,  
10 Germany; <sup>4</sup>VIB-Neuroelectronics Research Flanders (NERF), Belgium; <sup>5</sup>imec, Belgium; <sup>6</sup>Department of  
11 Epileptology and Neurology, University Hospital Tübingen, Tübingen, Germany; <sup>7</sup>Empirical Inference,  
12 Max Planck Institute for Intelligent Systems, Tübingen, Germany; <sup>8</sup>Machine Learning Group, Technical  
13 University of Berlin, Berlin, Germany; <sup>9</sup>Department of Experimental Psychology, University of Oxford,  
14 Oxford, UK; <sup>10</sup>Faculty of Psychology, UniDistance Suisse, Brig, Switzerland

---

\*These authors contributed equally to this work.

†These authors contributed equally to this work.

‡Currently at NVIDIA.

§Currently at Huawei.

For correspondence: Jakob.Macke@uni-tuebingen.de (Jakob H Macke); florian.mormann@ukbonn.de (Florian Mormann); llealtaixe@nvidia.com (Laura Leal-Taixé)

15

---

16 **Abstract** The human medial temporal lobe (MTL), a region implicated in memory and high-level cognition, contains  
17 neurons that respond selectively to stimuli belonging to specific categories, such as individual people, landmarks, or  
18 objects. However, these neurons have been largely studied via static, isolated presentations of stimuli. Therefore, it  
19 is unclear how neurons in the MTL respond to rich stimuli such as movies, and which dynamical stimulus features can  
20 be retrieved from neuronal population spiking activity. We studied single-unit responses from 2286 neurons recorded  
21 from the amygdala, hippocampus, entorhinal cortex, and parahippocampal cortex of 29 intracranially implanted pa-  
22 tients during the presentation of an 83-minute movie. We found only a few individual neurons that exhibited a classic  
23 selective response to semantic features. However, we successfully decoded the presence of characters, settings,  
24 and visual transitions from neuronal population activity. The information relevant for decoding varies across regions  
25 depending on the feature category, as visual transitions could be decoded from subsets of neurons with selective  
26 responses, whereas character and location features relied on distributed representations. Our results demonstrate  
27 an approach for reliably decoding movie features in the human MTL, and suggest that the brain uses a population  
28 code when representing character and location features.

29

---

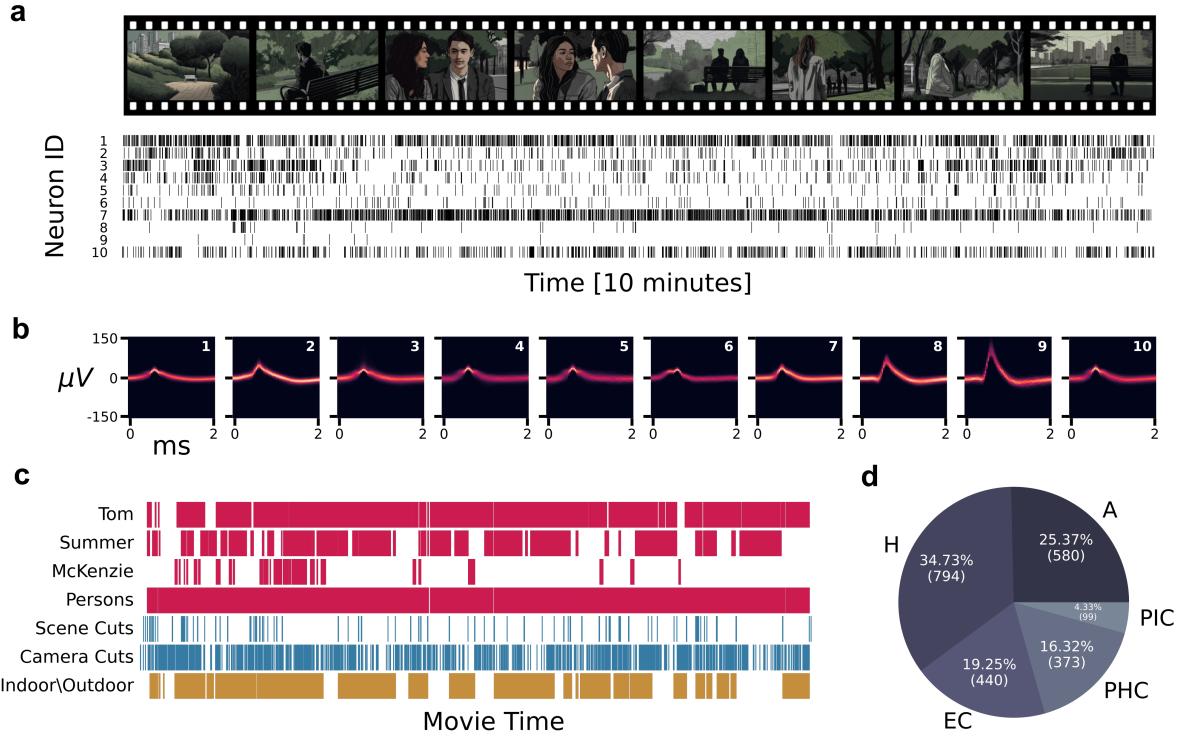
## 30 **Introduction**

31 The human medial temporal lobe (MTL) plays an integral role in the representation of semantic information. Single  
32 neurons in the MTL exhibit strong and highly selective tunings to categories, such as faces or locations<sup>1,2</sup>, and to spe-  
33 cific concepts, such as individual celebrities or objects<sup>3,4</sup>. These semantically tuned neurons relate to the processing  
34 of information at a conscious, declarative level, as their activity varies depending on perception<sup>5-7</sup>—when images are  
35 presented but unseen, these neurons exhibit reduced and delayed spiking compared to consciously seen images<sup>8</sup>.  
36 Such cells also support the formation of new memories<sup>9</sup>, and are involved in the retrieval of previously encoded  
37 experiences<sup>10,11</sup>.

38 Studies investigating the representation of semantic information in the human MTL have mostly focused on char-  
39 acterizing these neurons individually, often without considering their population dynamics. Many of these studies  
40 have identified semantically tuned cells by screening for stimulus-selective responses to static images depicting iso-  
41 lated persons or objects<sup>12,13</sup>. While this approach is effective for probing specific functional properties of individual  
42 neurons, it limits the generalizability of findings to more complex and dynamic contexts. Naturalistic and dynamic  
43 stimuli, such as movies, provide closer approximations to real-world environments, but also pose substantial chal-  
44 lenges, as they are presented continuously and depict complex evolving scenes. Several studies have used functional  
45 magnetic resonance imaging (fMRI) to study neural responses to natural movies. One line of work has focused on  
46 representations in early visual areas<sup>14,15</sup> and across the cortex<sup>16,17</sup>. Another line of work has identified synchrony  
47 in the brain states of separate individuals viewing a common movie<sup>18,19</sup>. Such states appear to be particularly well  
48 aligned in individual brains surrounding transition events within an ongoing continuous stimulus, as reported in fMRI  
49 studies<sup>20,21</sup>, intracranial field potentials along the human ventral visual pathway<sup>22</sup>, and in single neurons in the hu-  
50 man parahippocampal gyrus, hippocampus, and amygdala<sup>11</sup>. Most single neuron studies which use movies as stimuli  
51 have been based on only short 'snippets' of movies, and have so far largely examined representation on the level of  
52 the individual neuron<sup>23</sup> with few having addressed frame-wise representations in longer movie sequences<sup>24</sup>. How-  
53 ever, it remains an open question of how populations of neurons in the MTL collectively respond to naturalistic visual  
54 stimuli, such as those encountered in real-world environments, and which features of population activity encode  
55 information about specific components of such stimuli.

56 In this study, we investigate how information embedded in a naturalistic and dynamic stimulus is processed by  
57 neuronal populations in the human medial temporal lobe (MTL). Specifically, we asked the following questions: (1)  
58 Which aspects of a movie's content can be decoded from neuronal activity in the MTL? (2) Which brain regions are  
59 informative for specific stimulus categories (e.g. visual transitions or characters)? (3) Is the relevant information  
60 distributed across the neuronal population? We recorded the activity of neurons from patients with intracranially  
61 implanted electrodes as each watched the full-length commercial movie "500 Days of Summer". Our dataset is unique  
62 in both size and duration: we recorded a total of 2286 neurons across 29 patients during the complete presentation of  
63 the movie. To analyze the relationship between the neuronal activity and the film's content, we labeled the presence  
64 of main characters, whether a scene was indoors or outdoors, and visual transitions of the movie on a frame-by-frame  
65 basis.

66 We introduce a machine learning-based decoding pipeline that decodes a movie's visual content from population-  
67 level neuronal activity. At the single-unit level, individual neurons generally lacked reliable responses to the visual  
68 features, and we observed consistent stimulus-related changes in firing rates primarily during visual transitions. How-  
69 ever, at the population level, we achieved strong decoding performance across all visual features. For visual transi-  
70 tions, neurons exhibiting consistent changes in activity played a key role in population-level decoding. In contrast, no  
71 similar pattern emerged when decoding character identities. By analyzing the contributions of individual neurons, we  
72 identified distinct subsets of neurons that influenced decoding performance, extending beyond the subsets identi-  
73 fied by stimulus-aligned changes in firing activity. Remarkably, we found that restricting the analysis to a substantially  
74 smaller subset of these key neurons was sufficient to replicate the full population-level decoding performance. Taken  
75 together, our findings show that information about dynamic stimuli can be decoded from neuronal population activ-  
76 ity, even in the absence of strong single-neuron selectivity.

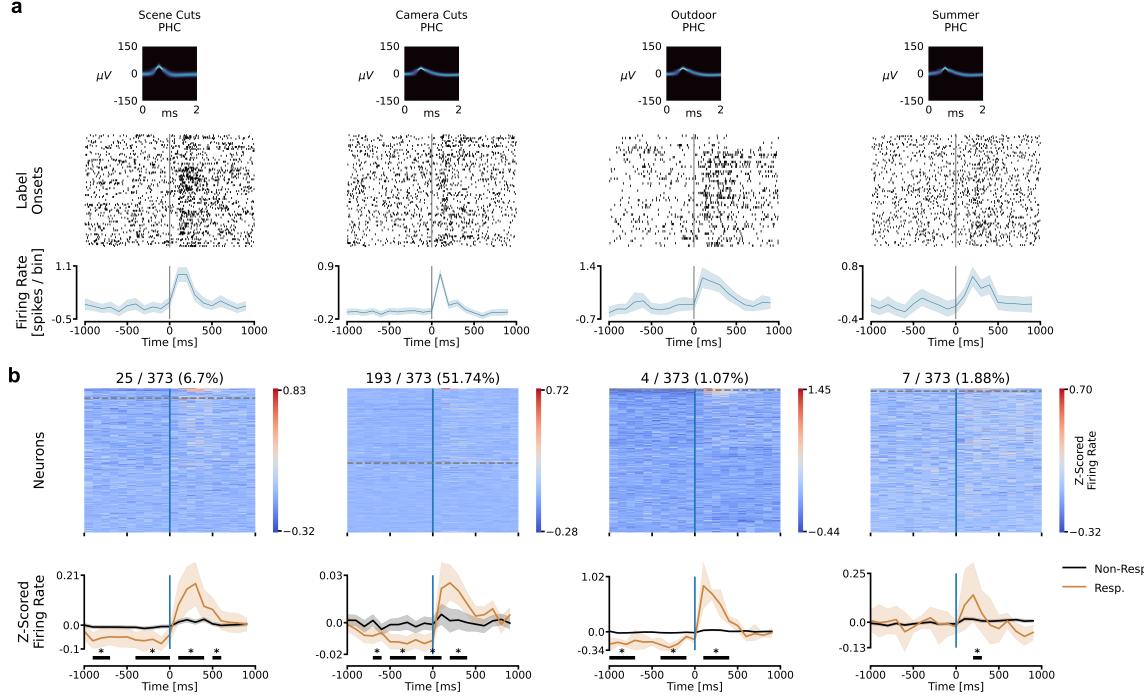


**Figure 1. Overview of dataset, features, and decoding approach** **a)** Example of the recorded neuronal activity. Patients watched the complete commercial film *500 Days of Summer* while neuronal activity was recorded via depth electrodes. Top row: example movie frames. Due to copyright, the original movie frames have been replaced with images generated using stable diffusion<sup>25</sup>. Bottom row: spike trains from ten amygdala neurons of a single patient, where each row shows data from an individual neuron (corresponding ID number given as a label). **b)** Spike density plot showing the waveforms of each neuron in **a** (corresponding neuron ID given in top right). Neurons shown include both single- and multi-neurons. **c)** Distribution of labels across the entire movie (runtime: 83 minutes). Occurrences of character-related features are in magenta, visual transition events in blue, and location events in yellow. **d)** Distribution of the 2286 neurons across the recorded regions (A: amygdala, H: hippocampus, EC: entorhinal cortex, PHC: parahippocampal cortex, PIC: piriform cortex) for all 29 patients.

## 77 Results

78 We recorded from 29 patients (17 female, ages 22 – 63) as each watched the movie *500 Days of Summer* (83 minutes) in its entirety (Fig. 1a). Patients were bilaterally implanted with depth electrodes for seizure monitoring, and spiking activity was recorded from a total of 2286 single- and multi-neurons across the amygdala (A; 580 neurons, 25.37%), hippocampus (H; 794, 34.73%), entorhinal cortex (EC; 440, 19.25%), parahippocampal cortex (PHC; 373, 16.32%), and piriform cortex (PIC; 99, 4.33%) (Fig. 1b). We pooled the neurons across patients (distribution shown in Supp. Fig. S1) and performed subsequent analyses on the resulting population. Due to the low number of neurons relative to the complete population, the PIC was excluded from subsequent region-wise analyses.

85 To determine which features of the movie are represented in MTL activity, we obtained frame-wise annotations of character presence, indoor/outdoor setting, and the occurrence of a visual transition, i.e. camera cuts and scene cuts (similar to "soft" and "hard" boundaries<sup>11</sup>) (Fig. 1c). All annotations concern the visual occurrence of a given feature in the frame, and do not consider feature content of the movie's audio. We used a mixture of manual and automated methods for obtaining these annotations (see Methods; sample frames in Supp. Fig. S2). For character-related content, we restricted the analysis to those characters most relevant to the movie's narrative (Tom, Summer, McKenzie) as determined by screentime. For the analysis of location, we investigate indoor and outdoor settings. Note that these two features are mutually exclusive, and the annotation of both are combined into a single label.



**Figure 2. Responsive single-neurons in the parahippocampal cortex** **a)** Example peri-stimulus activity for representative parahippocampal (PHC) neurons, for labels with a significant PHC response. Upper plots: spike density plot showing the waveforms a given responsive neuron (label name given as title). Middle plots: spike time rasters showing the neuron's activity surrounding the onset of the corresponding label throughout the movie. Note: onsets for Scene Cuts and Camera Cuts were randomly subsampled to match the number of Summer appearances. Lower plots: average firing rate across 100 ms bins, for 1000 ms before and 1000 ms after the onset or event. Solid lines show the mean across all neurons, within group, and the transparent area shows the 95% confidence interval. **b)** Region-wise single-neuron activity surrounding the onset of labeled entity. Number of cells exhibiting a significant response over the total number of PHC cells are given as the title, followed by the corresponding percentage. Upper plots (heatmaps): averages of peri-stimulus spike rates per neuron (spikes per 100 ms bin, z-scored across the pseudotrial) for 1000 ms before and 1000 ms after label onset. Each row of the heatmap represents the average binned activity for one neuron. Neurons are sorted in descending order by the p-value of the response—the dotted grey line shows the threshold for responsive neurons ( $p \leq 0.001$ ). Lower plots (line plots): average z-scored firing rate across bins. Neurons are separated into responsive (orange line) and non-responsive (black line). Solid lines show the mean for each group of neurons (responsive vs. non-responsive) and the transparent area depicts the 95% confidence interval. Significant differences between the responsive and non-responsive firing rates are shown as solid black lines (\*,  $p \leq 0.05$ , cluster permutation test).

### 93 **Stimulus-aligned responsive neurons found primarily in parahippocampal cortex**

94 To investigate whether the visual features of the movie are encoded at the level of individual neurons, we analyzed  
 95 stimulus-evoked changes in the firing rate of neurons after the onset of characters, indoor and outdoor scenes, and  
 96 visual transitions, for each MTL subregion separately (Fig. 2a, Supp. Fig. S3 - S5). We classified each individual neu-  
 97 ron as responsive or non-responsive according to a previously established criterion<sup>8</sup> which compares the spiking  
 98 activity after stimulus onset to that of a baseline period, adapted to the dynamic presentation. Specifically, for each  
 99 annotated feature, we identified all instances where the feature appeared following at least 1000 ms of absence and  
 100 remained continuously present for at least 1000 ms. We applied a cluster permutation test<sup>26</sup> to find time-points in  
 101 which the firing rate between the sets of responsive and non-responsive neurons differ (see Methods for additional  
 102 details). We found individual neurons with significant stimulus-evoked responses in all regions for visual transitions,  
 103 Persons, and the characters Tom and Summer, including Face-only onsets (bin-wise Wilcoxon signed-rank test, Simes-  
 104 corrected,  $\alpha = 0.001$ , see Supp. Table S1). Of these stimulus-evoked changes, neurons in the parahippocampal cortex  
 105 responded to the largest set of stimulus features (Summer, Persons, Camera Cuts, Scene Cuts, Outdoor;  $p \leq 0.001$ ,  
 106 cluster permutation test), as compared to the hippocampus (Tom, Camera Cuts, Outdoor), amygdala (Camera Cuts),

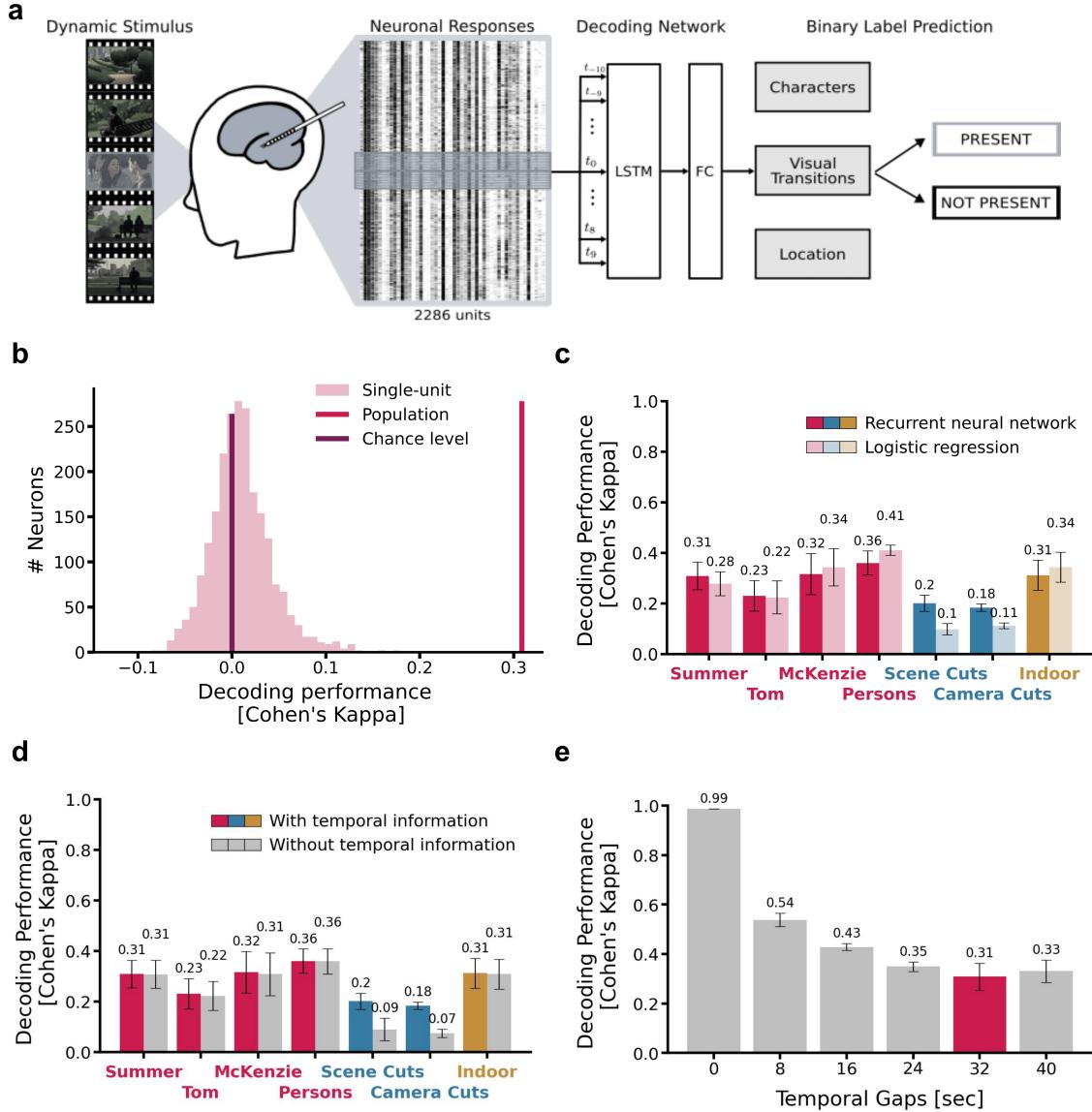
107 and entorhinal cortex (None). Over half of the parahippocampal neurons responded to Camera Cuts during the  
108 movie (193/373, 51.74%, bin-wise Wilcoxon signed-rank test, Simes-corrected,  $\alpha = 0.001$ ) (Fig. 2b), and stimulus-evoked  
109 changes in firing were consistent across the set of responsive neurons ( $p \leq 0.001$ , cluster permutation test). For the  
110 remaining regions, stimulus-evoked modulations were less consistent across individual neurons (see Supp. Fig. S5a).  
111 Comparatively few parahippocampal neurons responded to Scene Cuts (25/373, 6.70%), although there was nonethe-  
112 less a consistent pattern of modulation across responsive neurons ( $p = 0.018$ , cluster permutation test) (Fig. 2b). A  
113 similar pattern was observed for the onset of Outdoor scenes in the parahippocampal cortex (4/373, 1.07%). No re-  
114 sponses were detected for the onset of indoor scenes. For characters, significant stimulus-evoked modulation in the  
115 responsive neurons were only observed for Summer (PHC,  $p = 0.006$ , cluster permutation test) and, to a lesser extent,  
116 Tom (H,  $p = 0.022$ , cluster permutation test) (Fig. 2a, Supp. Fig. S3). Taken together, the parahippocampal cortex  
117 contained neuronal subsets that showed a consistent pattern of increased firing after the onset of a visual transition  
118 (Camera or Scene Cut) within the movie. Characters and character faces evoked a clear change in firing activity at  
119 the subpopulation-level in the parahippocampal cortex, and sparingly in the hippocampus, but not in other tested  
120 regions.

### 121 **Decoding of semantic content from population responses**

122 The observed pattern of responses in individual neurons aligned to feature onsets suggests that these cells primarily  
123 carry information relating to visual transitions, and, to a lesser degree, the characters Summer and Tom, and Persons.  
124 However, previous work suggests that there may be differences in the coding capacity at the single-neuron and popu-  
125 lation level<sup>27</sup>. Could there be stimulus-related information represented at the population level that is not apparent in  
126 the responses of individual neurons? To explore this, we decoded character presence, location, and visual transitions  
127 from the aggregated activity of the entire neuronal population using a recurrent neural network designed to capture  
128 both within-neuron and between-neuron activity patterns.

### 129 **Decoding from neuronal activity**

130 We aligned the activity of neurons across patients using the movie frames as a common time reference, generating a  
131 single neuronal pseudo-population, as done in previous work decoding from populations of single neurons<sup>28,29</sup>. We  
132 first evaluated single-neuron decoding performance for the main character, Summer, by fine-tuning a firing activity  
133 threshold for each neuron. Using the total spike count around frame onset (spanning 800 ms before and after onset),  
134 we optimized a threshold to the validation set from each cell and subsequently predicted Summer's appearance in  
135 the test set. This established a performance baseline for decoding character presence solely from the firing of indi-  
136 vidual neurons, without machine learning-based decoding algorithms. Decoding performances obtained by simply  
137 applying a threshold to single-neuron activity did not reliably predict the content, as the majority of neurons in the  
138 population performed near chance level (Fig. 3b). We then extended this analysis to a population-based approach,  
139 employing a Long Short-Term Memory (LSTM) network<sup>30</sup>, a deep neural network well-suited for processing dynamic  
140 sequential data (Fig. 3a). This population-based approach improved decoding performance for Summer, surpassing  
141 the performances achieved by individual neurons alone. Decoding performances for all tested labels significantly  
142 exceeded chance level (zero for Cohen's Kappa), as shown in Fig. 3c. The Persons label achieved the highest per-  
143 formance (mean  $\pm$  standard error of the mean, SEM:  $0.36 \pm 0.05$ ), followed by the location-related label Indoor/Outdoor  
144 ( $0.31 \pm 0.06$ ). Note that in contrast to the single unit analysis, there is no need to conduct two separate trainings for  
145 the Indoor/Outdoor label since the label combines both features and the network implicitly learns to differentiate  
146 between the two. Character-specific labels (Summer, Tom, McKenzie) showed comparable performances, ranging  
147 between 0.23 and 0.32. Labels related to visual transitions in the movie exhibited the lowest but nonetheless sig-  
148 nificant performances of  $0.2 \pm 0.03$  and  $0.18 \pm 0.01$ , respectively. Decoding results were consistent across metrics  
149 (F1 Score, PR-AUC, and AUROC shown in Supp. Table S2). We evaluated the statistical significance of our results by  
150 randomly permuting the test set labels ( $N = 1000$ ), demonstrating that all decoding performances were significantly  
151 above chance level at an alpha level of 0.001 (see Methods). To ensure that this significance was not an artifact of  
152 the permutation procedure, we conducted an additional test by circularly shifting the neuronal data relative to the  
153 movie features (see Methods for more detail). This approach preserved the temporal structure of each dataset while  
154 disrupting the relationship between the neuronal data and the stimulus. Results were consistent between random  
155 permutation and circular shifts (Supp. Fig. S11).



**Figure 3. Categories of labels can be decoded from the neuronal population activity** **a)** Overview of the neuronal decoding pipeline. Spiking data (individual neurons shown as columns) was sectioned into 1600 ms sequences, 800 ms before and after a frame onset (purple highlight; bins shown as horizontal lines, not to scale), and given as input to a two-layer Long Short-Term Memory (LSTM) network. The output of the fully connected layer (FC) predicts the presence of a given label in a frame. **b)** Assessment of individual-neuron decoding performance by classifying data samples into positive or negative predictions for the label Summer based on the firing activity of a neuron. **c)** Decoding performances on labels of the movie (reported performance using Cohen's Kappa, mean performance across five different data splits, error bars indicate standard error of the mean). Labels fall into one of three categories—characters (pink), visual transitions (blue), or location (yellow)—with a separate model for each label. All performances were significantly better than chance level with an alpha level of 0.001. Decoding performances for the logistic regression model in lighter colors. **d)** Impact of temporal information in spike trains for recurrent neural networks. Trained models were evaluated using temporally altered test data (sequence order shuffled, repeated 100 times). Colored bars depict performance without shuffling, while grey bars represent shuffled scenarios (reported performance using Cohen's Kappa, results show the mean performance across five different data splits, variance given as standard error of the mean). **e)** Decoding performance of the main movie character (Summer) for different temporal gap sizes between samples of training, validation, and test sets. Colored temporal gap of 32 s indicates the chosen gap size for all reported performances in this study.

156 To further test the hypothesis that decoding is based on population activity rather than individual neurons, we  
157 also trained an LSTM network on individual neurons. From the 46 neurons identified as responsive to Summer in the  
158 separate single neuron analysis, we selected a subset of neurons, ensuring an even distribution across both patients  
159 and regions. Similar to the threshold model, most models exhibited minimal prediction performances (see Fig. S6).  
160 In summary, our decoding network achieved consistent and statistically significant decoding performance on the  
161 population level exceeding chance level for all labeled movie features.

## 162 Choice of architecture

163 Our pipeline uses a recurrent neural network (LSTM) to process spiking data as a time series of event counts. We  
164 binned spike counts into 80 ms intervals, covering 800 ms before and after label onset, creating sequences with a  
165 total length of 20 and a dimensionality of 2286 neurons. We trained a separate model for the prediction of each  
166 label, with individually optimized hyperparameters. Given the high degree of imbalance for some labels (i.e., the  
167 character McKenzie only appears in 10 % of the movie's frames), we oversampled the minority class during training to  
168 mitigate the effects of the uneven distribution. We additionally employed a 5-fold nested cross-validation procedure  
169 and carefully selected samples to avoid correlations between samples, as discussed in more detail in the following  
170 section. See Methods for additional details regarding model training and architecture.

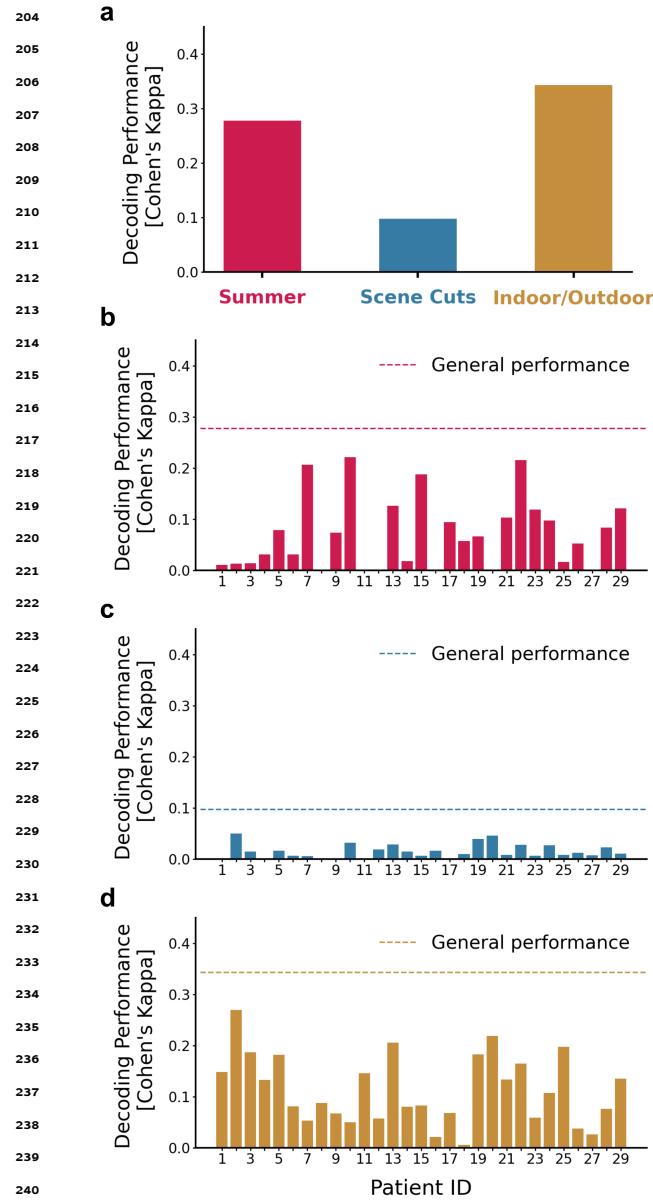
171 We compared the LSTM's decoding performance to a simpler logistic regression model, i.e. a linear method that  
172 does not consider the neuronal activity as a sequence of spike counts, and therefore ignores the temporal informa-  
173 tion and non-linear dynamics (Fig. 3c). Apart from this, the setup and data split for both pipelines were identical. The  
174 logistic regression model showed lower performances for the Scene and Camera Cut features (by 0.1 and 0.07, re-  
175 spectively), whereas no drop for character-related or location-related features was observed. To test our hypothesis  
176 that temporal information within spike sequences influences visual transition decoding, we assessed trained mod-  
177 els using temporally-modified test data (sequence order of the spike trains was randomly shuffled, see Methods for  
178 more details) (Fig. 3d). A pattern consistent with the LSTM decoding results emerged, with a noticeable decline in  
179 performance, especially for the visual transition labels.

## 180 Avoiding spurious decoding performance by introducing temporal gaps

181 Since each frame of the movie shares a high degree of similarity with neighboring frames, we controlled for the  
182 temporal correlations in the annotated features induced by the continuous nature of the stimulus. We divided the  
183 dataset into training, validation, and test sets, ensuring a gap of 32 s between samples from different sets to minimize  
184 temporal correlations (split visualized in Supp. Fig. S7). We investigated the impact of these gaps by decoding the  
185 character Summer using varying gap lengths while keeping the number of samples comparable. We observed that  
186 smaller gaps result in substantially higher decoding performance on the test set, raising concerns about potential data  
187 leakage between training and test sets. For instance, a random split without any temporal gaps achieves an almost  
188 perfect score of  $0.99 \pm 0.0005$ . However, as temporal gap sizes increase to 32 s, the performance drops precipitously  
189 to  $0.31 \pm 0.06$  (Fig. 3e; additional metrics in Supp. Fig. S9). This might explain the higher decoding performance  
190 for a comparable task in Zhang et al.<sup>24</sup>, which did not report the use of temporal gaps for model evaluation. All  
191 subsequently reported results refer to the performance on the held-out test data using 5-fold cross-validation, with  
192 data splits incorporating the most conservative temporal gap of 32 s (see Methods). Our analysis underscores the  
193 importance of appropriate architecture selection and careful data preparation in complex datasets such as ours, as  
194 these choices can exert a significant impact on the results.

## 195 Patient-wise decoding performance

196 The neuronal population analyzed thus far has been pooled from 29 patients, yielding a total of 2286 neurons. De-  
197 coding from a pooled population, rather than from individual patients, improves network stability by aggregating  
198 activity across a larger neuronal set and enhances the signal-to-noise ratio. However, using this pooled population  
199 (or "pseudo-brain") obscures the patient-wise contributions to decoding, which could vary due to the difference in  
200 neurons recorded per patient (units per patient range from 30 to 137) or due to differences in the semantic space  
201 of recorded neurons. To test for such differences, we assessed decoding performance on a per-patient basis, and  
202 analyzed each participant's neuronal population to see if key decoding information is widely distributed or driven by  
203 a particular subset.



**Figure 4. Patient-specific decoding performance** Decoding performances for the main character (Summer), visual transitions (Scene Cuts), and location (Indoor/Outdoor) are reported using Cohen's Kappa and compared to the performance obtained from the total population (pooled across all patients, dashed line). **a)** Decoding performance based on the total population of 2286 units, with neurons pooled across all patients. **b-d)** Patient-specific decoding performances for Summer, Scene Cuts and Indoor/Outdoor.

Decoding performance was obtained for three label categories—Summer, Scene Cuts, and Indoor/Outdoor—representing characters, visual transitions, and locations. To minimize computational load, we retrained a simpler logistic regression model on each patient's neurons, and achieved performance comparable to a more complex recurrent neural network but with lower computational costs. The results are illustrated in Fig. 4. Generally, decoding performance was lower at the individual patient level compared to the pooled neuronal population, with no single patient matching the performance of the aggregated data. For the Summer label, we observed substantial variability in decoding performance across patients, with some patients showing near-zero accuracy. However, certain patients (specifically 7, 10, 15, and 22) achieved decoding performances exceeding 0.2, compared to the overall pooled performance of 0.28. This variability was less pronounced for the Scene Cuts and Indoor/Outdoor labels. For the Scene Cuts label, the already low pooled performance declined further in the per-patient analysis, with patients 2, 10, 19, and 20 showing slightly better results, while most demonstrated minimal decoding accuracy. The Indoor/Outdoor label elicited a consistently higher accuracy across patients, matching the overall higher decoding performance achieved with the pseudo-brain population. Notably, patients 2, 13, 20 and 25 achieved performances exceeding 0.2 (Cohen's Kappa), compared to an overall pooled performance of 0.31, indicating robust neuronal responses in patient-specific subpopulations of neurons. Across all labels, the highest-performing patients vary, and no single patient showed consistently superior performance across all three.

### Parahippocampal cortex drives decoding of visual transitions and location

Continuously presented stimuli offer a rich array of features. Visual transitions, such as changes in filming angle or scenery, are a commonly studied feature that demarcate the event structure of the dynamic stimulus. In movies, these transitions are relatively well-defined since they consist of identifiable changes in pixel values between frames and are known to elicit time-locked changes in neural activity in fMRI<sup>20</sup>, iEEG<sup>22</sup>, and single neurons<sup>11</sup>. We investigated two types of frame-wise visual transitions: Scene Cuts (changes in scenery)

254 and Camera Cuts (changes in filming angle). As Scene Cuts consists of visual transitions between locations or points  
255 in time and demarcate narrative episodes within the movie, they are related to location but not exclusively. We  
256 compared this label to a more straightforward location-related feature, Indoor/Outdoor, which indicates whether a  
257 given frame depicts an indoor environment or not. Examples of Scene versus Camera Cuts as well as Indoor/Outdoor  
258 scenes are shown in Supp. Fig. S2.

259 To investigate region-wise differences, we trained separate decoders for neurons in the amygdala (A), hippocampus (H),  
260 entorhinal cortex (EC), and parahippocampal cortex (PHC) of the MTL. We excluded the piriform cortex (PIC)  
261 due to its relatively lower number of recorded neurons (Fig. 1b). We observed a clear dominance of the parahippocampal  
262 cortex for both types of visual transitions. The decoding performances reached  $0.21 \pm 0.02$  for Scene Cuts  
263 and  $0.19 \pm 0.03$  for Camera Cuts, respectively, when restricting the decoding to the parahippocampal cortex as opposed to  
264  $0.20 \pm 0.03$  and  $0.18 \pm 0.01$  when decoding from the full population. The other regions showed a lower  
265 but above-chance decoding performance. Similarly, the parahippocampal cortex yielded the highest performance  
266 for predicting indoor versus outdoor content and reached a performance of  $0.30 \pm 0.06$ , comparable to the performance  
267 on the full population ( $0.31 \pm 0.06$ ). Hippocampus was the second strongest region with a high performance  
268 of  $0.26 \pm 0.04$ . For entorhinal cortex and amygdala, we observed lower performances of  $0.15 \pm 0.05$  and  $0.1 \pm 0.03$ ,  
269 respectively.

270 In summary, the parahippocampal cortex consistently achieved the highest decoding performance for labels associated  
271 with visual transitions and location, in line with prior research on the MTL<sup>11</sup>. Given that both Scene and Camera  
272 Cuts are linked to sharp visual transitions, we anticipated that earlier processing stages in the MTL would show better  
273 decoding performances than later processing stages. However, despite the clear dominance of the parahippocampal  
274 cortex, our results show that other regions achieve lower, but nonetheless significant performance when detecting  
275 event structure and setting information.

## 276 Amygdala drives decoding of character presence

277 The MTL carries information about the identity of specific individuals, in addition to general person-related categories  
278 or attribute, primarily through the tuning of individual neurons<sup>31-33</sup>. Unlike visual transitions, character identities are  
279 a semantic feature which rely on both visual attributes and higher-level abstract representations. To investigate  
280 character-driven representations at the population level, we analyzed neuronal activity during the presence of the  
281 movie's three main characters, Summer, Tom, and McKenzie, as well as the more general concept of any character appearance  
282 (Persons, see example frames in Supp. Fig. S2). While Summer's appearance throughout the movie frames  
283 is balanced (50/50), the remaining labels are highly imbalanced: Tom and Persons appear in the majority of frames  
284 (80/20 and 95/5), while McKenzie is predominantly absent (10/90) (Fig. 1c). Despite the imbalances, we observed significant  
285 decoding performances for all four character labels ranging between 0.23 and 0.36 (Fig. 3c). Notably, decoding  
286 performance for character identities—despite being abstract and variable—exceeded that of visual transitions (0.20  
287 and 0.18).

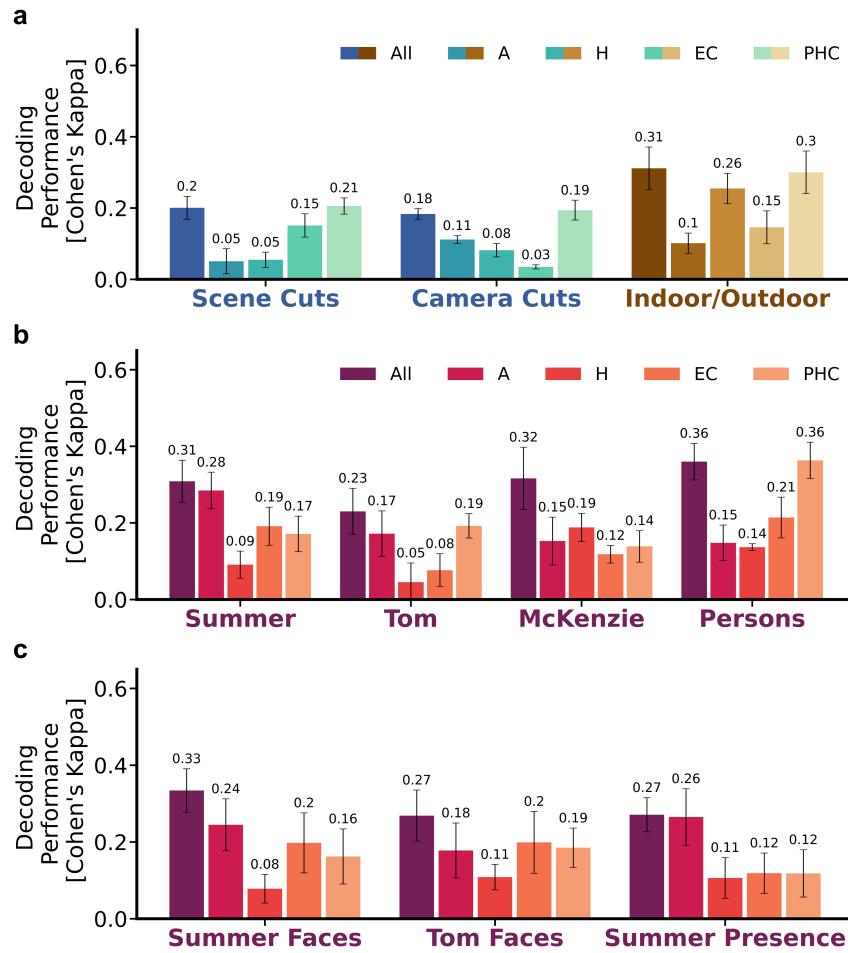
## 288 Distribution of information across MTL regions

289 To investigate whether characters were primarily processed in a specific MTL region or in all regions equally, we  
290 conducted a similar analysis as before by retraining on region-specific activity (Fig. 5a). Our results show that all four  
291 tested regions carry information about the character's identities, enabling decoding at above-chance levels ( $p < 0.001$ ,  
292 permutation test, see Methods). The amygdala and parahippocampal cortex showed the highest decoding performances  
293 for Summer and Tom, respectively, approaching levels similar to decoding from the full population. However,  
294 the distribution of information among the other regions was less consistent and varied across labels (Fig. 5a). The  
295 hippocampus had the lowest performance for Summer ( $0.09 \pm 0.04$ ) and Tom ( $0.05 \pm 0.05$ ), while the entorhinal cortex  
296 performed lowest for McKenzie ( $0.12 \pm 0.02$ ). For Persons, the parahippocampal cortex dominated with decoding  
297 performance comparable to the full population ( $0.36 \pm 0.05$ ), while other regions ranged between 0.14 and 0.21.

## 298 Differences in character's visual appearances

299 Since the labels indicate the presence of a given feature within a natural scene rather than a single exemplar shown  
300 in isolation, the visual appearance of the labeled entity varied substantially during the movie.

301 To better control for visual appearance and examine decoding differences 302 across various levels of character presence, we created the additional labels Summer Faces, Tom Faces, and 303 Summer Presence (see Methods for details on annotation creation). As with 304 the character labels, we trained a decoding network 305 on both the full neuronal population as well as the four individual regions (Fig. 5b). 306 Face labels for both characters elicited a slight improvement in performance 307 compared to the full neuronal populations, with Summer increasing from  $0.31 \pm 0.06$  to  $0.33 \pm 0.06$  and 308 Tom increasing from  $0.23 \pm 0.06$  to  $0.27 \pm 0.07$ . Regionally, the distribution 309 remained consistent with the general character labels, except for Tom Faces, 310 where the entorhinal cortex had the highest performance. The hippocampus 311 performed weakest for both Summer Faces ( $0.08 \pm 0.04$ ) and Tom Faces ( $0.11 \pm 0.03$ ). The Summer Presence label had an overall 312 performance of  $0.27 \pm 0.04$ , slightly lower than the general Summer label, with 313 the amygdala clearly dominating the regional distribution ( $0.26 \pm 0.07$ ). For face labels, distributions were more similar across regions than for 314 the general character labels. The amygdala outperformed other regions in decoding the abstract presence of Summer 315 but performed poorly for general person appearances, where the parahippocampal cortex performed best. These 316 findings align with previous research showing that semantically-tuned cells in the human MTL can flexibly activate 317 when their preferred conceptual category is indirectly invoked<sup>34</sup>.



**Figure 5. Semantic information is distributed differently across MTL regions based on category** Decoding performances for semantic features, by MTL region. All performances were significantly better than chance level with an alpha level of 0.001 (reported performance using Cohen's Kappa, mean performance across five different data splits, error bars indicate standard error of the mean. **a)** Decoding performances for visual transitions and location features. **b)** Character visibility could be decoded from the entire population of neurons, and with variable performance when training only on individual MTL regions. **c)** Decoding performances for face-specific character appearances and Presence features, by region.

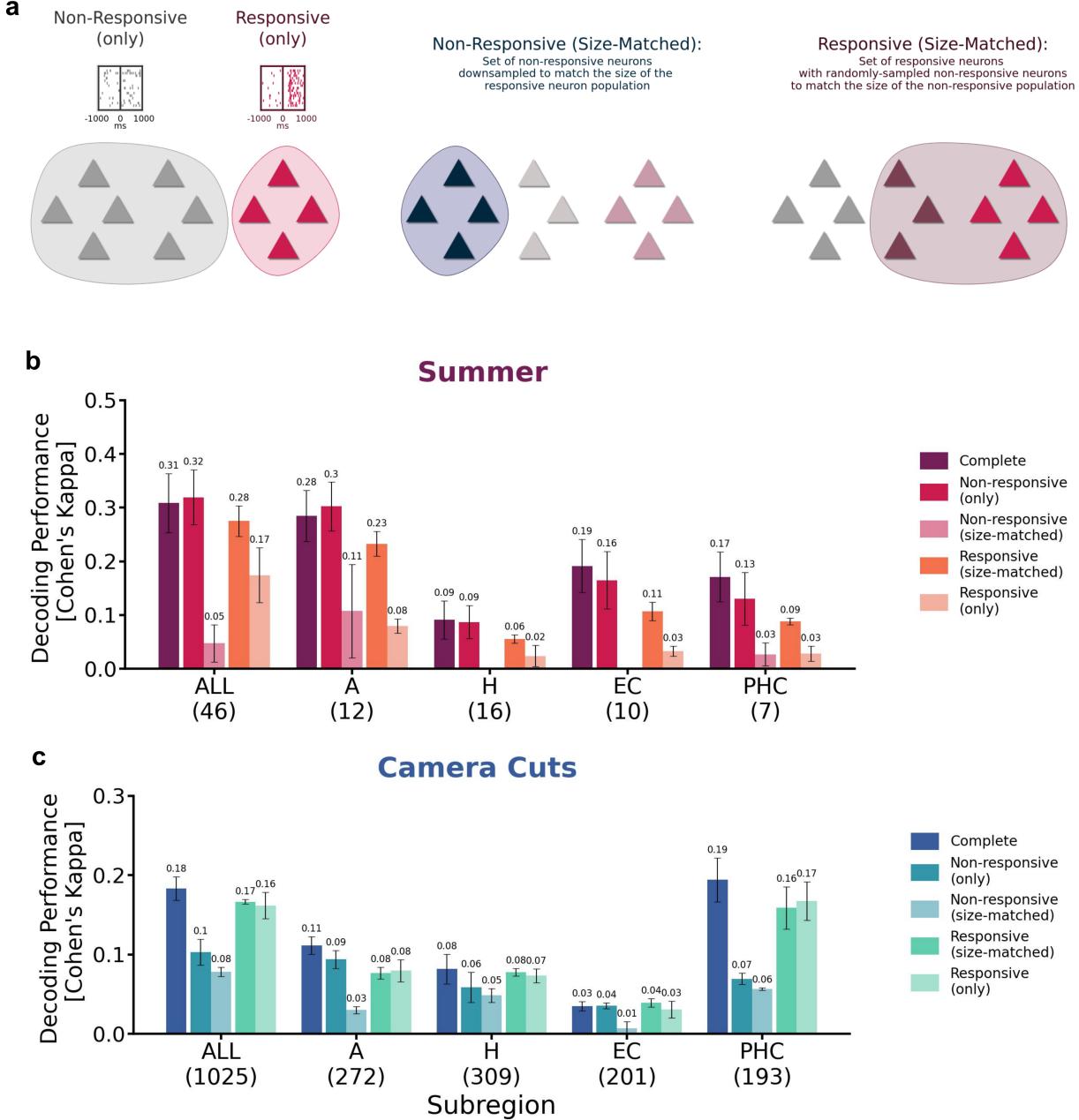
346 **Responsive neurons drive decoding of visual transitions but not decoding of characters**  
 347 Although only a subset of neurons modulated firing in response to the onset of a given movie feature, we nevertheless 348 observed significant decoding performance from the full population. This effect could result from two scenarios: 349 a) information is distributed throughout the neuronal population and decoding does not disproportionately rely on 350 neurons with post-onset increases in firing, or b) the subpopulation of responsive neurons informs the decoder while

351 non-responsive neurons are ignored. We tested each scenario by dividing the full population into two corresponding  
352 subsets—non-responsive and responsive neurons—and re-training a neural network on each subset. We analyzed  
353 the labels Summer and Camera Cuts, which represent the character-related and visual transition categories, respec-  
354 tively. We then compared the prediction performance of each re-trained model to that of the full population to  
355 determine if the decoding of a given label took the entire population into account (scenario a) or relied on responsive  
356 neurons (scenario b) (Fig. 6). The subsequent analysis evaluates these subsets both at the full population level and  
357 within the restricted context of MTL regions.

358 **Decoding with non-responsive versus responsive neurons**

359 First, we tested prediction performance using only the non-responsive neurons to determine if similar decoding could  
360 be obtained without neurons which significantly modulated firing after the onset of a feature. For this subpopula-  
361 tion (*Non-responsive (only)*), a separate LSTM was retrained and tested. Despite the exclusion of responsive neurons,  
362 these subpopulations yielded performances comparable to those of the complete population (Fig. 6a, *Complete*) for  
363 the character Summer. Minimal differences in performance were observed across MTL regions, with only the en-  
364 torhinal and parahippocampal cortex showing qualitative decreases. For comparison, we additionally retrained using  
365 only responsive neurons as input, and again tested the decoding performance (*Responsive (only)*). In contrast to the  
366 minimal differences observed in the *Non-responsive (only)* model, restricting to only responsive neurons produced  
367 a decrease in overall performance across the entire MTL and all regions. This general decrease suggests that the  
368 subset of individually responsive neurons is not the primary driver of the decoding performance observed in the full  
369 population model. Note that the total number of responsive neurons was less than the total non-responsive, so this  
370 effect may be influenced by an overall decrease in neuronal data. This difference in totals is directly addressed below  
371 by a size-matching procedure.

372 A different pattern emerged for the Camera Cuts feature, which exhibited the greatest performance drop when  
373 responsive neurons were excluded, for the entire MTL and for all subregions but the entorhinal cortex (Fig. 6b). This  
374 pattern was most pronounced for the parahippocampal cortex, indicating that responsive neurons carry valuable  
375 information for processing Camera Cuts in the movie. This hypothesis is further supported by the decoding perfor-  
376 mances obtained when restricting the decoding to only responsive neurons. Despite the restricted number of input  
377 neurons, the performance dropped marginally, and remained comparable to that of the complete population for the  
378 entire MTL, as well as for the amygdala, hippocampus, and entorhinal cortex regions.



**Figure 6. Responsive neurons drive performance for visual transitions, but not characters.** To assess the contribution of the responsive neurons on decoding (identified in *Stimulus-aligned responsive neurons found primarily in parahippocampal cortex*), we compared the decoding performance for subpopulations which did or did not contain these neurons. **a)** Illustration of neuronal sets used in the decoding comparisons (triangles represent neurons). An example of the complete population is shown in the left-most section, which depicts *Non-responsive (only)* and *Responsive (only)* cells, with a simulated example of a respective peri-stimulus time histogram (onset raster, grey and magenta). A *Non-responsive (size-matched)* group (middle section) was randomly subsampled from the total population to have a size-matched comparison to the total set of responsive neurons. The *Responsive (size-matched)* set (right-most section) consisted of all responsive neurons padded with randomly selected non-responsive neurons to match the total *Non-responsive (only)* population. **b,c)** Decoding performances for discussed subpopulations for the character label Summer and Camera Cuts. Number of responsive neurons for the respective subpopulation reported in parentheses.

**379 Size-matched neuronal populations: comparing responsive and non-responsive subsets**

**380** Since the total number of responsive neurons was lower than that of non-responsive neurons, we tested the per-  
**381** formance using size-matched versions of both non-responsive and responsive subpopulations. To match the size  
**382** of the responsive neurons, we randomly selected an equivalent number of non-responsive neurons (*Non-responsive*  
**383** (*size-matched*)) and trained and tested a separate neural network. This process was repeated three times with dif-  
**384** ferent random selections, and we report the average performance. For both labels, Summer and Camera Cuts, the  
**385** smaller size-matched non-responsive subpopulation showed an expected decrease in performance compared to the  
**386** full non-responsive subpopulation. However, the results diverged when comparing the size-matched populations:  
**387** For Summer, the size-matched non-responsive neurons performed comparably to the responsive neurons within  
**388** individual MTL regions. Only for the complete population did the responsive neurons show a clear improvement.  
**389** Conversely, for the Camera Cuts, restricting to only the responsive neurons improved performances, with decoding  
**390** predictions of all but the amygdala surpassing those of the size-matched non-responsive set of neurons.

**391** A similar pattern emerged for a size-matched version of the responsive neurons (*Responsive (size-matched)*), which  
**392** we formed within region by padding the set of responsive neurons with randomly selected non-responsive neurons.  
**393** For Summer, decoding from this subpopulation consistently showed a performance drop compared to the total set  
**394** of non-responsive neurons across all tested regions. In contrast, for Camera Cuts, the size-matched subpopulation of  
**395** responsive neurons achieved similar or better performance than the non-responsive neurons across all regions, with  
**396** the complete population and parahippocampal region showing a clear dominance of the subpopulation containing  
**397** responsive neurons. We additionally evaluated performances for the labels Tom, Scene Cuts and Indoor/Outdoor,  
**398** which matched the effects found for Summer and Camera Cuts (Supp. Fig. S10).

**399**

**400** In summary, our findings indicate that responsive neurons play distinct roles for different features. Neurons respon-  
**401** sive to visual transitions appeared to carry information not equally present in other neurons. On the contrary, for  
**402** character- and location-related labels, individual responsive neurons contributed less to decoding, and information  
**403** appeared to be distributed either across the entire population or a subset of neurons distinct from the previously  
**404** identified responsive neurons.

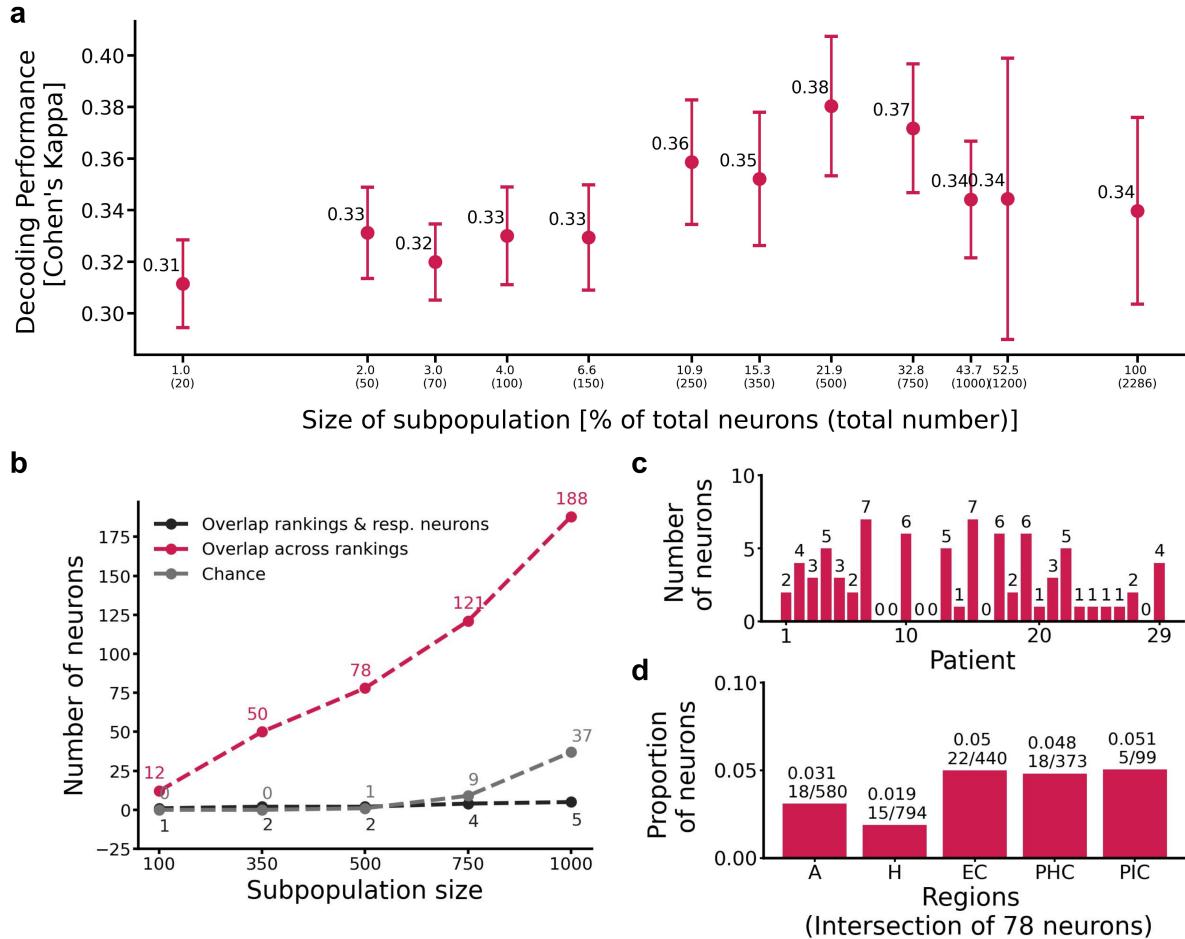
**405 Relevant information is carried by a smaller subpopulation of 500 neurons**

**406** We observed that subsets of neurons with stimulus-selective responses to characters did not account for the decoding  
**407** performance of the same character. Previous research in sensory information processing suggests that relevant stim-  
**408** ulus information is often encoded by only a subset of neurons within a population<sup>27,35,36</sup>. To explore this further, we  
**409** adopted a more data-driven approach to define neuronal subsets, ranking their importance using weights extracted  
**410** from a trained logistic regression model and investigated the minimally sufficient number of neurons required for  
**411** successful decoding.

**412 Ranking of neurons for the character label Summer**

**413** The weights of a logistic regression model can be used to assess each neuron's importance in decoding, allowing for  
**414** the creation of a ranking across all neurons (see Methods for more details). In contrast to an LSTM, logistic regression  
**415** models are computationally less expensive to train and achieve comparable decoding results for all movie features,  
**416** except for Scene Cuts and Camera Cuts, which exhibited reduced but above-chance decoding performance (Fig. 3c).  
**417** For the character label "Summer," we generated a neuron ranking from the trained logistic regression model and  
**418** defined subsets of neurons by selecting those with the highest rankings.

**419** The above LSTM and logistic regression models were trained on population data using 5-fold cross-validation, with  
**420** alternating test sets for each of the five splits. However, this approach precludes an independent ranking of neurons  
**421** across splits, as test data from one split may overlap with training data from another. To address this, we expanded  
**422** the five original splits to 20, where each subgroup of four splits shared a common test set but varied in the allocation  
**423** of training and validation data. Any analysis relying on a subset of neurons derived from logistic regression weights  
**424** was exclusively assessed using the four splits that produced the ranking and shared held-out test data (see more  
**425** details in Methods, and Supp. Fig. S8).



**Figure 7. 500 neurons are sufficient to reach peak decoding performance** **a)** Decoding performances for the character Summer for subpopulations of top-performing neurons, testing various sizes ranging from 1% to 100% of the full population (absolute numbers of neurons are reported in parenthesis). Mean performance across different splits is reported, and standard error of the mean is visualized by the error bars. **b)** Number of overlapping neurons across rankings for different sizes of subpopulations of top-performing neurons (pink). As a baseline, we compare the number of overlapping neurons to the number expected by chance (grey), and we observed a notably higher intersection of top-ranked neurons across the splits. Additionally, the overlap between the intersection of top-ranked neurons and the previously defined responsive neurons is shown (black). **c, d)** Overlapping neurons (in total 78) in subpopulations of 500 top-performing neurons for each ranking were distributed across patients and MTL regions.

**426 Training on pre-selections of top-ranked neurons**

**427** We trained on progressively smaller subpopulations of top-ranked neurons for the character Summer (Fig. 7a) and  
**428** observed an increase in performance when restricting the input activity to smaller populations (peak at 0.38 (Cohen's  
**429** Kappa) for 500 neurons, 21.9% of the total population). The decoding performance reported here for the entire  
**430** population shows a slight variation from the previously reported value of 0.31 due to the modified nested cross-  
**431** validation procedure. Further reduction of the population lead to a decrease in performance, yet high decoding  
**432** performance persisted even in small subpopulations of neurons.

**433** We observed that a ranking procedure which did not use a common test set was subject to potential cross-talk  
**434** between data splits, which substantially impacted and distorted the results. Using a selection of neurons derived  
**435** from non-independent training and test data led to a stark increase in performance, nearly doubling the original  
**436** performance of the character Summer to 0.52 (Cohen's Kappa) when restricting to a subpopulation of 150 top-ranked  
**437** neurons. This again underscores the need to carefully prepare the data for paradigms such as ours, where data

438 samples are highly correlated, as ignoring dependencies between training and test data can greatly skew results (see  
439 Supp. Fig. S13).

440 Top-ranked neurons and their distribution across patients and MTL regions

441 As the selection process involved five distinct rankings of neurons, we investigated the consistency of the neuronal  
442 composition across rankings. The overlap of neurons within subpopulations of top-ranked neurons assessed across  
443 different sizes is shown in Fig. 7b. Analyzing the top-performing 500 neurons from each of the five rankings revealed  
444 a common set of 78 neurons. We compared this observed overlap to that expected by chance with random subpop-  
445 ulation selections (see Methods), finding a notably higher overlap. This suggests the presence of common neurons  
446 crucial for decoding the label Summer. Additionally, we also compared these selected neurons to those previously  
447 classified as responsive. The overlap between the two groups of neurons was small, suggesting that the respon-  
448 sive neurons defined by onset-related changes in activity may not be the primary drivers of decoding performance in  
449 general, aligning with previous observations for character labels when the decoding pipeline was restricted to respon-  
450 sive neurons. As restricting to 500 neurons yielded the highest decoding performance, we subsequently analyzed the  
451 resulting intersection of 78 neurons across rankings. These 78 neurons were distributed across both patients and  
452 regions (Fig. 7c,d), with no single patient or area of the MTL driving the performance. Visualizations of the spiking ac-  
453 tivity surrounding the onset of Summer for these neurons do not reveal a clear pattern of stimulus-evoked increases  
454 in firing (Supp. Fig. S12).

455 Our findings reveal that a core subpopulation of approximately 500 neurons drives decoding performance, while  
456 additional neurons mainly contribute redundant or noisy information. These neurons are distributed across patients  
457 and MTL regions, presenting an important avenue for future research to investigate the mechanisms underlying their  
458 organization and the specific functional roles they play in decoding processes.

## 459 Discussion

460 We investigated how the human brain processes semantic and event structure in a naturalistic setting by analyzing  
461 the activity of neurons in the medial temporal lobe during the presentation of a full-length commercial movie. Al-  
462 though earlier work has established and characterized the role of MTL neurons in semantic representation, and the  
463 processing of dynamic stimuli via fMRI and intracranial electroencephalography, few studies have investigated how  
464 human single neurons process dynamic stimuli and none have addressed the relationship between representations  
465 on the single-neuron level and the population level.

466 By analyzing changes in each neuron's activity aligned with the onset of labeled movie features, we identified  
467 groups of individual cells that adjust their firing rates in response to specific features. The most pronounced re-  
468 sponses occurred following changes in camera angles and scenes, as these features induced activity changes in the  
469 largest number of cells. Outdoor scenes and the two main characters also elicited consistent single-unit responses, al-  
470beit in far fewer neurons. This lack of explicit single-neuron responses to characters, which might otherwise suggest  
471 selective and invariant representations such as those in concept neurons, could be explained by the study partici-  
472 pants' lack of familiarity with the movie. Most participants had neither seen the movie before nor encountered the  
473 actors in other media prior to its presentation as part of this study. This interpretation aligns with previous research  
474 showing that neuronal selectivity to individuals varies with familiarity and personal relevance, as photos depicting  
475 personally known individuals are more likely to elicit selective responses in MTL neurons<sup>37</sup>.

476 We anticipated that the decoding performance for each label would reflect the pattern observed at the single-  
477 neuron level: Visual transitions would be the most accurately predicted feature, followed by setting, with character  
478 presence showing the lowest prediction accuracy. Although not every feature elicited explicit responses from a signif-  
479 icant portion of individual neurons, the network, which takes the collective population activity as input, successfully  
480 decoded all tested features with above-chance performance. Interestingly, the decoding performance varied across  
481 features and contradicted the pattern found in the individual neurons. Despite eliciting the highest proportion of  
482 responses at the single-unit level, visual transitions showed the lowest decoding accuracy out of all tested features.  
483 Conversely, characters showed the highest decoding accuracy despite there being few individually responsive neu-  
484 rons. Both approaches link neuronal responses to the movie content but differ markedly in their focus, as the single-  
485 unit analysis targets the specific onset of features, emphasizing their initial activation, while the population approach  
486 processes data continuously, decoding both the onset and sustained presence of features. This methodological dif-  
487 ference may affect the results and contribute to the observed contradiction between them. To bridge these findings,  
488 we examined how individual responsive neurons contribute to the decoding network. We hypothesized that the de-  
489 coding performance for each label would be primarily driven by the subset of neurons that exhibit increased activity  
490 in response to that label. This hypothesis held true for the decoding of visual transitions, as the set of individually  
491 responsive neurons disproportionately contributed to the decoding performance when using the full population, and  
492 performance was strongly affected by their removal from the training population. In contrast, character decoding  
493 does not rely on individually responsive cells, as removing neurons that responded strongly to character onset had  
494 little impact on performance. When analyzing the network and its prediction behavior, we identified a subset of units  
495 which contributed most to the decoding of character information across models, and found that these units had little  
496 overlap with the set of neurons which increased firing after character onset. Together, these findings suggest that  
497 character-related representations relied on a population code, while visual transitions were encoded by the activity  
498 of specific neurons.

499 When training separately on different regions of the MTL, we observed variations in decoding performance de-  
500 pending on the specific content being decoded. The parahippocampal cortex achieved the highest performance in  
501 detecting visual transitions, while the amygdala performed best in predicting character-related information. These  
502 results support previous findings which identified that certain regions are more likely to respond to certain categories  
503 of images in static screenings. For example, previous studies have shown that the amygdala preferentially responds  
504 to images containing faces<sup>1,32</sup>, and contains cells that selectively respond to whole faces as opposed to discrete facial  
505 features<sup>38</sup>. We extended this analysis to examine different levels of character appearance, distinguishing between  
506 face visibility and the general presence of a character in a scene (regardless of face visibility). Although, the amygdala  
507 demonstrated strong performance in both cases, it most clearly drove the decoding for the general presence of a  
508 character, rather than specifically to face visibility. This effect could be due again to the unfamiliarity of the movie

509 and its actors, as face-specific responses have been shown to form as a function of exposure<sup>39</sup>.

510 Previous work has found that the parahippocampal cortex is especially sensitive to scene information<sup>40</sup>, as op-  
511 posed to objects<sup>40,41</sup>, with a higher likelihood of neuronal activation when scenes feature stronger spatial layout cues,  
512 such as depth and a recognizable background<sup>42</sup>. Zheng et al. (2022) identified generalized cells, termed "boundary  
513 cells", in the parahippocampal gyrus, hippocampus, and amygdala which modulate their firing after any visual trans-  
514 sition event. In our study, we observed significant responses to camera cuts in these same regions, which we inter-  
515 pret as analogous to "soft boundaries"<sup>11</sup>. However, significant responses to scene cuts were only observed in the  
516 parahippocampal cortex, whereas Zheng et al. reported responses to their analogous feature ("hard boundaries")  
517 in all measured MTL regions. In addition to their preference for scene-related images, parahippocampal neurons  
518 have been shown to respond more frequently to outdoor images compared to indoor ones<sup>42</sup>. In our dataset, a small  
519 subset of parahippocampal neurons increased firing in response to the onset of outdoor scenes, whereas none  
520 showed increased activity for indoor scene onset. Despite their limited number, these responsive neurons achieved  
521 well above-chance decoding performance, albeit falling short of the performance achieved by the complete neuronal  
522 population. Further work is needed to more accurately determine if the activity of single scene-selective parahip-  
523 pocampal neurons in our dataset can be explained by the onset of location-related content.

524 Visual transitions and character content additionally differed in their use of sequence information. Through a  
525 comparison between decoding architectures, we found that temporal information in the spike trains only mattered  
526 for the prediction of visual transitions, and that ignoring or even scrambling the sequence information had little effect  
527 on character and location features. Although using temporal dynamics improved decoding performance for visual  
528 transitions, the temporal dynamics in our data, particularly those inherent to the movie stimulus, also present a sig-  
529 nificant confounding factor. Both the movie and recorded neuronal activity are subject to a high degree of correlation  
530 in time and thus require extra consideration when formulating a pipeline to ensure that the training and test data did  
531 not contain adjacent, highly correlated samples. The correlation between training and test samples artificially inflates  
532 decoding performance, and does not reflect actual generalization to unseen data. In related research by Zhang et al.  
533 (2023), where temporal distance was not considered, high decoding performances were reported on a similar task.  
534 Based on our analyses, we anticipate that their reported decoding accuracy is overestimated, and would change if  
535 sufficiently large gaps were introduced.

536 As our dataset consists of single-neuron activity pooled across 29 subjects, each with activity from an average of  
537 approximately 80 recorded neurons, the extent to which claims can be made about what an individual brain does  
538 is limited. In addition, the participants watched a full-length movie in an clinical setting, where neuronal activity is  
539 not solely focused on visual stimuli and likely processes additional information. Although we cannot know whether  
540 the neuronal population that we sampled is representative of the human MTL generally, it is one of the largest sam-  
541 ples collected to date, both in terms of neurons and patients and offers a unique opportunity to understand the  
542 processing and representation of information in the MTL population activity. Despite the inherent limitations, which  
543 preclude near-perfect performances in the decoding task, the presented approach demonstrates that movie con-  
544 tent can nevertheless be successfully decoded from such a sub-sampled population of neurons. Future work on this  
545 dataset could leverage more advanced network architectures to explicitly model between-neuron dynamics within  
546 each patient, which could better explain the gain in performance achieved moving from the single-neuron level to  
547 the population level. Additionally, in this work we focused specifically on the *visual* content of the movie. A clear next  
548 direction would be to integrate the auditory information of the movie, and possibly disentangle the contribution of  
549 visual versus audio information streams to the neuronal representation of movie features in the MTL.

## 550 Materials and Methods

### 551 Participants and recording

552 The study was approved by the Medical Institutional Review Board of the University of Bonn (accession number  
553 095/10 for single-unit recordings in humans in general and 243/11 for the current paradigm) and adhered to the  
554 guidelines of the Declaration of Helsinki. Each patient gave informed written consent for the implantation of micro-  
555 wires and for participation in the experiment.

556 We recorded from 46 patients with pharmacologically intractable epilepsy (ages 19 - 62, median age 37; 26 fe-  
557 male patients). Patients were implanted with depth electrodes<sup>31</sup> for locating the seizure onset zone for potential  
558 later resection. Micro-wire electrodes (AdTech, Racine, WI) were implanted inside the shaft of each depth electrode.  
559 Signal from the micro-wires was amplified using a Neuralynx ATLAS system (Bozeman, MT), filtered between 0.1 Hz  
560 and 9000 Hz, and sampled at a rate of 32 kHz. Spike sorting was performed via Combinato<sup>43</sup> using the default pa-  
561 rameters and the removal of recording artifacts such as duplicated spikes and signal interference was performed via  
562 the Duplicate Event Removal package<sup>44</sup>. After all data were processed, neuronal signals, experimental variables, and  
563 movie annotations were uploaded to a tailored version of Epiphyte<sup>45</sup> for analysis. Due to disruptions in the movie  
564 playback caused by clinical interruptions, 13 patient sessions were excluded from further analysis.

### 565 Task and stimuli

566 Patients were shown a German dubbing of the commercial movie *500 Days of Summer* (2009) in its entirety (83 min-  
567 utes). This film was chosen because the actors portraying the main characters were relatively unfamiliar to a general  
568 German audience at the time of the initial recordings. The movie was shown in an uncontrolled clinical setting, where  
569 neither gaze nor attention were directly monitored and was presented in letterbox format without subtitles on a lap-  
570 top using a modified version of the open-source Linux package, *FFmpeg*<sup>46</sup>, with a frame rate of 25 frames per second.  
571 Due to the length of the movie and the possibility of clinical interruptions, patients and staff were allowed to freely  
572 pause the playback. Discontinuity in playback was controlled for within Epiphyte<sup>45</sup>. Pauses and skips in the movie  
573 playback were identified through the output of the modified *FFmpeg* program and used as a basis of exclusion for  
574 patients. Patients were excluded if they did not watch the entire movie, or watched the movie discontinuously.

### 575 Movie annotations

576 In order to relate the content of the movie to the recorded neuronal activity, we labeled various features on a frame-  
577 by-frame basis. These labels are binary and cover the following features:

- 578 • **Main characters:** *Summer, Tom, McKenzie*

579 Frames were labeled as positive if the character could be clearly distinguished by either appearance or context.  
580 Characters and persons were considered only on a visual basis (i.e., a frame in which Tom is speaking but not  
581 visible is labeled as not containing Tom).

- 582 • **Faces:** *Summer, Tom*

583 Instances of a character's face. Positive samples are frames where the character's face is shown, while negative  
584 samples are frames where the character's face is not visible at all. All other frames are excluded.

- 585 • **Presence:** *Summer*

586 Indicates the character's general presence in the scene, even if the character is not visible in the frame. For  
587 instance, frames are labeled as positive if the character is part of the scene but is not visible in that particular  
588 frame due to factors like the camera angle.

- 589 • **Visual transitions:** *Camera Cuts, Scene Cuts*

590 Marks visual transitions in the movie. Scene Cuts correspond to changes in scenery, while Camera Cuts are  
591 primarily based on changes in the visual stimuli.

- 592 • **Location:** *Inside/Outside*

593 Distinguishes between indoor and outdoor locations in the movie. Scenes that do not clearly fit into either  
594 category are excluded from the annotation.

- 595 • **Persons**

596 General appearance of any person(s)

597 Main character, Presence, Persons, Location, and Scene Cut labels were obtained manually using the open-source  
598 annotation program *Advene*<sup>47</sup>. For face labels of the characters, we developed a deep-learning pipeline for face detec-  
599 tion and classification. As backbone we used a pre-trained neural network for face detection and feature extraction<sup>48</sup>.  
600 We extended the pipeline by a classification network consisting of fully-connected layers combined with ReLU activa-  
601 tion functions. The classification network was fine-tuned on the movie frames to classify detected faces of the main  
602 characters, including a "not known" class for faces not belonging to the main characters. In the fine-tuning process,  
603 we used the manually created character labels for the characters. Camera Cuts were labeled automatically using the  
604 open-source algorithm *PySceneDetect*<sup>49</sup>, run with default parameters and manually reviewed. Camera Cuts mark a  
605 cut in the movie by labeling the first frame after cut onset as positive, resulting in cut events associated with a single  
606 frame. To adjust for the temporal latency in brain activity, cuts in the movie were associated with frames occurring  
607 within 520 ms of the cut onset. This adjustment smooths the cut labels, rendering them more comparable to what  
608 we anticipate in neuronal responses.

#### 609 **Calculation of single-neuron response statistics**

610 The "baseline" period was defined as 1000 ms prior to the onset (e.g., the entry of a character into frame) and the  
611 "stimulus" period was 1000 ms after the onset or appearance. Pseudo-trials with baseline periods containing frames  
612 depicting the label of interest were excluded. Responsive neurons were identified using a modified bin-wise signed-  
613 rank test<sup>8</sup>. The spiking activity across pseudo-trials was aligned to the stimulus onset. The baseline period was  
614 binned by 100 ms, and the normalized firing rate of the baseline period was compared to nineteen overlapping 100  
615 ms bins defined across the stimulus period using a Wilcoxon signed-rank test (alpha = 0.01, SciPy wilcoxon<sup>50</sup>, Simes  
616 corrected<sup>51</sup>). Additionally, a neuron was required to have spiked during at least one third of the total pseudo-trials  
617 to be tested (otherwise, assigned  $p = 1$ ).

#### 618 **Cluster permutation test**

619 A cluster permutation test<sup>26</sup> was used to test the difference in firing rates between the responsive and non-responsive  
620 subsets of neurons. Using the calculated response statistics, neurons were divided into two conditions: responsive  
621 ( $p \leq 0.01$ ) and non-responsive ( $p > 0.01$ ) for a given label. Activity for a neuron was averaged across bins, yielding a  
622 single vector of mean spike counts (spikes / 100 ms) spanning both baseline and stimulus periods for each neuron.  
623 This vector was then z-scored relative to its own mean and standard deviation. Mean spike count vectors were com-  
624 bined across conditions, yielding two datasets:  $A_{resp}$  and  $A_{non}$ , matrices containing the summarized, binned activity  
625 for all responsive and non-responsive neurons, respectively. A bin-wise comparison between  $A_{resp}$  and  $A_{non}$  was per-  
626 formed using a two-sided t-test for independent samples (ttest\_ind, SciPy), producing a t-stat and p-value for each  
627 bin. Clusters were defined as temporally adjacent bins with  $p \leq 0.005$  and t-stats were summed within clusters. The  
628 procedure was adapted to allow testing of multiple clusters, so no clusters were excluded at this stage. One set of  
629 1000 permutations were performed by randomly assigning neurons to  $A_{resp}$  and  $A_{non}$  such that the total number of  
630 neurons in each group was conserved and the bin-wise testing procedure was performed on each permuted dataset.  
631 Each cluster was compared to the resulting histogram of permuted t-stats. P-values for each cluster were defined as  
632 the number of permutations with a higher summed t-stat relative to the total number of permutations, and a p-value  
633 less than 0.05 was considered significant.

#### 634 **Decoding from population responses**

635 For decoding, we used population responses (input) to predict the corresponding concept labels (output). We ex-  
636 cluded the credits from the dataset and focused solely on the narrative content. The movie was presented at a frame  
637 rate of 25 frames per second, with each frame lasting 40 ms. In total, 125,743 frames were shown. The spiking activity  
638 of the recorded neurons was binned using a bin size of 80 ms, corresponding to the activity of two consecutive frames.  
639 Each bin was then labeled based on the first frame within that interval.

640 We sampled activity of the neurons before and after the onset of each frame and found that using a window of 800  
641 ms before and 800 ms after onset yielded the best decoding performance. This resulted in data samples comprising  
642 a binary label for the concept and a spike train of length 20 (10 bins before and after the onset of the frame). Given  
643 the full population of 2286 recorded neurons, each input data sample had a dimensionality of 2286.

## 644 **Architecture**

645 To decode concepts from the sequence of neuronal activity, we used a long short-term memory (LSTM) network<sup>30</sup>,  
646 which is well-suited to process the dynamical structure of the dataset. The output of the LSTM was fed into several  
647 fully-connected layers with ReLU activation functions to obtain binary label predictions. We found that pre-processing  
648 the raw spiking data with a linear layer of same size as the input, combined with a batch normalization layer, improved  
649 performance. We used the binary cross-entropy loss function, and optimized our network using Adam optimizer<sup>52</sup> in  
650 Pytorch<sup>53</sup> with default settings (first and second order moment equal to 0.9 and 0.999, respectively). We obtained the  
651 best results with a 2-layer LSTM, with hidden size of 32. We adapted other hyperparameters, e.g., number of linear  
652 layers, hidden sizes, batch size, learning rate, weight decay, and dropout rate, for each label.

653 We trained each network for 700 epochs and used the validation set to estimate the model's ability to generalize to  
654 unseen data. As is common practice, we selected the model with the best performance on the validation set and used  
655 this for evaluation on the unseen test data. Some labels were highly imbalanced and models trained on these labels  
656 were biased towards predicting the majority class. To ensure unbiased predictions and to optimize performance, it  
657 is common to force the batches of data samples presented in the training process to be balanced by oversampling  
658 the minority class. This oversampling technique was applied to all imbalanced labels, comprising all labels except the  
659 label for the main character Summer.

## 660 **Data split**

661 To ensure that the decoding performance was not affected by correlations between data samples, we carefully split  
662 the dataset into training, validation, and test sets. We used 5-fold nested cross-validation, and assigned 70% of the  
663 data to training, 15% to validation, and 15% to testing in each split. To avoid correlations between samples, we as-  
664 signed samples from consecutive segments of the movie to each set (train/val/test) and excluded 32 seconds of  
665 data between each set (see Fig. S7). The choice of excluding 32 s was based on the results for the main character  
666 Summer (Fig. 3c). The total dataset contained 45,200 samples, resulting in sets of 30,800/7,200/7,200 samples for  
667 training/validation/test. We trained a network using the training data, optimized its hyperparameters using the vali-  
668 dation set, and evaluated its performance on the corresponding test set. We report the final decoding performance  
669 as the average performance on all five test sets.

## 670 **Evaluation metrics**

671 For each semantic feature, we compared the model's prediction against the binary class labels. While accuracy is a  
672 simple evaluation metric, it is not suitable in our case due to the highly imbalanced distribution of most labels, making  
673 it challenging to compare accuracy metrics across different labels. We report all decoding performances using the  
674 Cohen's Kappa metric<sup>54</sup> which measures the agreement between the ground-truth labels and the predictions of the  
675 network, where performance equal to zero is interpreted as chance-level and performance equal to one is interpreted  
676 as complete agreement. Cohen's Kappa is defined as

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (1)$$

677 where  $p_0$  defines the relative observed agreement and  $p_e$  the hypothetical probability of chance agreement among  
678 prediction and labels. The Cohen's Kappa metric can take negative values, too, implying that the predictions are  
679 worse than chance level. The common chance performance equal to zero makes the Cohen's Kappa a useful metric  
680 to compare performances across all labels.

681 Additionally, we report the F1 Score, Area under the Precision Recall Curve (PR-AUC) and Area Under the Receiver  
682 Operating Characteristic (AUROC) metric in the Supplementary for all experiments. We briefly explain these metrics,  
683 including potential advantages and disadvantages for our analysis:

### 684 **F1 Score**

685 F1 Score is a metric combining precision and recall by calculating the harmonic mean between these two. Precision  
686 and recall are determined based on a classification threshold of 0.5. Designed to perform well on imbalanced datasets,  
687 the F1 Score is particularly useful for evaluating our decoding tasks. However, the baseline for chance performance  
688 with this metric is not consistent and varies depending on the label distribution.

689 PR-AUC

690 The PR-AUC metric, representing the area under the precision-recall curve, extends the F1 Score by evaluating perfor-  
691 mance across different threshold settings. Similarly to the F1 Score, it can be used for imbalanced datasets. However,  
692 as for the F1 Score, the PR-AUC metric is sensitive to changes in the class distributions, resulting in varying chance  
693 performance across labels. Being a reliable performance metric in general for our various classification problems,  
694 one has to bear in mind that a comparison of performances across concepts can be misleading due to the different  
695 chance baselines.

696 AUROC

697 The Receiver Operating Characteristics (ROC) curve represents the trade-off between the true positive rate and false  
698 positive rate for various threshold settings. AUROC as the area under the ROC curve is a performance measure that  
699 is used in settings where one equally cares about positive and negative classes. Performances range in [0, 1] and it is  
700 insensitive to changes in the class distribution, which means chance performance is given by a value of 0.5. However,  
701 the metric is generally not used for highly imbalanced classification problems and an evaluation of specific labels in  
702 our analysis such as McKenzie (class distribution is 90/10) should be taken with caution.

703 **Single-neuron decoding performance**

704 For a fair assessment of decoding from population responses compared to single-neuron activity, we evaluated the  
705 decoding performance of individual neurons under a setup comparable to that of the LSTM-based decoding network.  
706 Instead of relying on full-population responses, we established a performance based solely on the firing rates of each  
707 single neuron. Employing the same data splits, cross-validation approach, and binning procedure used for the LSTM  
708 network, we summed the binned firing rates of a neuron surrounding frame onset for the 1600 ms time window  
709 utilized by the decoding network. We then individually selected the best threshold for each neuron's activity on  
710 the validation set, which was subsequently used to evaluate the neuron's performance on the hold-out test set. We  
711 reported the final performance as the average of the five results obtained from the 5-fold cross-validation procedure.

712 **Permutation tests for decoding**

713 To determine if the reported decoding performances were significantly better than chance, we performed two sets  
714 of permutation tests—first, we randomly shuffled the labels of the held-out test set (Test Set Shuffle), and second,  
715 we shifted the labels while preserving the order of the test labels (Circular Shift). For both tests, the input to the  
716 decoding network remained unchanged from the non-permuted version. The only modifications made were to the  
717 corresponding feature label in the test set, which were changed in the ways outlined above, and then compared to  
718 the original network's prediction scores. The dynamic nature of the visual stimuli implies not only a strong correla-  
719 tion within the neuronal activity but also suggests a temporal correlation for the feature labels. By modifying only the  
720 labels, we ensured that the temporal information embedded in the neuronal data remained unaffected by the per-  
721 mutation test. In the following, we test the significance of both scenarios: one where the temporal correlation within  
722 the labels is disrupted (Test Set Shuffle), and another where the temporal correlation within the labels is maintained  
723 (Circular Shift).

724 **Test set shuffle**

The first type of permutation, and the one reported in the main text, consisted of randomly shuffling labels in the test  
set and evaluating the predictions of the models on those. We compared the performance of the model on the held-  
out test set to a null distribution generated by evaluating the model on the test set with shuffled labels ( $N = 1000$ ).  
We calculated the probability of our observed performance under the null distribution to obtain the p-value:

$$p = \frac{k + 1}{N + 1}$$

725 where  $k$  is the number of performances on permuted data outperforming the ground-truth performance of the model.  
726 This p-value provided the basis of comparison for describing significant decoding results in our main analyses. By  
727 preserving the temporal structure during the model's inference process and only disrupting temporal correlations  
728 within the concept labels used for evaluating the model's predictions, we consider this assessment of significance to  
729 be the most appropriate for our data setup.

730 Circular shift

731 The second type of permutation was performed as a comparison for the above *Test Set Shuffle*, as it is a standard  
732 method in human single-neuron studies (see<sup>55, 56, 44</sup>). Circular shift permutations maintain both the temporal rela-  
733 tionship of neuronal data as well as that of the stimulus information, here the concept labels. Rather than randomly  
734 shuffling the labels in the test set, we applied a circular shift of labels that maintained the intrinsic temporal structure.  
735 The shifting size was randomly chosen  $N = 1000$  times to obtain a null distribution akin to the previous test. This  
736 distribution was then utilized to compute the probability of our ground-truth performance and derive the p-value  
737 for assessing significance. For studies involving static stimulus presentation, wherein the stimulus is largely uncor-  
738 related with itself, this test provides a useful way to disentangle stimulus-related effects from those endemic to the  
739 time-series information. For a comparison between the permutation results, see Supplementary Materials, Fig. S11.

740 **Impact of temporal information on the decoding**

741 To assess the significance of temporal information in neuronal activity sequences, we evaluated the trained models  
742 using temporally-altered test data. The input to the decoding network included neuronal activity from 800 ms before  
743 to 800 ms after the onset of the frame, divided into 20 bins. For the temporally-altered test data, we randomly shuffled  
744 the sequence order of the 20 bins (applying a consistent permutation for all neurons and data samples) and evaluated  
745 the pre-trained models on the modified test data. This procedure was repeated 100 times, and the performance was  
746 averaged. The final performance was calculated by averaging the results across the five data splits, using the standard  
747 error of the mean (SEM) as the measure of variance.

748 **Logistic regression models and evaluation of neuron contribution to decoding**

749 We compared the decoding performance of the LSTM to that of a logistic regression model. The dataset used to  
750 train the logistic regression model was identical, barring one key change: the spike trains provided to the LSTM,  
751 initially of dimension (20, 2286), were reduced to a single bin representing the total number of spikes around a frame.  
752 The data then had a revised shape of (1, 2286), and no longer incorporated temporal information. We trained a  
753 logistic regression model using the `liblinear` solver implementation in Scikit-learn<sup>57</sup>. For training, we employed an  
754 L1 penalty and z-scored the neuronal activity per neuron using the mean and standard deviation of the training data.  
755 We utilized a nested 5-fold cross-validation, and separately optimized the regularization strength for each data split.  
756 The final decoding performance was calculated by averaging the performance across all five test sets.

757 **Logistic regression weights for evaluating individual neuron's contribution**

758 Applying an L1 penalty during training enforced feature sparsity, which facilitated the interpretation of input feature  
759 importance through the model's coefficients. A ranking of neurons was generated by evaluating the coefficients of  
760 the trained logistic regression models. Caution is needed when combining logistic regression weights from models  
761 trained on different splits due to the alternating test sets in each split, which sometimes include data used for training  
762 in another split's test data (see Fig. S7). To avoid any interference between training and test data across splits while still  
763 accounting for the temporal variation in our data, we further modified the splits used for cross-validation. The original  
764 five splits were extended to 20 splits in the following way: each of the original five splits was further divided into four  
765 sub-splits which shared a common test set but alternated the division of training and validation data (see visualization  
766 in Fig. S8). Any analysis based on neuron selection using logistic regression weights was evaluated exclusively on the  
767 corresponding four splits that generated the ranking and shared held-out test data. Final decoding performances  
768 for such analyses were derived through a nested cross-validation procedure. This involved initially averaging the  
769 decoding performances of models that shared a common test set (i.e. averaging across each set of four sub-splits) and  
770 then averaging the resulting five performances. In short, our training procedure for the logistic regression analysis  
771 consisted of training  $5 \times 4 = 20$  splits, and thus 20 models, where each group of 4 sub-splits shared a common test  
772 set.

773 Neurons were ranked according to their logistic regression coefficients across each set of four sub-splits. Given  
774 that there are a total of five such groups, this lead to five distinct rankings. To obtain each ranking, we combined the  
775 coefficients of the four trained models using a two-step procedure:

776 1. Partition the neurons into separate subsets, based on the number of models for which a neuron had a non-  
777 zero coefficient (e.g. one group of neurons which had non-zero coefficients on all four sub-splits, then all three

778 sub-splits, etc.).

779 2. Within each subset of four models, use the average of the absolute coefficients across the five splits to obtain  
780 a subset-specific ranking.

781 By concatenating the partitions of ranked neurons, we obtained a comprehensive ranking of all neurons. The neuron  
782 ranked highest displayed non-zero coefficients in all four models (corresponding to four sub-splits) and possessed  
783 the greatest average absolute coefficient value among neurons activated in all four models.

#### 784 **Intersection of top-performing neurons and chance-level overlap**

785 In our analysis, we restricted the decoding to subpopulations of neurons that were derived through a ranking of  
786 weights of a trained logistic regression model. To ensure that the selection of neurons was independent of the test  
787 data, the neuron selection procedure was based on five rankings derived from distinct data splits, each paired with  
788 fixed test data. We evaluated the intersection of neurons across subsets of top-ranked neurons from the five rankings,  
789 evaluated for varying subpopulation sizes (Fig. 7b). For instance, a comparison of the top-performing 500 neurons  
790 from each of the five rankings revealed a set of 78 common neurons.

791 As a reference point for comparison, we report the average count of overlapping neurons anticipated when ran-  
792 domly selecting sets of 500 neurons five times from the entire population (denoted as chance level of overlapping  
793 neurons in Fig. 7). In the previously mentioned scenario involving subpopulation sizes of 500, the expected number  
794 of overlapping neurons is equivalent to one. Mathematically, this is computed as follows: the full population consists  
795 of  $N = 2286$  neurons. We refer to the size of the subpopulation as  $k = 500$  and the number of total rankings  $m = 5$ .  
796 For each neuron  $n_i$  in the subpopulation, we define a random variable  $X_{i,500}$  as follows:

$$X_{i,500} = \begin{cases} 1, & \text{if neuron } n_i \text{ lies in all five subpopulations} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

We observe that  $\mathbb{P}(X_{i,500} = 1) = \left(\frac{k}{N}\right)^m = \left(\frac{500}{2286}\right)^5$ . The expected value of  $X_{i,500}$  is given by

$$\mathbb{E}(X_{i,500}) = 1 \cdot \mathbb{P}(X_{i,500} = 1) + 0 \cdot \mathbb{P}(X_{i,500} = 0) = \mathbb{P}(X_{i,500} = 1)$$

We define  $X_{500} = \sum_{i=1}^N X_{i,500}$  as the number of overlapping neurons across all five rankings. Since the random variables  
are independently and identically distributed, this implies

$$\mathbb{E}(X_{500}) = \sum_{i=1}^N \mathbb{E}(X_{i,500}) = \sum_{i=1}^N \mathbb{P}(X_{i,500} = 1) = 2286 \cdot \left(\frac{500}{2286}\right)^5 \approx 1.1443$$

Analogous calculations for  $k = 100, 350, 750, 1000$  yield

$$\begin{aligned} \mathbb{E}(X_{100}) &= 2286 \cdot \left(\frac{100}{2286}\right)^5 \approx 0.0004 \\ \mathbb{E}(X_{350}) &= 2286 \cdot \left(\frac{350}{2286}\right)^5 \approx 0.1923 \\ \mathbb{E}(X_{750}) &= 2286 \cdot \left(\frac{750}{2286}\right)^5 \approx 8.6896 \\ \mathbb{E}(X_{1000}) &= 2286 \cdot \left(\frac{1000}{2286}\right)^5 \approx 36.6180 \end{aligned}$$

797 Thus, we derive the chance baselines as 0, 0, 1, 9, and 37 for subpopulation sizes of 100, 350, 500, 750, and 1000, respec-  
798 tively.

#### 799 **Decoding on regions of the MTL**

800 We compared the decoding performance when using the activity of all 2286 recorded neurons to the performance  
801 when only using activity from specific regions of the MTL. These regions are the amygdala (580 neurons), hippocampus  
802 (794), entorhinal cortex (440), and parahippocampal cortex (373). To use activity from a particular region, we limited

803 ourselves to the activity of neurons in that region and reduced the input dimension to match the number of neurons  
804 in the region. The network architecture and data splits remained the same as when using activity from the full pop-  
805 ulation, but the hyperparameters were optimized for the reduced dataset and given label. Training, validation, and  
806 test set sizes remained the same as the full dataset condition. In summary, decoding from different regions differed  
807 from full population decoding primarily due to reduced input data dimensionality: from a spike train of length 20 and  
808 dimension 2286 to a spike train of the same length but with a dimension reduced according to the number of neurons  
809 in the specific region.

## 810 **Acknowledgments**

811 We would like to thank Tamara Müller and Aleksandar Levic for their contributions to the data analysis framework.  
812 This work was supported by the German Federal Ministry of Education and Research (DeepHumanVision, FKZ: 031L0197A-  
813 C; Tübingen AI Center, FKZ: 01IS18039), as well as the German Research Foundation (DFG) through Germany's Excel-  
814 lence Strategy (Cluster of Excellence Machine Learning for Science, EXC-Number 2064/1, PN 390727645) and SFB1233  
815 (PN 276693517), SFB 1089 (PN 227953431), SPP 2411 (PN: 520287829), and MO 930/4-2.

## 816 **Data Availability Statement**

817 The data will be made fully available in a separate publication following the release of this manuscript. Code will be  
818 made available on GitHub.

## 819 **Author Contributions**

820 **Conceptualization:** Franziska Gerken, Alana Darcher, Pedro J Gonçalves, Rachel Rapp, Ismail Elezi, Johannes Niediek,  
821 Stefanie Liebe, Jakob H Macke, Florian Mormann, Laura Leal-Taixé

822 **Data acquisition:** Alana Darcher, Johannes Niediek, Thomas P Reber, Marcel S Kehl, Stefanie Liebe, Florian Mormann

823 **Data curation:** Alana Darcher, Marcel S Kehl

824 **Methodology:** Franziska Gerken, Alana Darcher, Pedro J Gonçalves, Rachel Rapp, Ismail Elezi, Stefanie Liebe, Jakob  
825 H Macke, Florian Mormann, Laura Leal-Taixé

826 **Formal analysis:** Franziska Gerken, Alana Darcher

827 **Funding:** Jakob H Macke, Florian Mormann, Laura Leal-Taixé

828 **Software:** Franziska Gerken, Alana Darcher, Johannes Niediek, Thomas P Reber

829 **Project administration:** Jakob H Macke, Florian Mormann, Laura Leal-Taixé

830 **Supervision:** Pedro J Gonçalves, Ismail Elezi, Stefanie Liebe, Jakob H Macke, Florian Mormann, Laura Leal-Taixé

831 **Writing - original draft:** Franziska Gerken, Alana Darcher

832 **Writing - review and editing:** Pedro J Gonçalves, Rachel Rapp, Ismail Elezi, Stefanie Liebe, Jakob H Macke, Florian  
833 Mormann, Laura Leal-Taixé

834

## 835 **References**

836 [1] G. Kreiman, C. Koch, and I. Fried. Category-specific visual responses of single neurons in the human medial temporal lobe.  
837 *Nature Neuroscience*, 3(9):946–953, 2000.

838 [2] A. Kraskov, R.Q. Quiroga, L. Reddy, I. Fried, and C. Koch. Local field potentials and spikes in the human medial temporal lobe  
839 are selective to image category. *Journal of Cognitive Neuroscience*, 19(3):479–492, 2007.

840 [3] R.Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single neurons in the human brain.  
841 *Nature*, 435(7045):1102–1107, 2005.

842 [4] T.P. Reber, M. Bausch, S. Mackay, J. Boström, C.E. Elger, and F. Mormann. Representation of abstract semantic knowledge in  
843 populations of human single neurons in the medial temporal lobe. *PLOS Biology*, 17(6):1–17, 06 2019.

844 [5] G. Kreiman, I. Fried, and C. Koch. Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings  
845 of the National Academy of Sciences*, 99(12):8378–8383, 2002.

846 [6] R.Q. Quiroga, R. Mukamel, E.A. Isham, R. Malach, and I. Fried. Human single-neuron responses at the threshold of conscious  
847 recognition. *Proceedings of the National Academy of Sciences*, 105(9):3599–3604, 2008.

848 [7] R.Q. Quiroga, A. Kraskov, F. Mormann, I. Fried, and C. Koch. Single-cell responses to face adaptation in the human medial  
849 temporal lobe. *Neuron*, 84(2):363–369, 2014.

850 [8] T.P. Reber, J. Faber, J. Niediek, J. Boström, C.E. Elger, and F. Mormann. Single-neuron correlates of conscious perception in the  
851 human medial temporal lobe. *Current Biology*, 27(19):2991–2998, 2017.

852 [9] S. Mackay, T.P. Reber, M. Bausch, J. Boström, C.E. Elger, and F. Mormann. Concept and location neurons in the human brain  
853 provide the ‘what’and ‘where’in memory formation. *Nature Communications*, 15(1):7926, 2024.

854 [10] B.P. Staresina, T.P. Reber, J. Niediek, J. Boström, C.E. Elger, and F. Mormann. Recollection in the human hippocampal-entorhinal  
855 cell circuitry. *Nature Communications*, 10(1):1503, 2019.

856 [11] J. Zheng, A.G.P. Schjetnan, M. Yebra, B.A. Gomes, C.P. Mosher, S.K. Kalia, T.A. Valiante, A.N. Mamelak, G. Kreiman, and  
857 U. Rutishauser. Neurons detect cognitive boundaries to structure episodic memories in humans. *Nature Neuroscience*,  
858 25(3):358–368, 2022.

859 [12] H.G. Rey, M.J. Ison, C. Pedreira, A. Valentin, G. Alarcon, R. Selway, M.P. Richardson, and R.Q. Quiroga. Single-cell recordings in  
860 the human medial temporal lobe. *Journal of Anatomy*, 227(4):394–408, 2015.

861 [13] U. Rutishauser, L. Reddy, F. Mormann, and J. Sarnthein. The architecture of human memory: insights from human single-  
862 neuron recordings. *Journal of Neuroscience*, 41(5):883–890, 2021.

863 [14] S. Nishimoto, A.T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J.L. Gallant. Reconstructing visual experiences from brain activity  
864 evoked by natural movies. *Current Biology*, 21(19):1641–1646, 2011.

865 [15] H. Wen, J. Shi, Y. Zhang, K.-H. Lu, J. Cao, and Z. Liu. Neural encoding and decoding with deep learning for dynamic natural vision.  
866 *Cerebral Cortex*, 28(12):4136–4160, 10 2017.

867 [16] A. G. Huth, S. Nishimoto, A. T. Vu, and J. L. Gallant. A continuous semantic space describes the representation of thousands of  
868 object and action categories across the human brain. *Neuron*, 76(6):1210–1224, 2012.

869 [17] J. V. Haxby, M. I. Gobbini, and S. A. Nastase. Naturalistic stimuli reveal a dominant role for agentic action in visual representation.  
870 *NeuroImage*, 216:116561, 2020.

871 [18] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach. Intersubject synchronization of cortical activity during natural vision.  
872 *Science*, 303(5664):1634–1640, 2004.

873 [19] J.N. Van der Meer, M. Breakspear, L.J. Chang, S. Sonkusare, and L. Cocchi. Movie viewing elicits rich and reliable brain state  
874 dynamics. *Nature Communications*, 11(1):5004, 2020.

875 [20] J.M. Zacks, T.S. Braver, M.A. Sheridan, D.I. Donaldson, A.Z. Snyder, J.M. Ollinger, R.L. Buckner, and M.E. Raichle. Human brain  
876 activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6):651–655, 2001.

877 [21] C. Baldassano, J. Chen, A. Zadbood, J.W. Pillow, U. Hasson, and K.A. Norman. Discovering event structure in continuous narrative  
878 perception and memory. *Neuron*, 95(3):709–721.e5, 2017.

879 [22] L. Isik, J. Singer, J.R. Madsen, N. Kanwisher, and G. Kreiman. What is changing when: Decoding visual information in movies  
880 from human intracranial recordings. *NeuroImage*, 180:147–159, 2018.

881 [23] H. Gelbard-Sagiv, R. Mukamel, M. Harel, R. Malach, and I. Fried. Internally generated reactivation of single neurons in human  
882 hippocampus during free recall. *Science*, 322(5898):96–101, 2008.

883 [24] Y. Zhang, Z.M. Aghajan, M. Ison, Q. Lu, H. Tang, G. Kalender, T. Monsoor, J. Zheng, G. Kreiman, V. Roychowdhury, and I. Fried.  
884 Decoding of human identity by computer vision and neuronal vision. *Scientific Reports*, 13(1):1–16, 2023.

885 [25] StabilityAI. Stable diffusion. <https://beta.dreamstudio.ai/generate>, 2024.

886 [26] E. Maris and R. Oostenveld. Nonparametric statistical testing of eeg-and meg-data. *Journal of Neuroscience Methods*, 164(1):177–  
887 190, 2007.

888 [27] S. Panzeri, J.H. Macke, J. Gross, and C. Kayser. Neural population coding: combining insights from microscopic and mass signals.  
889 *Trends in Cognitive Sciences*, 19(3):162–172, March 2015. Publisher: Elsevier.

890 [28] J. Kamiński, S. Sullivan, J.M. Chung, I.B. Ross, A.N. Mamelak, and U. Rutishauser. Persistently active neurons in human medial  
891 frontal and medial temporal lobe support working memory. *Nature Neuroscience*, 20(4):590–601, 2017.

892 [29] M. Jamali, B.L. Grannan, E. Fedorenko, R. Saxe, R. Báez-Mendoza, and Z.M. Williams. Single-neuronal predictions of others'  
893 beliefs in humans. *Nature*, 591(7851):610–614, 2021.

894 [30] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–80, 12 1997.

895 [31] I. Fried, K.A. MacDonald, and C.L. Wilson. Single neuron activity in human hippocampus and amygdala during recognition of  
896 faces and objects. *Neuron*, 18(5):753–765, 1997.

897 [32] F. Mormann, J. Niediek, O. Tudusciuc, C.M. Quesada, V.A. Coenen, C.E. Elger, and R. Adolphs. Neurons in the human amygdala  
898 encode face identity, but not gaze direction. *Nature Neuroscience*, 18(11):1568–1570, 2015.

899 [33] R. Cao, X. Li, N.J. Brandmeir, and S. Wang. Encoding of facial features by single neurons in the human amygdala and hippocam-  
900 pus. *Communications Biology*, 4(1):1394, 2021.

901 [34] M. Bausch, J. Niediek, T.P. Reber, S. Mackay, J. Boström, C.E. Elger, and F. Mormann. Concept neurons in the human medial  
902 temporal lobe flexibly represent abstract relations between concepts. *Nature Communications*, 12(1):6164, 2021.

903 [35] Jose A. Garcia-Lazaro, L.A.C. Belliveau, and N.A. Lesica. Independent population coding of speech with sub-millisecond precision.  
904 *Journal of Neuroscience*, 33(49):19362–19372, December 2013.

905 [36] R.A.A. Ince, S. Panzeri, and C. Kayser. Neural codes formed by small and temporally precise populations in auditory cortex.  
906 *Journal of Neuroscience*, 33(46):18277–18287, November 2013.

907 [37] Indre V. Viskontas, Rodrigo Quiroga, and Itzhak Fried. Human medial temporal lobe neurons respond preferentially to  
908 personally relevant images. *Proceedings of the National Academy of Sciences*, 106(50):21329–21334, 2009.

909 [38] U. Rutishauser, O. Tudusciuc, D. Neumann, A. N. Mamelak, A. C. Heller, I. B. Ross, L. Philpott, W. W. Sutherland, and R. Adolphs.  
910 Single-unit responses selective for whole faces in the human amygdala. *Current Biology*, 21(19):1654–1660, 2011.

911 [39] R. Cao, J. Wang, P. Brunner, J. T. Willie, X. Li, U. Rutishauser, N. J. Brandmeir, and S. Wang. Neural mechanisms of face familiarity  
912 and learning in the human amygdala and hippocampus. *Cell Reports*, 43(1), 2024.

913 [40] V. Troiani, A. Stigliani, M.E. Smith, and R.A. Epstein. Multiple object properties drive scene-selective regions. *Cerebral Cortex*,  
914 24(4):883–897, 2014.

915 [41] D.D. Dilks, J.B. Julian, A.M. Paunov, and N. Kanwisher. The occipital place area is causally and selectively involved in scene  
916 perception. *Journal of Neuroscience*, 33(4):1331–1336, 2013.

917 [42] F. Mormann, S. Kornblith, M. Cerf, M.J. Ison, A. Kraskov, M. Tran, S. Knieling, R. Quiroga, C. Koch, and I. Fried. Scene-  
918 selective coding by single neurons in the human parahippocampal cortex. *Proceedings of the National Academy of Sciences*,  
919 114(5):1153–1158, 2017.

920 [43] J. Niediek, J. Boström, C.E. Elger, and F. Mormann. Reliable analysis of single-unit recordings from the human brain under noisy  
921 conditions: tracking neurons over hours. *PLOS One*, 11(12):e0166598, 2016.

922 [44] G. Dehnen, M.S. Kehl, A. Darcher, T.T. Müller, J.H. Macke, V. Borger, R. Surges, and F. Mormann. Duplicate detection of spike  
923 events: A relevant problem in human single-unit recordings. *Brain Sciences*, 11(6):761, 2021.

924 [45] A. Darcher, R. Rapp, and T. Mueller. Epiphyte. <https://github.com/mackelab/epiphyte>, May 2024.

925 [46] FFmpeg. <https://ffmpeg.org/>, 2012.

926 [47] O. Aubert and Y. Prié. Advene: active reading through hypervideo. *Proceedings of ACM Hypertext'05*, pages 235–244, 2005.

927 [48] T. Esler. Facenet pytorch. <https://github.com/timesler/facenet-pytorch.git>.

928 [49] B. Castellano. PySceneDetect. <https://github.com/Breakthrough/PySceneDetect>.

929 [50] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright,  
930 S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng,  
931 E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H.  
932 Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing  
933 in Python. *Nature Methods*, 17:261–272, 2020.

934 [51] R.J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 12 1986.

935 [52] D.P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014. Published as a conference paper at the 3rd Interna-  
936 tional Conference for Learning Representations, San Diego, 2015.

937 [53] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf,  
938 E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style,  
939 high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran  
940 Associates, Inc., 2019.

941 [54] J. Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

942 [55] A.D. Ekstrom, M.J. Kahana, J.B. Caplan, T.A. Fields, E.A. Isham, E.L. Newman, and I. Fried. Cellular networks underlying human  
943 spatial navigation. *Nature*, 425(6954):184–188, 2003.

944 [56] J.F. Miller, I. Fried, N. Suthana, and J. Jacobs. Repeating spatial activations in human entorhinal cortex. *Current Biology*,  
945 25(8):1080–1085, 2015.

946 [57] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg,  
947 J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python.  
948 *Journal of Machine Learning Research*, 12:2825–2830, 2011.