nature communications

Article

https://doi.org/10.1038/s41467-025-57220-y

Memorability of novel words correlates with anterior fusiform activity during reading

Received: 16 April 2024

Accepted: 13 February 2025

Published online: 23 February 2025

Check for updates

Oscar Woolnough [1,2] & Nitin Tandon [1,2,3]

Our memory for the words we already know is best predicted by their associated meanings. However, the factors that influence whether we will remember a new word after we see it for the first time are unclear. We record memory performance for 2100 novel pseudowords across 1804 participants during a continuous recognition task. Participants show significant agreement across individuals for which novel words were memorable or forgettable, suggesting an intrinsic memorability for individual pseudowords. Pseudowords that are similar to low-frequency known words, with sparse orthographic neighbourhoods and rarely occurring letter pairs, are more memorable. Further, using intracranial recordings in 36 epilepsy patients we show a region in the anterior fusiform cortex that shows sensitivity to the memorability of these pseudowords. These results suggest that known words in our lexicon act as a scaffold for remembering novel word forms, with rare and unique known words providing the best support for novel word learning.

In our everyday lives, we frequently encounter words that we have never seen before[1]. This can include building our vocabulary, new words coming into common use, or names of new people, companies, or drugs. But what factors affect our ability to remember these words after we first see them? In order to learn a new written word we need to be able to remember the specific combination of letters associated with the word and distinguish this from other similar word forms. This process begins with an initial rapid familiarisation period, believed to be mediated by medial temporal and hippocampal memory circuits[2,3]. During this period, before cortical consolidation, memory retrieval of novel words is slower, less efficient, and less accurate than for known words. Probing which novel word forms are more efficiently encoded during this period will teach us more about how this initial phase of word learning occurs.

Intrinsic memorability[4] of an item is a measure of the likelihood, across multiple individuals, of remembering that item when reencountering it or attempting to recall it. Robust intrinsic memorability of specific items has been demonstrated for many classes of visual stimuli including words[5–7], faces[8], scenes[9,10], objects[11], symbols[12], and artwork[13]. The memorability of these visual stimuli modulates neural activation across ventral occipitotemporal cortex (vOTC), with more memorable images typically showing greater activation, most notably within category-selective regions such as fusiform face area[4].

Memorability for known words is most prominently influenced by the semantics we have learned to associate with their word form[5–7]. Though when we first encounter novel words they have no inherent meaning to us and we cannot easily utilise semantic memory to encode them. Robust memorability has been demonstrated for meaningless phase-scrambled images, suggesting memorability can exist in the absence of semantics[10]. However, this leaves an open question as to, in the absence of meaning, which other factors drive our ability to remember novel word forms.

Within vOTC, novel pseudowords typically show greater neural activation than known words while reading[14–19], likely as pseudoword processing is less efficient than the processing of known words through this route. vOTC is sensitive to a pseudoword's orthographic neighbourhood and bigram frequency[18,20–22] and is highly sensitive to the frequency at which known words occur in natural language[17,18,22–25]. Consequently, these factors are likely candidates for driving differences in memorability between pseudowords.

Here, to derive a measure of novel word memorability we perform an online study of pseudoword memory. Using these data, we derive

[1]Vivian L. Smith Department of Neurosurgery, McGovern Medical School at UT Health Houston, Houston, TX, USA. [2]Texas Institute for Restorative Neurotechnologies, University of Texas Health Science Center at Houston, Houston, TX, USA. [3]Memorial Hermann Hospital, Texas Medical Center, Houston, TX, USA. e-mail: oscar.woolnough@uth.tmc.edu

empirical memorability scores for 2100 unique novel pseudowords and, from this, reveal the factors underlying the successful recognition of previously encountered pseudowords. Further, to find neural correlates of these memorability effects, we use intracranial recordings in 36 individuals, with 3259 electrodes in their left, language-dominant hemisphere, as they silently read sentences and lists made of pseudowords, to map the sensitivity of vOTC to pseudoword memorability.

Results

A total of 1804 participants performed a continuous recognition task on the online platform Prolific. Participants were each presented with 256 pseudowords in succession and were asked to press a button if any of the items were repeated (Fig. 1a). Fifty words were repeated, with a >30 s lag between the prime and repeat. We calculated the memorability of each word as its corrected recognition (CR) score across participants, measured by the hit rate minus the false alarm rate. Mean hit rate across participants was 52 ± 25% (mean ± SD), false alarm rate was 10 ± 9%, and mean response time was 805 ± 97 ms (Fig. 1b). To assess the consistency of memorability across participants we calculated the split-half consistency, across 1000 permutations (Fig. 1c). This demonstrated that random participant halves had significant agreement on the memorability of individual pseudowords (Spearman-Brown corrected $\rho = 0.38$, $p < 0.001$).

Memorability of novel words

There was a large variability between the memorability of individual items (Mean CR = 0.42 ± 0.1, range 0.11–0.78) (Figs. 2a, b, 3).

We modelled the memorability of individual items as a function of several sublexical factors, using multiple linear regression (MLR). We included factors known to modulate reading speed and neural activation for pseudowords, including letter length, orthographic neighbourhood (OLD20[26], the mean number of character-edits to a pseudoword's 20 nearest known word neighbours), the upper and lower frequency bounds of a pseudoword's closest known word neighbours, and letter co-occurrence statistics (bi- and quadri-gram frequencies).

This model showed significant effects of orthographic neighbourhood, bigram and quadrigram frequencies, and the frequency of the pseudoword's lowest-frequency neighbour (Table 1). The highest memorabilities were seen for pseudowords with a sparse orthographic neighbourhood, low-frequency bigrams but high-frequency quadrigrams, and low-frequency neighbours.

Highly memorable six-to-eight-letter long pseudowords showed significantly sparser orthographic neighbourhoods between two to five character-edits away (Fig. 4). Five-letter words only showed these differences out to four character-edits, and two character-edits for four-letter words, suggesting an initial length dependency that saturates for pseudowords six letters or longer.

We found that the observed memorability scores were a significant predictor of response time for correctly identified items (LME, $t(47{,}064) = -8.1$, $\beta = -70$, $p < 0.001$, 95% CI −87 to −53), with more memorable items resulting in faster response times.

For the subset of the pseudowords that had published lexical decision time data[27] (1300 items), we also ran a model including the mean decision time as a regressor. This model demonstrated a strong, significant effect of lexical decision time on memorability (MLR, $t(1292) = 5.42$, $\beta = 0.27$, $p < 0.001$, 95% CI 0.17 to 0.36), with pseudowords with longer lexical decision times showing higher memorability.

We used the pseudoword predictors to attempt to predict known word memorability scores from a previous study[5]. An MLR model fit to the item level memorability scores for the pseudowords resulted in a low predictive power for known word memorability (Spearman, $\rho(236) = 0.06$). This value was significantly lower than the predictive power within the pseudoword set, predicting a held-out set of stimuli

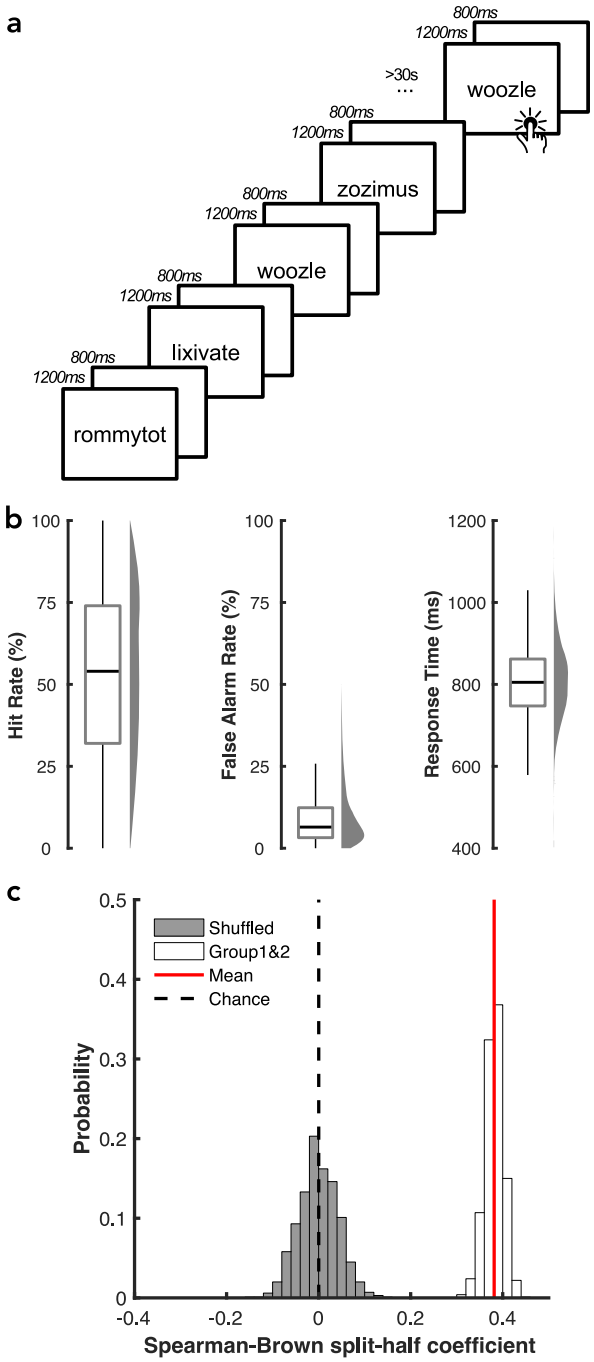


Fig. 1 | Experimental design and behavioural analysis. a Schematic representation of the continuous recognition task. **b** Distributions of participant hit rates, false alarm rates and response times during the continuous recognition task. Box plots represent quartiles, with whiskers extending to ±1.5 × IQR. Density plots were calculated using the random average shifted histogram method. **c** Spearman-Brown split-half analysis of the memorability results, testing 1000 randomly split halves of the data against shuffled data.

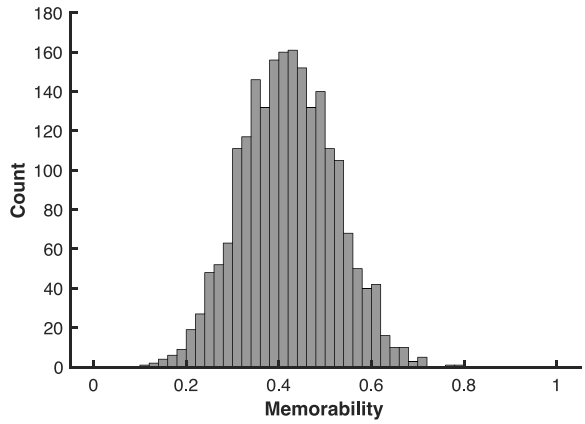
the same size as the known word set (Monte Carlo, 1000 iterations, $p < 0.001$, 95% CI 0.15–0.34).

Neural representation of novel word memorability

Thirty-six participants with intracranial electrodes placed in their left, language-dominant hemisphere for the localisation of intractable epilepsy were visually presented with eight-word lists of words or

https://doi.org/10.1038/s41467-025-57220-y

a



b

Low Memorability		High Memorability	
colp	0.14	snig	0.68
gilks	0.13	alvil	0.71
fonned	0.11	jubjub	0.71
feplete	0.16	streech	0.70
brastles	0.16	slaytime	0.78

c

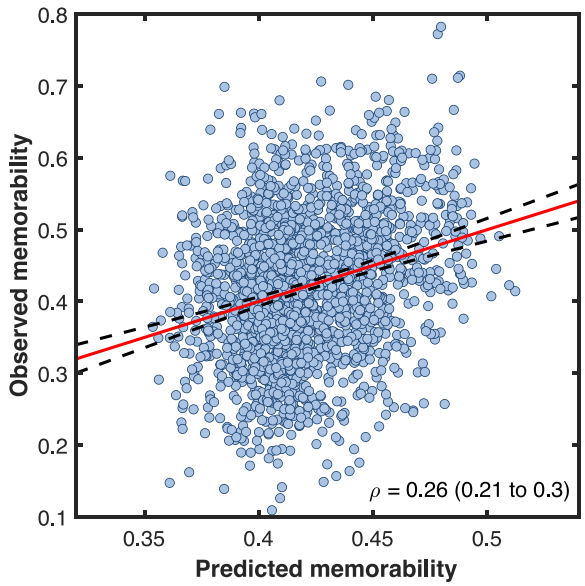


Fig. 2 | Quantification of novel word memorability. a Histogram of pseudoword memorabilities within our 2100-item corpus. **b** Exemplar items showing the highest and lowest memorability items for each word length. **c** Predictability of item memorability based on multiple linear regression, showing the Spearman rank correlation coefficient with 95% confidence interval.

pseudowords or structured pseudoword sentences. Thirty-four participants had stereotactic EEG electrodes (sEEGs) and two had subdural grid electrodes (SDEs). Words were presented in rapid serial visual presentation format at a comfortable reading pace (500 ms/item), followed by a two-alternative forced choice decision between a presented vs. non-presented word from the preceding list. Each participant was presented with up to 687 pseudowords with derived memorability scores from the online study (566 unique items). Task accuracy was high across all trial types (word list, 92 ± 6%; pseudoword sentence, 89 ± 8%; pseudoword list, 85 ± 10%). In trials where both of the forced choice options had memorability scores from the online study, participants were significantly more likely to answer correctly when the correct item had a high memorability score (mixed-effects logistic regression, $t(382) = 2.37$, $\beta = 5.09$, $p = 0.018$, 95% CI 0.86 to 9.31); however, the memorability of the foil did not have a significant effect on responses ($t(382) = -1.16$, $\beta = -1.68$, $p = 0.25$, 95% CI −4.51 to 1.16).

We generated population-level maps to tease apart the factors driving cortical activation using surface-based linear mixed-effects (sbLME) analysis. This model used the mean broadband gamma activity (BGA; 70–150 Hz) from 200–400 ms after the onset of each word in normalised cortical surface space. sbLME enables the separation of contributions of multiple linguistic features to BGA while adjusting for unequal sampling and inter-individual variations in activation. The features used for these analyses included the memorability score from the online study and lexicality. Effects of word position in the sequence, orthographic neighbourhood, bigram and quadrigram frequencies, and lowest frequency neighbour were regressed out to probe the effects of memorability and lexicality beyond those of their related sublexical statistics.

First, we applied an sbLME model to quantify the effects of memorability, using exclusively pseudoword trials. This revealed a distinct region in anterior fusiform displaying significant sensitivity to pseudoword memorability (Fig. 5a). In contrast, a comparable sbLME model contrasting words and pseudowords, revealed a broad effect across most of ventral temporal cortex, extending from early visual cortex to anterior fusiform, with greater activation in response to pseudowords than words (Fig. 5b).

To visualise the time course of their responses, we used ROIs centred on three segments of the fusiform gyrus—posterior fusiform (pFus), mid-fusiform (mFus), and anterior fusiform (aFus)—to isolate the spatiotemporal properties of the memorability response. For each ROI, we analysed activity over time using linear mixed-effects (LME) models. We observed significant modulation of activity in aFus by memorability from 280 to 390 ms following word onset, with more memorable pseudowords showing lower activation (Fig. 5d). In contrast, pFus and mFus did not show significant modulation by memorability.

We additionally tested ROIs in inferior parietal sulcus (IPS), precentral sulcus (pCS), and inferior frontal gyrus (IFG), regions known to be crucial for pseudoword processing; however, we observed no significant modulation by memorability (Fig. 6).

Discussion

In this study, we provide an empirical quantification of novel word memorability for a large corpus of pseudowords. This memorability is repeatable across individuals, and, at corpus level, memorability can be predicted by a pseudoword's relationship to words already in the lexicon, but with a large degree of variability within individual items. Additionally, we demonstrate a region in the anterior fusiform whose activity is sensitive to the memorability of pseudowords while reading.

Memory for new words, the initial stage toward learning novel words, appears to use our existing lexicon as a scaffold, with novel words that are close to distinctive, unique words resulting in stronger memorability. Pseudowords similar to low-frequency known words that themselves have low orthographic neighbourhoods (e.g., lyrx-*lynx*, sombat-*wombat*, dinoseur-*dinosaur*) were more memorable. The presence of low-frequency bigrams was also predictive of high memorability (e.g., jisc, snylark, unzie). Pseudowords that were perceived as more word-like, as evidenced by longer lexical decision times, were also more memorable (e.g., addessed, 0.65, 914 ms; piecrast, 0.62, 903 ms) while less word-like forms were less memorable (e.g., sloximal, 0.37, 606 ms; delp, 0.25, 617 ms). These results replicate an effect seen with known words[5], that retrieval of more memorable items is not only more accurate but also faster.

Verbal memory for written words is commonly thought of as dependent on both semantic memory and the orthographic lexicon,

https://doi.org/10.1038/s41467-025-57220-y

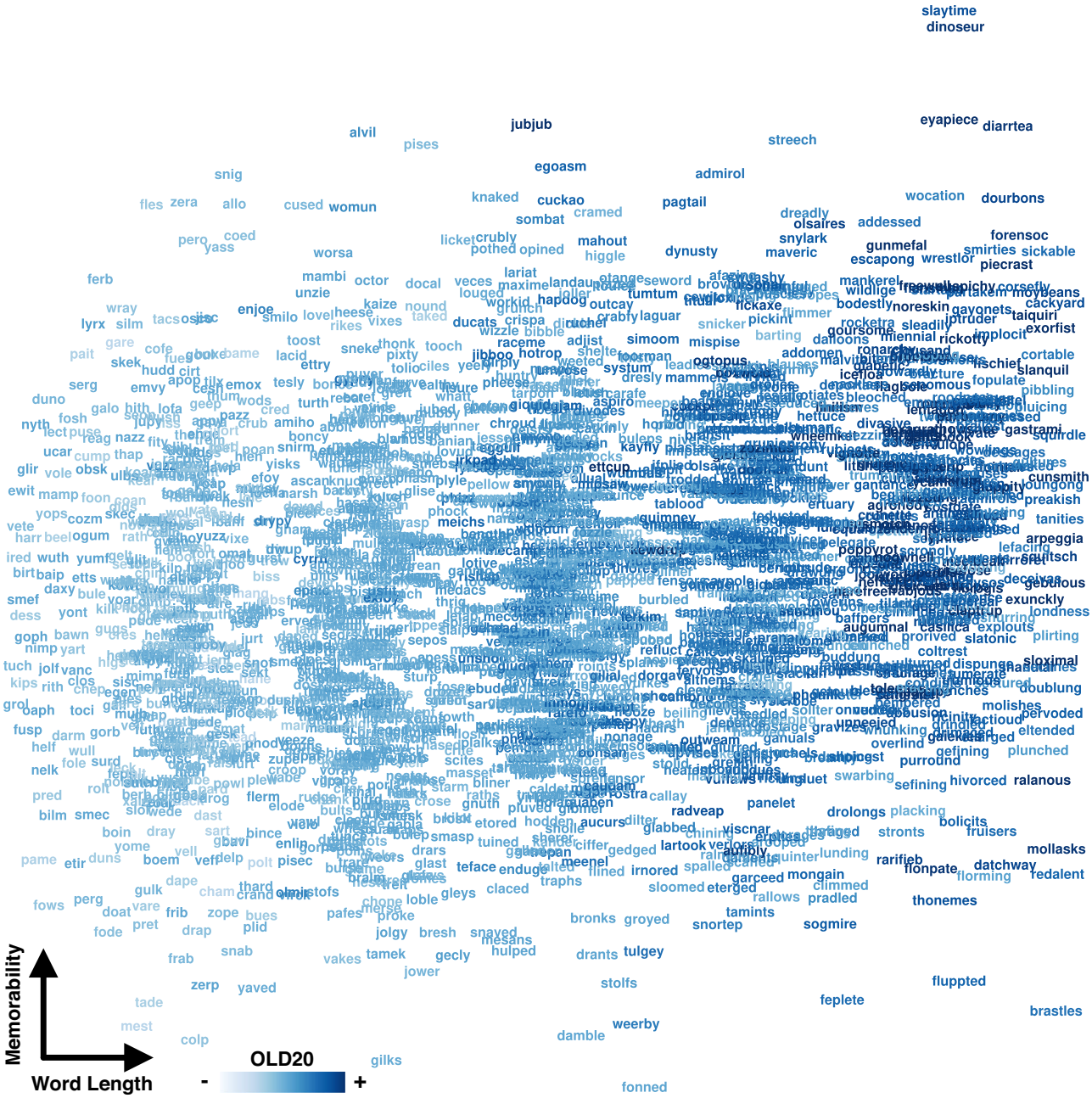


Fig. 3 | Distribution of novel word memorabilities. Spatial map of the pseudoword corpus with the vertical axis representing memorability and horizontal axis distributed by word length. Words are coloured by orthographic neighbourhood, with sparser neighbourhoods being darker.

Table 1 | Multiple linear regression of novel word memorability

	β (95% CI)	*t*	*p*-value
Intercept	0.42 (0.41–0.42)	199	0
Length	−0.00018 (−0.006 to −0.006)	−0.06	0.95
Orthographic neighbourhood (OLD20)	0.039 (0.027–0.052)	6.02	2×10^{-9}
Highest frequency neighbour	−0.0009 (−0.005 to 0.0036)	−0.38	0.71
Lowest frequency neighbour	0.0095 (0.0029–0.016)	2.82	0.005
Mean bigram frequency	−0.013 (−0.022 to −0.028)	−2.51	0.012
Mean quadrigram frequency	0.072 (0.026–0.12)	3.04	0.002

MLR model (df = 2093, $r^2 = 0.07$) of memorability during the continuous recognition task.

the long-term memory storage of known word forms[28]. The dependence of pseudoword memorability on orthographic similarity to known word forms implies the influence of the orthographic lexicon, as the novel words have no inherent meaning that can be encoded by semantic memory. However, access to the orthographic lexicon while reading typically engages a region in mFus more posterior to the aFus cluster seen here, as shown both functionally[18,24,29,30] and causally[31–33]. This could suggest distinct processes within the orthographic lexicon for the recognition of known orthographic forms and the encoding of novel orthographic forms. This could also suggest the memorability cluster represents a distinct domain-general memory region that may not be specific to novel words but is dependent on inputs from the orthographic lexicon during the encoding process for novel words.

Using intracranial recordings during reading aloud of words and pseudowords we have previously shown traditional lexical stream regions, such as mFus, show greater activation to pseudowords than

https://doi.org/10.1038/s41467-025-57220-y

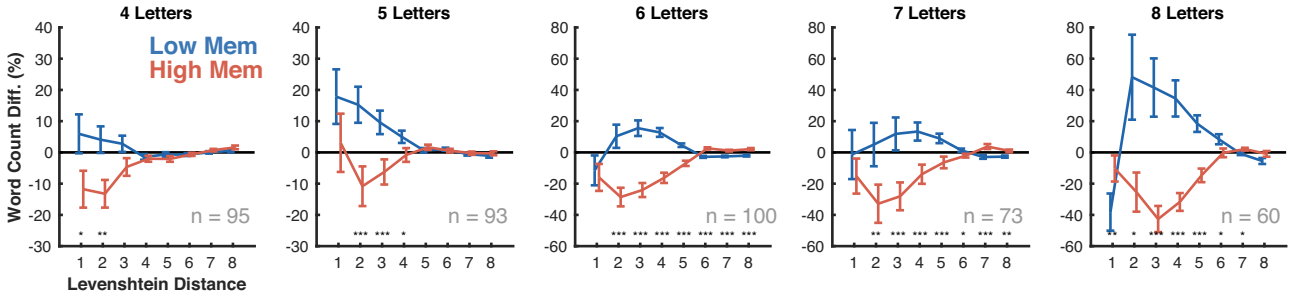


Fig. 4 | Neighbourhood density of memorable pseudowords. Difference (mean ± SE) in neighbourhood size at varying Levenshtein (character-edit) distances for pseudowords of each word length. Low (<20th percentile; blue) and high (>80th percentile; red) memorability score pseudowords' neighbourhood sizes were calculated as their percentage difference from median memorability (40–60th percentile) pseudowords within each word length. Low and high memorability neighbourhood sizes were compared using two-tailed Wilcoxon rank-sum tests; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Number of pseudowords per quintile is shown for each word length.

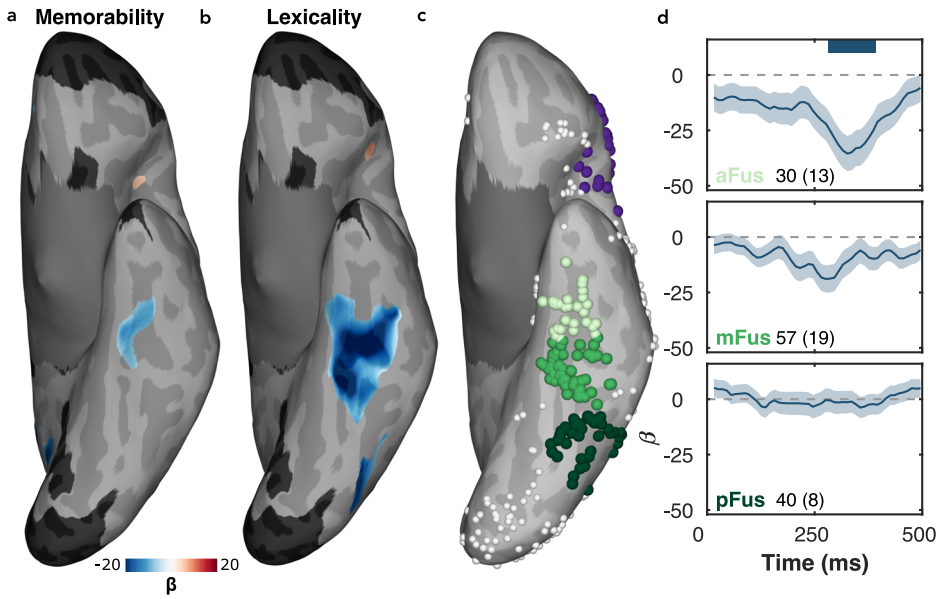


Fig. 5 | Intracranial correlates of novel word memorability. a, **b** sbLME models for **a** memorability and **b** lexicality, showing significant clusters ($p < 0.01$, Monte Carlo cluster corrected) of sensitivity. Regions in black did not have sufficient coverage for reliable sbLME results (<3 patients). **c** Anatomical ROI definitions for posterior (pFus; dark green), mid (mFus; green), and anterior (aFus; light green) fusiform cortex. Pseudoword responsive electrodes not included in an ROI shown in white. **d** LME beta values (±SE) for the effect of memorability within each ROI, calculated at each time point independently (10 ms resolution). Effects of word position in the sequence, orthographic neighbourhood, bigram and quadrigram frequencies, and lowest frequency neighbour were regressed out. Number of electrodes and patients per cluster is shown. Coloured bars represent regions of significance from the LME analyses ($q < 0.01$ FDR-corrected).

real words[17]. We also observe disruption of the ability to read pseudowords when vOTC is stimulated[31–33], suggesting a causal role in pseudoword reading. This is contrary to some predictions based on dual-stream cognitive models which predict the lexical route should only be sensitive to known words[15]. Here, we show the importance of neighbouring known words on the memory of novel pseudowords, with higher memorability words showing sparser orthographic neighbourhoods, out to five character-edits away. This could suggest that when pseudowords are read then the internal representations of known words up to five character-edits away are activated to aid processing, though activation of too many comparable orthographic representations may be deleterious to memory. While novel pseudowords can be pronounced using grapheme-to-phoneme correspondences, there is considerable inter-individual and within-individual variability in which pronunciations are used, not just utilising the most probable correspondences[34]. This could suggest we use internal representations of this broad range of orthographically similar known words in vOTC to facilitate our processing of novel pseudowords.

We found a region of anterior fusiform whose pseudoword-induced activation is sensitive to memorability. This region is more anterior to where we have previously shown the initial distinction between words and pseudowords occurs[17,22], likely suggesting a subsequent process. Previous intracranial work has shown that more memorable known words result in faster reinstatement of neural activity in the anterior temporal lobe during memory retrieval[5]. Resection of left, but not right, anterior temporal lobe is also associated with impairments to recognition of previously encountered pseudowords[35]. This region is also in close proximity to regions of parahippocampal gyrus and perirhinal cortex that are sensitive to the familiarity of faces and scenes[36,37]. While sublexical route regions, such as IPS and pCS, are typically more associated with novel word processing than lexical route regions, we find no evidence for sublexical route involvement in pseudoword memorability. It is possible that this association naturally emerges due to vOTC's proximity to medial temporal and entorhinal memory structures critical for verbal memory[38]. This raises the question of what role the sublexical route plays during the early stages of novel word learning.

https://doi.org/10.1038/s41467-025-57220-y

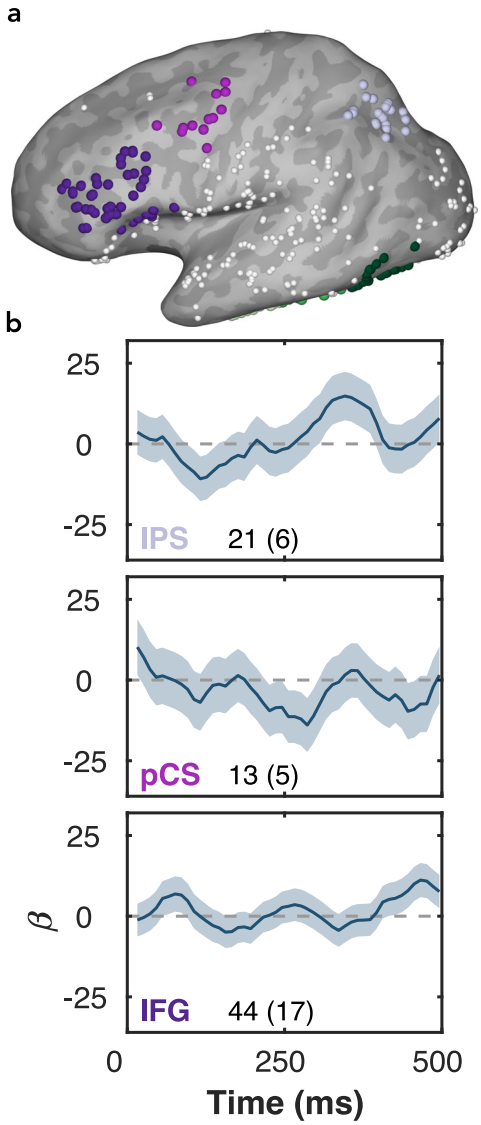


Fig. 6 | Novel word memorability outside of vOTC. a Anatomical ROI definitions for inferior frontal gyrus (IFG; dark purple), precentral sulcus (pCS; purple), and inferior parietal sulcus (IPS; light purple). Pseudoword responsive electrodes not included in an ROI shown in white. **b** LME beta values (±SE) for the effect of memorability within each ROI, calculated at each time point independently (10 ms resolution). Effects of word position in the sequence, orthographic neighbourhood, bigram and quadrigram frequencies, and lowest frequency neighbour were regressed out. Number of electrodes and patients per cluster is shown. None of these ROIs showed any periods of significant sensitivity to memorability (all $q > 0.05$ FDR-corrected).

While words, faces, and scenes are perceived as coherent high-level objects they are a combination of multiple lower-level constituent parts. The principles seen here in novel words may transfer across to other high-level visual categories, with more memorable items being similar to rarely seen but otherwise familiar exemplars, and containing constituent parts that are low probability given the context of the whole image. Given the differences seen here between known and novel word memorability, it is likely these principles would be most relevant for novel items without a strong, extant semantic association. Our models based on parameters from the pseudowords showed significant ability to predict pseudoword memorability, however, these models were poor predictors of known word memorability. This is suggestive that known words indeed rely on additional semantic factors to aid in memorisation[5], and are less reliant on lower-level orthographic and lexical neighbourhood effects that are required for pseudoword memory.

Prior fMRI work has shown modulations of activation within vOTC based on memorability of faces and scenes[4], with greater activation for more memorable items. In contrast, here we show a reduction in activation during the presentation of more memorable pseudowords. This could relate to more memorable pseudowords being more word-like, resulting in more word-like activation patterns. However, the neural signatures for lexicality appear much more widespread across the cortex, and the memorability effect is much more focal. This reduction in activation could instead reflect the decreased effort in memorising these pseudowords[15], with a lexical scaffold providing increased efficiency. More memorable items tend to be more efficiently encoded, recognised, and retrieved[5,39,40]. This may also be influenced by the task, as our participants were viewing and encoding the words knowing there would be a subsequent recall required, while fMRI participants in the prior study were just passively viewing the images.

Many recent additions to the Merriam-Webster dictionary (e.g., yeet, rizz, janky, adorkable) follow the trends seen here for highly memorable pseudowords; word-like forms with small orthographic neighbourhoods and the presence of low-frequency bigrams. This could suggest that highly memorable novel words are more likely to propagate between individuals and be incorporated into the shared lexicon.

Here, we show that while known word memorability appears most dependent on semantic memory, novel word memorability appears to be driven by the orthographic lexicon. This suggests that early word learning processes before cortical consolidation begins, are dependent on the orthographic lexicon. This leaves the open question of how the memorability of novel words changes throughout the learning process, as words gradually transition from hippocampal and orthographic lexicon-dependent processes to cortical and semantically driven processes.

Methods

Online experiment

A total of 1804 participants (809 male, 977 female, 18 other, 18–60 years, mean age 38 ± 11 years) provided informed consent to participation through the online experimental platform Prolific. To be recruited, the participants had to be monolingual English speakers, reside in the United States, aged between 18 and 60 years, and with at least a 90% approval rating. Participants were excluded if they did not provide any correct key presses in the memory task, if their false alarm rate was >3.5 SD above the mean (>50%), or if their mean response time was less than 300 ms. The number of participants was determined to get at least 40 ratings for each pseudoword in the corpus. All experimental procedures were reviewed and approved by the Committee for the Protection of Human Subjects (CPHS) of the University of Texas Health Science Center at Houston as Protocol Number HSC-MS-06-0385.

Our test corpus consisted of 2100 unique, phonotactically legal word forms. Pseudoword selection was based on pseudowords used in previous intracranial studies by our lab[17,18,22], pseudowords from the English Lexicon Project with existing lexical decision data[27], and nonsense words from classic literature (e.g., Lewis Carroll, Roald Dahl, Dr. Seuss, A.A. Milne, Spike Milligan). A small number of very low-frequency real words (<100), whose meanings are not widely known, were also included in the test set. All tested stimuli had between 4 and 8 letters and had a SUBTLEXus frequency lower than 1 occurrence per million words.

Participants engaged in a continuous recognition task, viewing a stream of 256 words, and had to press the 'R' key whenever they identified a repeated word. Each word was displayed for 1200 ms with an 800 ms blank inter-stimulus interval. Fifty of these words were targets, where their repetition occurred >30 s after the prime. Twenty

shorter duration (<30 s) repetitions were included to maintain vigilance. All the remaining stimuli were shown only once and were drawn from the same pool of pseudowords. Word selection for each participant was pseudorandom, aiming to use each word as a target for at least 40 participants, and ensuring a >1 orthographic Levenshtein distance between all stimuli within a given set. The experiment was implemented using oTree[41], took approximately 10 min, and participants were compensated $2 for their time.

Memorability for each word was quantified as the Corrected Recognition (CR) score, the Hit Rate minus the False Alarm rate. Trials with a response time faster than 3.5 SD less than the mean (<220 ms) were considered to be misses (0.34% of trials). To test consistency between participants we calculated the split-half consistency across 1000 permutations. For each permutation, the participants were randomly split into two halves, CR was calculated for each word and the Spearman rank correlation between the halves was calculated, corrected by the Spearman-Brown split-half reliability correction. A null distribution was calculated by randomly shuffling CR values between words within one of the participant halves.

Intracranial participants

Thirty-six patients (17 male, 20–66 years, mean age 36 ± 10 years, 2 left-handed, IQ 96 ± 12, Age of Epilepsy Onset 22 ± 12 years) participated in the experiments after giving written informed consent. All participants were semi-chronically implanted with intracranial electrodes for seizure localisation of pharmaco-resistant epilepsy. Participants were excluded if they had confirmed right hemisphere language dominance, only had electrode coverage of the right hemisphere, or had a significant additional neurological history (e.g., previous resections, MR imaging abnormalities such as malformations or hypoplasia). Sample size was determined to provide sufficient coverage (>3 patients) of all investigated regions of interest. All experimental procedures were reviewed and approved by the Committee for the Protection of Human Subjects (CPHS) of the University of Texas Health Science Center at Houston as Protocol Number HSC-MS-06-0385.

Electrode implantation and data recording

Data were acquired from either subdural grid electrodes (SDEs; 2 patients) or stereotactically placed depth electrodes (sEEGs; 34 patients). SDEs were subdural platinum-iridium electrodes embedded in a silicone elastomer sheet (PMT Corporation; top-hat design; 3 mm diameter cortical contact) and were surgically implanted via a craniotomy[42,43]. sEEGs were implanted using a Robotic Surgical Assistant (ROSA; Medtech, Montpellier, France)[44,45]. Each sEEG probe (PMT corporation, Chanhassen, Minnesota) was 0.8 mm in diameter and had 8–16 electrode contacts. Each contact was a platinum-iridium cylinder, 2.0 mm in length and separated from the adjacent contact by 1.5–2.43 mm. Each patient had 12–20 such probes implanted. Following surgical implantation, electrodes were localised by co-registration of pre-operative anatomical 3 T MRI and post-operative CT scans in AFNI[46]. Electrode positions were projected onto a cortical surface model generated in FreeSurfer[47], and displayed on the cortical surface model for visualisation[42]. Intracranial data were collected during research experiments starting on the first day after electrode implantation for sEEGs and two days after implantation for SDEs. Data were digitised at 2 kHz using the NeuroPort recording system (Blackrock Microsystems, Salt Lake City, Utah), imported into Matlab, initially referenced to the white matter channel used as a reference for the clinical acquisition system and visually inspected for line noise, artefacts, and epileptic activity. Electrodes with excessive line noise or localised to sites of seizure onset were excluded. Each electrode was re-referenced to the common average of the remaining channels. Trials contaminated by inter-ictal epileptic spikes were discarded.

Stimuli and experimental design

Patients undertook a silent sentence reading task[18,22]. Participants were presented with eight-word sentences and word lists using a rapid serial visual presentation format. A 1000 ms fixation cross was presented followed by each word presented one at a time, each for 500 ms. Words were presented in all capital letters, in Arial font with a height of 150 pixels (~2.2° visual angle) on a 2880 × 1800, 15.4" LCD screen positioned at eye-level, 2–3' from the patient, using PsychToolbox for MATLAB. To maintain the participants' attention, after each sentence, they were presented with a two-alternative forced choice, deciding which of two presented words was present in the preceding sentence, responding via a key press. Only trials with a correct response were used for analysis. Stimuli were presented in blocks containing 40 real sentences, 20 Jabberwocky sentences, 20 word lists and 20 pseudoword lists in a pseudorandom order. Each participant completed between 2 and 4 blocks.

Signal analysis

Analyses were performed by first bandpass filtering raw data of each electrode into broadband gamma activity (BGA; 70–150 Hz) following removal of line noise (zero-phase 2nd order Butterworth bandstop filters). A frequency domain bandpass Hilbert transform (paired sigmoid flanks with half-width 1.5 Hz) was applied, and the analytic amplitude was smoothed (Savitzky-Golay finite impulse response, 3rd order, frame length of 201 ms). The resultant BGA time course was then downsampled to 100 Hz using a non-overlapping sliding window average. BGA is presented here as percentage change from baseline level, defined as the period −500 to −100 ms before word 1 presentation.

Statistical analysis

For ROI-based analyses, electrodes were tested to determine whether they were responsive during pseudoword presentation. We decided on an analysis window of 200–400 ms post word onset as our prior work shows this interval is critical for pseudoword reading[17,22]. For each electrode, we calculated the Bayes factor (BF) of the mean pseudoword-induced activity for this window, combining across all trials, word positions and experimental conditions, compared to the mean activity in each of two 200 ms pre-sentence baseline windows (−500 to −300 ms, −300 to −100 ms). Electrodes were considered responsive if they showed strong evidence ($\ln(BF_{10}) > 2.3$, BGA > 10%) of a deviation from baseline during stimulus presentation. Of the 3259 electrodes located in left, language-dominant cortex, 457 electrodes (in 32 patients) were considered responsive. ROIs were selected based on prior intracranial studies of reading[17,18,22]. ROI centres were defined on the cortical surface, and all responsive electrodes within a set geodesic radius of this point were included[48]. Centres of mass for each of the left-hemispheric ROIs in Talairach space were as follows: pFus, -40 -59 -13; mFus, -31 -36 -18; aFus, -35 -21 -23; IFG, -41 27 17; pCS, -40 -2 46; IPS, -28 -59 35.

Linear mixed-effects (LME) models were used to dissociate multiple factors modulating BGA over time. Hierarchical random effects were used, with the random effect of individual electrodes grouped by patient. For time-resolved LME analyses, we used LME models at each time point (10 ms resolution) and corrected for multiple comparisons using a Benjamini-Hochberg false detection rate (FDR) threshold of $q < 0.01$.

Surface-based linear mixed-effects (sbLME) modelling

sbLME[18] was used to map electrode activations for each trial onto the standardised population brain surface using each electrode's presumed "recording zone", an exponentially decaying geodesic radius[49,50]. This resulted in a surface-based activation map for each patient for each trial. Linear mixed-effects (LME) models were then used at each vertex of the standardised surface, providing a beta and

t-statistic estimate at each vertex. LME models are an extension on a multiple linear regression, incorporating fixed effects for fixed experimental variables and random effects for uncontrolled variables. Fixed effects used include memorability, word length, orthographic neighbourhood, lexicality, and word position and are stated in the relevant analyses. Models included a random effect of patient, allowing a random intercept for each patient to account for differences in mean response size between patients. Results were thresholded at a t-statistic greater than 2 and coverage of at least 3 patients. Cluster significance was computed at a corrected alpha-level of 0.01, using family-wise error rate corrections for multiple comparisons. The minimum criterion for family-wise error rates was determined by white-noise clustering analysis (Monte Carlo simulations, 1000 iterations) of data with the same dimension and smoothness as that analysed[49].

Linguistic analysis

We quantified word frequency as the base-10 log of the SUBTLEXus frequency[51]. This resulted in a frequency of 1 meaning 10 instances per million words and 4 meaning 10,000 instances per million words. Orthographic neighbourhood (OLD20)[26] was calculated as the mean number of character-edits required to convert a word into its 20 nearest neighbours. The tested neighbourhood consisted of all words from the SUBTLEXus dataset with a frequency >−1 and at least three letters long. Highest and lowest frequency neighbours were determined as the upper and lower bounds of the frequencies of a pseudoword's closest orthographic neighbours. For example, coltrest's closest neighbours each have an edit distance of 2, with contest (freq = 1.27) and coldest (freq = −0.07) representing the upper and lower bounds, respectively. Bigram and quadrigram frequencies were calculated as the mean probabilities of occurrence of each 2- or 4-letter block within each word, based on the frequency of occurrence in the SUBTLEXus corpus.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Trial-level data and derived memorability scores supporting this paper are available at https://doi.org/10.17605/OSF.IO/DFVJY. Intracranial datasets generated from this research are not publicly available due to their containing information non-compliant with HIPAA, and the human participants from whom the data were collected have not consented to their public release. They are available on request from Nitin Tandon (nitin.tandon@uth.tmc.edu).

Code availability

Code for analysing the data is available at https://doi.org/10.17605/OSF.IO/DFVJY.

References

1. Brysbaert, M., Stevens, M., Mandera, P. & Keuleers, E. How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Front. Psychol.* **7**, 1116 (2016).
2. Davis, M. H. & Gaskell, M. G. A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 3773–3800 (2009).
3. Gaskell, M. G. An integrated framework for the learning, recognition and interpretation of words. *Q. J. Exp. Psychol.* **77**, 2365–2384 (2024).
4. Bainbridge, W. A., Dilks, D. D. & Oliva, A. Memorability: a stimulus-driven perceptual neural signature distinctive from memory. *Neuroimage* **149**, 141–152 (2017).
5. Xie, W., Bainbridge, W. A., Inati, S. K., Baker, C. I. & Zaghloul, K. A. Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. *Nat. Hum. Behav.* **4**, 937–948 (2020).
6. Madan, C. R. Exploring word memorability: how well do different word properties explain item free-recall probability? *Psychon. Bull. Rev.* **28**, 583–595 (2021).
7. Aka, A., Bhatia, S. & McCoy, J. Semantic determinants of memorability. *Cognition* **239**, 105497 (2023).
8. Bainbridge, W. A., Isola, P. & Oliva, A. The intrinsic memorability of face photographs. *J. Exp. Psychol. Gen.* **142**, 1323–1334 (2013).
9. Isola, P., Xiao, J., Torralba, A. & Oliva, A. What makes an image memorable? In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 145–152 (2011).
10. Lin, Q., Yousif, S. R., Chun, M. M. & Scholl, B. J. Visual memorability in the absence of semantic content. *Cognition* **212**, 104714 (2021).
11. Hovhannisyan, M. et al. The visual and semantic features that predict object memory: concept property norms for 1000 object images. *Mem. Cogn.* **49**, 712–731 (2021).
12. Roberts, B. R. T., MacLeod, C. M. & Fernandes, M. A. Symbol superiority: why $ is better remembered than 'dollar'. *Cognition* **238**, 105435 (2023).
13. Davis, T. M. & Bainbridge, W. A. Memory for artwork is predictable. *Proc. Natl Acad. Sci. USA* **120**, e2302389120 (2023).
14. Fiebach, C. J., Friederici, A. D., Müller, K. & Von Cramon, D. Y. fMRI evidence for dual routes to the mental lexicon in visual word recognition. *J. Cogn. Neurosci.* **14**, 11–23 (2002).
15. Taylor, J. S. H., Rastle, K. & Davis, M. H. Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychol. Bull.* **139**, 766–791 (2013).
16. Liu, Y. et al. Early top-down modulation in visual word form processing: evidence from an intracranial SEEG study. *J. Neurosci.* **41**, 6102–6115 (2021).
17. Woolnough, O. et al. A spatiotemporal map of reading aloud. *J. Neurosci.* **42**, 5438–5450 (2022).
18. Woolnough, O. et al. Spatiotemporally distributed frontotemporal networks for sentence reading. *Proc. Natl Acad. Sci. USA* **120**, e2300252120 (2023).
19. White, A. L., Kay, K. N., Tang, K. A. & Yeatman, J. D. Engaging in word recognition elicits highly specific modulations in visual cortex. *Curr. Biol.* **33**, 1308–1320.e5 (2023).
20. Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E. & Buchanan, L. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage* **33**, 739–748 (2006).
21. Vinckier, F. et al. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron* **55**, 143–156 (2007).
22. Woolnough, O. et al. Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nat. Hum. Behav.* **5**, 389–398 (2021).
23. Kronbichler, M. et al. The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage* **21**, 946–953 (2004).
24. White, A. L., Palmer, J., Boynton, G. M. & Yeatman, J. D. Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proc. Natl Acad. Sci. USA* **116**, 10087–10096 (2019).
25. Huizeling, E., Arana, S., Hagoort, P. & Schoffelen, J. M. Lexical frequency and sentence context influence the brain's response to single words. *Neurobiol. Lang.* **3**, 149–179 (2021).
26. Yarkoni, T., Balota, D. & Yap, M. Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychon. Bull. Rev.* **15**, 971–979 (2008).
27. Balota, D. A. et al. The English lexicon project. *Behav. Res. Methods* **39**, 445–459 (2007).

28. Coltheart, M. Are there lexicons? *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* **57**, 1153–1171 (2004).
29. Glezer, L. S., Kim, J., Rule, J., Jiang, X. & Riesenhuber, M. Adding words to the brain's visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. *J. Neurosci.* **35**, 4965–4972 (2015).
30. Lochy, A. et al. Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. *Proc. Natl Acad. Sci. USA* **115**, E7595–E7604 (2018).
31. Woolnough, O. et al. Intraoperative localization and preservation of reading in ventral occipitotemporal cortex. *J. Neurosurg.* **137**, 1610–1617 (2022).
32. Woolnough, O. & Tandon, N. Dissociation of reading and naming in ventral occipitotemporal cortex. *Brain* **147**, 2522–2529 (2024).
33. Hirshorn, E. A. et al. Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc. Natl Acad. Sci. USA* **113**, 8162–8167 (2016).
34. Ulicheva, A., Coltheart, M., Grosseck, O. & Rastle, K. Are people consistent when reading nonwords aloud on different occasions? *Psychon. Bull. Rev.* **28**, 1679–1687 (2021).
35. Falk, M. C., Cole, L. C. & Glosser, G. Pseudoword and real word memory in unilateral temporal lobe epilepsy. *J. Clin. Exp. Neuropsychol.* **24**, 327–334 (2002).
36. Woolnough, O. et al. Category selectivity for face and scene recognition in human medial parietal cortex. *Curr. Biol.* **30**, 2707–2715 (2020).
37. Yonelinas, A., Hawkins, C., Abovian, A. & Aly, M. The role of recollection, familiarity, and the hippocampus in episodic and working memory. *Neuropsychologia* **193**, 108777 (2024).
38. Jacobs, J. et al. Direct electrical stimulation of the human entorhinal region and hippocampus impairs memory. *Neuron* **92**, 983–990 (2016).
39. Bainbridge, W. A. & Rissman, J. Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Sci. Rep.* **8**, 8679 (2018).
40. Mohsenzadeh, Y., Mullin, C., Oliva, A. & Pantazis, D. The perceptual neural trace of memorable unseen scenes. *Sci. Rep.* **9**, 6033 (2019).
41. Chen, D. L., Schonger, M. & Wickens, C. oTree—an open-source platform for laboratory, online, and field experiments. *J. Behav. Exp. Financ.* **9**, 88–97 (2016).
42. Pieters, T. A., Conner, C. R. & Tandon, N. Recursive grid partitioning on a cortical surface model: an optimized technique for the localization of implanted subdural electrodes. *J. Neurosurg.* **118**, 1086–1097 (2013).
43. Tong, B. A., Esquenazi, Y., Johnson, J., Zhu, P. & Tandon, N. The brain is not flat: conformal electrode arrays diminish complications of subdural electrode implantation, a series of 117 cases. *World Neurosurg.* **144**, e734–e742 (2020).
44. Tandon, N. et al. Analysis of morbidity and outcomes associated with use of subdural grids vs stereoelectroencephalography in patients with intractable epilepsy. *JAMA Neurol.* **76**, 672–681 (2019).
45. Rollo, P. S., Rollo, M. J., Zhu, P., Woolnough, O. & Tandon, N. Oblique trajectory angles in robotic stereo-electroencephalography. *J. Neurosurg.* **135**, 245–254 (2020).
46. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
47. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).
48. Kadipasaoglu, C. M. et al. Development of grouped icEEG for the study of cognitive processing. *Front. Psychol.* **6**, 1008 (2015).
49. Kadipasaoglu, C. M. et al. Surface-based mixed effects multilevel analysis of grouped human electrocorticography. *Neuroimage* **101**, 215–224 (2014).
50. McCarty, M. J., Woolnough, O., Mosher, J. C., Seymour, J. & Tandon, N. The listening zone of human electrocorticographic field potential recordings. *eNeuro* **9**, ENEURO.0492-21.2022 (2022).
51. Brysbaert, M. & New, B. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* **41**, 977–990 (2009).

Acknowledgements

The authors would like to thank Wilma Bainbridge for her advice and feedback in implementing the online memorability experiments. We express our gratitude to all the patients who participated in this study, the neurologists at the Texas Comprehensive Epilepsy Program who participated in the care of these patients, and the nurses and technicians in the Epilepsy Monitoring Unit at Memorial Hermann Hospital who helped make this research possible. This work was supported by the National Institute of Neurological Disorders and Stroke (U01NS128921 to N.T.).

Author contributions

Conceptualisation: O.W.; Methodology: O.W.; Data curation: O.W.; Software: O.W.; Formal Analysis and Visualisation: O.W.; Writing—Original Draft: O.W.; Writing—Review and Editing: O.W., N.T.; Funding Acquisition: N.T.; Neurosurgical Procedures: N.T.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-57220-y.

Correspondence and requests for materials should be addressed to Oscar Woolnough.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.