



Published in final edited form as:

*Nat Hum Behav.* 2020 September ; 4(9): 937–948. doi:10.1038/s41562-020-0901-2.

## Memorability of Words in Arbitrary Verbal Associations Modulates Memory Retrieval in the Anterior Temporal Lobe

Weizhen Xie<sup>1,\*</sup>, Wilma A. Bainbridge<sup>2,3,\*</sup>, Sara K. Inati<sup>4</sup>, Chris I. Baker<sup>3</sup>, Kareem A. Zaghloul<sup>1,\*</sup>

<sup>1</sup>Surgical Neurology Branch, NINDS, National Institutes of Health, Bethesda, MD 20892, USA

<sup>2</sup>Department of Psychology, University of Chicago, Chicago, IL 60637, USA

<sup>3</sup>Laboratory of Brain and Cognition, NIMH, National Institutes of Health, Bethesda, MD 20892, USA

<sup>4</sup>Office of the Clinical Director, NINDS, National Institutes of Health, Bethesda, MD 20892, USA

### Abstract

Despite large individual differences in memory performance, people remember certain stimuli with overwhelming consistency. This phenomenon is referred to as the memorability of an individual item. It remains unknown, however, whether memorability also affects our ability to retrieve associations between items. Here, using a paired associates verbal memory task, we combine behavioural data, computational modelling, and direct recordings from the human brain to examine how memorability influences associative memory retrieval. We find that certain words are correctly retrieved across participants irrespective of the cues used to initiate memory retrieval. These words, which share greater semantic similarity with other words, are more readily available during retrieval and lead to more intrusions when retrieval fails. Successful retrieval of these memorable items, relative to less memorable ones, results in faster reinstatement of neural activity in the anterior temporal lobe. Collectively, our data reveal how the brain prioritizes certain information to facilitate memory retrieval.

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Corresponding authors: Weizhen Xie ([weizhen.xie@nih.gov](mailto:weizhen.xie@nih.gov)); Wilma A. Bainbridge ([wilma@uchicago.edu](mailto:wilma@uchicago.edu)); Kareem A. Zaghloul ([kareem.zaghloul@nih.gov](mailto:kareem.zaghloul@nih.gov)).

#### Author Contributions

W. X. conceptualized the study and wrote the paper, with advice from K. A. Z., W. A. B., and C. I. B. W. X. proposed the computational model and analyzed the iEEG data. W. A. B. collected and analyzed the online crowd-sourced data. S. K. I. oversaw iEEG data acquisition and provided clinical assessment of iEEG waveforms and seizure focus localization. K. A. Z. performed all surgical procedures and supervised the study. C. I. B. provided additional funding support for the online crowd-sourced study. All authors provided critical comments.

#### Data Availability

Processed data used in this study can be found at: <https://neuroscience.nih.gov/ninds/zaghloul/downloads.html>.

#### Code Availability

Custom code that supports the findings of this study is available from the first author upon request.

#### Competing Interests

The authors declare no competing interests.

Although people's past experiences and overall memory ability tend to vary<sup>1</sup>, human observers consistently remember some stimuli better than others<sup>2-4</sup>. For example, with just a single viewing, some face and scene images are found to be easily remembered while others are easily forgotten<sup>4-6</sup>. This attribute of each visual image has been referred to as its memorability, and the memorability of different images can account for as much as half of the overall variance in visual recognition performance across individuals<sup>4</sup>. Notably, the memorability of visual images cannot simply be explained by variations in lower-level stimulus features such as colour or spatial frequency<sup>5</sup>. Stimulus memorability also does not appear to depend upon an observer's level of engagement with a memory task, such as their attention or level of processing<sup>7</sup>, or upon their familiarity with or subjective preference for a particular item<sup>4</sup>.

While it is not known why some items are more memorable than others, recent neuroimaging studies have demonstrated that activation patterns in the ventral processing stream can reliably differentiate memorable from forgettable images even in the absence of an explicit memory task<sup>5,8,9</sup>. Thus, it is possible that the memorability of visual stimuli may reflect an intrinsic stimulus property that bridges both perceptual processing and memory formation<sup>7</sup>. In this case, memorable images can be more efficiently perceived, encoded, and recognized<sup>5,8,9</sup>. This framework suggests that stimulus memorability may play a critical role in mnemonic representations and processes beyond other factors that have traditionally been found to affect memory performance, such as task context<sup>10,11</sup> and an observer's emotional state<sup>12,13</sup>.

Despite these recent findings, however, contemporary studies of memorability have largely been limited to investigating the role of memorability in the recognition of single items. This raises the question as to whether memorability is only a phenomenon that arises during perception. The extent to which stimulus memorability may affect recall, when the to-be-remembered task content is not explicitly perceived, is less clear<sup>14</sup>. As opposed to recognition, recall critically relies on our ability to form and remember associations between items<sup>15</sup>. Once formed, these associations provide a powerful cue for recalling items from memory, even when no perceptual information regarding the to-be-retrieved content is provided<sup>16</sup>. It remains unclear whether stimulus memorability can also affect memory retrieval based on this recall process, given that memorability of associative task content has received surprisingly little attention in the literature.

Here, we use a paired associates verbal memory task to investigate the role of stimulus memorability in associative memory retrieval. Our goal is to investigate whether stimulus memorability can sufficiently influence memory retrieval, even when successful retrieval of an item occurs in the absence of its perceptual representation and instead is primarily driven by the association formed between the item and its retrieval cue. We use verbal stimuli in this study to take advantage of the highly associative nature of verbal content<sup>17</sup> and to additionally ask whether memorability is also a stimulus property of words, and therefore semantic concepts, rather than strictly visual images. We examine data captured from 30 participants with temporal lobe intracranial electrodes placed for seizure monitoring, and complement these data with online crowd-sourced findings from 2623 participants recruited through the experimental platform Amazon Mechanical Turk. In both cases, we find that

certain words are more successfully recalled across participants, and are therefore more memorable, even when paired with arbitrary retrieval cues.

We next construct a computational model to characterize the memorability of each word based on intrinsic semantic properties of the words themselves. In this model, we hypothesize that semantic relationships among words influence how a word is retrieved from memory<sup>16,18–20</sup>. Because some words tend to be semantically more interconnected with others, they may serve as prior locations in the semantic network to initiate a more efficient memory search during retrieval<sup>21</sup>. As a result, these words would emerge earlier during the memory search process, making them less susceptible to cumulative mnemonic interference, and hence more memorable. We test several core predictions that emerge from this hypothesis. First, computationally, observed memorability of a word should be predicted by a memory search model metric that captures how likely a word would be searched given any arbitrary retrieval cues. Second, behaviourally, memorable words should be more rapidly reported, as they tend to come to mind more easily. However, this would also imply that memorable words can lead to more intrusion errors when retrieval fails. Last, neurally, successful retrieval of memorable words should involve rapid reactivation of initially encoded memory content recorded by intracranial electrodes in semantically imbued brain regions, such as the anterior temporal lobe (ATL)<sup>22</sup>. Together, by testing these predicted effects of stimulus memorability on associative memory retrieval using modelling, behavioural, and electrophysiological data, this study adds theoretical and empirical insights into the research of human memory and the emerging literature on memorability<sup>7</sup>.

## Results

### Memorability of Words in Arbitrary Verbal Associations

Thirty participants (10 females;  $32.73 \pm 1.93$  [mean  $\pm$  s.e.m] years old; Supplementary Table 1) with implanted intracranial electroencephalogram (iEEG) electrodes studied lists of six arbitrary pairs of words in a paired associates verbal memory task. For each participant, the word pairs in each list were constructed by randomly sampling words from a pool of 300 common nouns<sup>23,24</sup>, and thus these word pairs were not identical across participants. Following each study list and a distraction period (~20 seconds), we presented one word from each pair as a retrieval cue in a random order and participants attempted to recall the paired word (i.e., retrieval target; Fig. 1A). Participants on average completed  $216 \pm 16$  total trials and successfully retrieved the correct word on  $30.22\% \pm 4.15\%$  of trials, while vocalizing an incorrect word or intrusion on  $17.01\% \pm 2.89\%$  of trials (also see Supplementary Tables 2 to 4).

To evaluate whether certain target words were successfully recalled across participants, we examined the split-half reliability of memorability estimates of the words in the current study using a resampling procedure<sup>4</sup>. In brief, in each of 5000 iterations, we split the retrieval data across participants into random halves. In each half, we calculated the probability of successful memory recall for each target word across participants. This provides an estimate of memorability for each word when it is the retrieval target. We calculated the Spearman-Brown<sup>25,26</sup> corrected rank-order correlations ( $\rho$ ) of the memorability estimates across all words between the two random halves to quantify the

split-half consistency in each iteration. If words that were remembered by one half of the participants also tended to be remembered by the other half, the split-half consistency measure across iterations should be a positive value on average<sup>27</sup>. We, therefore, used a one-tailed test to evaluate this prediction by comparing the mean of split-half correlation coefficients (Fisher's Z transformed Spearman correlation) across 5000 iterations against surrogate correlation coefficients calculated by shuffling the word ranks in one of the halves (Fig. 1B). We found that this mean estimate of split-half correlation coefficients was significantly larger than surrogate correlation coefficients in the 30 iEEG participants ( $\rho_z = 0.19$ , bootstrapped 95% confidence interval, CI: [0.01, 0.40],  $p = 0.035$ , one-tailed; Fig. 1B).

We further confirmed that the memorability of individual target words was conserved across participants by performing the same analysis using data captured from an online Amazon Mechanical Turk study. In this study, 2,623 participants (1556 female;  $36.27 \pm 0.30$  years old) performed the paired associates verbal memory task with the same word pairs. They on average successfully retrieved the target words on  $51.65\% \pm 0.68\%$  trials. In this online sample, we also found that the memorability of the retrieval target words was also consistent across split halves of the participants, in that the mean of split-half correlations across 5000 iterations was significantly larger than the correlations observed using surrogate word ranks ( $\rho_z = 0.23$ , bootstrapped 95% CI: [0.06, 0.44],  $p = 0.011$ , one-tailed; Fig. 1B). Critically, we performed a correlation analysis between the memorability of words in the two datasets and found that words that were more memorable as retrieval targets in the iEEG sample were also more memorable in the online sample ( $\rho = 0.20$ , 95% CI: [0.09, 0.31],  $p = 0.00047$ , two-tailed,  $n = 300$  words). These data suggest that words, when used as retrieval targets and paired with arbitrary retrieval cues from the word pool, reliably vary in their levels of memorability. This observation is not limited by the clinical characteristics of a special group of participants (patients with epilepsy) in the iEEG sample.

### Modelling Memorability of Words in Arbitrary Verbal Associations

To account for the observations above, we modelled the memorability of target words in the paired associates memory task. In our model, we assume that associative memory retrieval involves a search process, and this process follows a sampling rule based on how semantically similar a target word is to a cue word. This semantic similarity property between target and cue words is referred to as the matching strength between them in previous memory search models<sup>16,20</sup>. According to these models, the likelihood of searching a target word based on a retrieval cue is the ratio of the given cue-and-target matching strength relative to the sum of the matching strengths of the cue with all possible search targets (Equation 1; see Methods). Building upon this likelihood function, memorability of a target word can then be formally defined as the aggregated likelihood of retrieving a certain target word given any arbitrary retrieval cues (Equation 3). That is, a more memorable target word should have a stronger relative matching strength on average with any other words.

We simulated this model by calculating the semantic similarity between every two words as a proxy for their matching strength<sup>18,28</sup>. To do so, we represented each word as a feature vector defined by the Global Vectors (GloVe) for word representation<sup>29</sup>. Based on aggregated global word-word co-occurrence statistics in large text corpuses (e.g., Wikipedia,

newspapers, and books), the GloVe uses an unsupervised learning algorithm to estimate structures of a word's vector space in the natural linguistic environment<sup>29</sup>. The cosine similarity between word vectors may then provide an effective method for measuring the semantic similarity (Equation 2), hence matching strength, of the corresponding words<sup>18</sup>. Accordingly, a more memorable target word that has a stronger relative matching strength on average with any other words should locate more centrally in a semantic space. Based on these measures, we modelled the predicted memorability of each target word using the cue words that were randomly chosen in the current study (Fig. 2A).

We found that the predicted memorability estimates from our model were significantly correlated with the observed memorability values of the 300 words in both the iEEG sample ( $\rho = 0.20$ , 95% CI: [0.09, 0.30],  $p = 0.00057$ , two-tailed) and the online sample ( $\rho = 0.21$ , 95% CI: [0.10, 0.31],  $p = 0.00029$ , two-tailed, Fig. 2B). We next evaluated several alternative accounts to check whether the observed memorability of words could be better explained by other lower-level word features, such as concreteness ratings<sup>30</sup> and word frequency (ranked in the Corpus of Contemporary American English)<sup>31</sup>. As shown in multiple regression analyses that included all predictors simultaneously in a model (see Table 1 and Supplementary Tables 5 and 6 for additional details), the predicted memorability estimates explained a significant amount of variance in the observed memorability scores in both the iEEG sample ( $\beta = 0.19$ , 95% CI: [0.06, 0.32],  $t(296) = 2.97$ ,  $p = 0.003$ , two-tailed) and the online sample ( $\beta = 0.18$ , 95% CI: [0.05, 0.30],  $t(296) = 2.79$ ,  $p = 0.006$ , two-tailed). However, we did not observe significant evidence that word frequency and concreteness could predict observed memorability of words (see Table 1 for details).

### Memorability of Words in Arbitrary Verbal Associations Modulates Memory Retrieval

Given that relative matching strength appears to account for the observed memorability of the retrieved target words, we hypothesized that words with higher overall matching strength should be more readily available during the retrieval search process. We therefore tested core behavioural and neural predictions that directly emerged from this hypothesis.

**Memorable Words Are Reported Faster**—We first tested the prediction that, if memorable items are more readily available during memory search, then those items should be retrieved more quickly irrespective of recall accuracy. To test this prediction, we correlated participants' trial-by-trial response times and observed memorability scores of the vocalized words in the iEEG sample. We limited this analysis to participants who had responded to more than 10 words to obtain a better estimate of the rank-order correlation (range of trial counts: 11 to 265 trials, on average  $115 \pm 13$  trials per participants, see Supplementary Table 2). Using a resampling test at the subject level (random-effect), we found that target word memorability was significantly correlated with shorter response times across participants in the iEEG sample (Fisher's Z transformed Spearman correlation:  $\rho_z = -0.06$ , bootstrapped 95% CI: [-0.09, -0.03],  $p = 0.0004$ , two-tailed,  $n = 27$ ; Fig. 3A; also see Extended Data Fig. 1). Similar analyses limited to correct trials only ( $\rho_z = -0.03$ , bootstrapped 95% CI: [-0.09, 0.02],  $p = 0.23$ , two-tailed) or intrusion trials only ( $\rho_z = 0.12$ , bootstrapped 95% CI: [-0.01, 0.24],  $p = 0.06$ , two-tailed) did not yield statistically significant results. Nonetheless, these results suggest that the response time of associative

memory retrieval is modulated by the memorability of the vocalized items, and that both correctly and incorrectly vocalized items are critical aspects of the overall memory search process that reveal this association.

**Memorable Words Can Lead to More Intrusion Errors**—We next tested the potential cost of target word memorability on memory retrieval. If highly memorable verbal content is indeed more readily available during memory search, then memorable non-target words should more easily intrude into the mind when the retrieval of a target word fails. We tested this prediction by comparing the average observed memorability of intruded words in each participant to the median observed memorability of all words from the word pool. We again limited our analysis to only those iEEG participants who had more than 10 intrusions trials to obtain a reasonable estimate of memorability for the intruded items (range of trial counts: 13 to 93 trials, on average  $37 \pm 4$  trials per participants, see Supplementary Table 3). Using a resampling test at the subject level (random-effect), we found that the average memorability of the intruded words across participants (0.366, bootstrapped 95% CI: [0.357, 0.373],  $n=21$ ) was significantly higher than the median memorability of the entire word pool (0.353; bootstrapped  $p = 0.0016$ , two-tailed; complementary one-sample  $t$ -test:  $t(20) = 3.25$ ,  $p = 0.0040$ , two-tailed,  $r_{\text{equivalent}} = 0.59$  [0.21, 0.81], Fig. 3B). We further confirmed that higher memorability also led to greater intrusions in the online participants. To reduce the influence of the shorter online experiment (3 study and test lists, ~5 min in duration, see Methods) on a lower likelihood of intrusions, we counted the total number of intrusions for each word across participants, normalized by the number of times the word was presented in the entire study, and correlated this measure with observed memorability of each word. We found that the proportion of a word intruding on retrieval was significantly correlated with the memorability of the word in the online participants ( $\rho = 0.13$ , 95% CI: [0.02, .24],  $p = 0.026$ , two-tailed).

**Memorable Targets Are Reinstated Earlier During Memory Retrieval**—Finally, we were interested in examining whether the memorability of the target words was related to the iEEG data. Previous evidence has demonstrated that successful memory retrieval involves reinstating patterns of neural activity that were present during encoding<sup>32,33</sup>. If the memorability of the target word guides memory search, then target word memorability should be related to how quickly neural reinstatement arises during memory retrieval. To examine this prediction, we first used patterns of iEEG power distributed across five frequency bands in all electrodes in the anterior and posterior temporal lobe (ATL and PTL, respectively) to compute neural reinstatement between every encoding and retrieval time point in every trial (200 ms windows, 20 ms step; Figs. 4A,B; see Methods). Similar to previous reports<sup>32,33</sup>, we observed significantly greater neural reinstatement during correct compared with incorrect trials in both the ATL and PTL immediately before recall vocalization (Figs. 4C,E). Based on group-level comparisons (corrected for family-wise error rate using permutation tests at  $\alpha = .05$  level; see Methods), we identified an encoding time region of interest (tROI) that could reliably differentiate correct versus incorrect recall trials, separately for the ATL (420 ms to 2460 ms after cue onset) and for the PTL (220 ms to 3700 ms after cue onset). We then averaged the measures of neural reinstatement across these group-level encoding tROIs to generate a continuous time measure of neural

reinstatement across different retrieval time points for each participant (Figs. 4D,F). Post-hoc comparisons confirmed that these neural reinstatement values averaged across encoding tROIs and the retrieval period from the probe onset to the median response time were significantly higher in the correct, relative to the incorrect, trials in both the ATL ( $t(18) = 3.39$ ,  $p = 0.0032$ , two-tailed,  $r_{\text{equivalent}} = 0.62$  [0.23, 0.84]) and the PTL ( $t(17) = 4.23$ ,  $p = 0.00056$ , two-tailed,  $r_{\text{equivalent}} = 0.72$  [0.38, 0.89]).

For correct trials, we then divided all target words into memorable and forgettable words based on a median split of memorability values across the entire word pool, and separately calculated their continuous time measures of neural reinstatement in the ATL and the PTL (Figs. 5A,C). We found that the overall temporal profile of neural reinstatement for memorable words, as compared with forgettable ones, emerged earlier during the retrieval period in the ATL (Fig. 5A). We computed the latency required to reach 50% of the area under each time series of neural reinstatement<sup>34</sup>, and found that in the ATL, memorable items had a 50% fractional area latency that was significantly shorter than that for forgettable items (bootstrapped means for 50% fractional area latency: 905 ms vs. 1050 ms,  $n = 19$ , bootstrapped 95% CI of latency difference: [-300 ms, -20 ms],  $p = 0.040$ , two-tailed; Fig. 5B). Complementing this, we found that the time series of reinstatement also exhibited significant differences in the peak latency between memorable and forgettable items in the ATL (jack-knife peak latency<sup>35</sup>: 765 ms vs. 1264 ms,  $t(18) = 2.26$ ,  $p = 0.036$ , two-tailed,  $r_{\text{equivalent}} = 0.47$  [0.02, 0.76]). We, however, observed no evidence for a statistically significant difference in the temporal profiles of neural reinstatement in the PTL between memorable and forgettable target words (bootstrapped means for 50% fractional area latency, 887 ms vs. 897 ms,  $n = 18$ , bootstrapped 95% CI of latency difference: [-80 ms, 60 ms],  $p = 0.63$ , two-tailed; jack-knife peak latency: 606 ms vs. 558 ms,  $t(17) = 0.81$ ,  $p = 0.43$ , two-tailed,  $r_{\text{equivalent}} = 0.21$  [-0.29, 0.61]; Figs. 5C,D). Critically, among participants who had both ATL and PTL electrodes ( $n = 18$ ), the differences in latency measures between memorable and forgettable target words were significantly greater in the ATL than that in the PTL (bootstrapped 50% fractional area latency,  $p = 0.031$ , one-tailed; jack-knife peak latency,  $t(17) = 2.47$ ,  $p = 0.012$ , one-tailed,  $r_{\text{equivalent}} = 0.51$  [0.06, 0.79]).

These data suggest that trials with memorable target words demonstrate earlier reinstatement of neural activity. We therefore further investigated this relationship between target word memorability and neural reinstatement by examining trial-by-trial neural reinstatement profiles during an early (between 25% to 50% of response time on a trial) and a late (between 75% to 100% of response time on a trial) retrieval time window (Figs. 6A,B). If memorable items involve earlier reinstatement, then the difference in neural reinstatement between memorable and forgettable items should be more pronounced in the early retrieval time window. However, the later retrieval time window may be dominated by other factors, such as those related to searching for alternative items, deciding to stop memory search, or response preparation. All these factors may be unrelated to the memorability of target words. Indeed, across participants, we found that neural reinstatement in the ATL was significantly correlated with target word memorability in the early retrieval time window ( $\rho_z = 0.07$ , bootstrapped 95% CI: [0.02, 0.13],  $p = 0.0092$ , two-tailed,  $n = 19$ ), but not in the late retrieval time window ( $\rho_z = -0.03$ , bootstrapped 95% CI: [-0.14, 0.07],  $p = 0.60$ , two-tailed,  $n = 19$ ) across trials (range of trial counts: 14 to 242 trials, on average  $76 \pm 14$  trials per

participants, see Supplementary Table 4). Across 18 participants who had PTL electrodes, we found no significant evidence for the correlation between target word memorability and neural reinstatement patterns in the PTL (see Fig. 6C and Extended Data Fig. 2), either in the early ( $\rho_z = 0.004$ , bootstrapped 95% CI:  $[-0.06, 0.07]$ ,  $p = 0.92$ , two-tailed) or late ( $\rho_z = -0.01$ , bootstrapped 95% CI:  $[-0.09, 0.06]$ ,  $p = 0.89$ , two-tailed) retrieval time windows. We captured these spatial (ATL vs. PTL) and temporal (early vs. late response time windows) distinctions in the relationship between target word memorability and neural reinstatement using a focal one-tailed contrast analysis<sup>36</sup>, which allows us to test only one set of outcomes to avoid omnibus arguments and multiple comparisons<sup>36,37</sup>. We found a significantly stronger correlation between target word memorability and neural reinstatement in the early retrieval time window in the ATL as compared with all other conditions (see Methods for details;  $n = 18$ , bootstrapped  $p = 0.0016$ , one-tailed;  $t$  contrast:  $t(17) = 2.88$ ,  $p = 0.0050$ , one-tailed,  $r_{\text{equivalent}} = 0.57$  [0.14, 0.82], see Fig. 6C and Extended Data Fig. 2). These findings provide strong evidence that early neural reinstatement in the ATL is related to target word memorability as participants successfully retrieve associative verbal content.

## Discussion

Using a paired associates verbal memory task across two different samples of participants, we find that certain words are more likely to be successfully retrieved, irrespective of the arbitrary cues used to initiate memory retrieval. Our computational model, which is grounded in semantic similarity among words, well characterizes the memorability of words and makes several core predictions supported by behavioural and neural data. That is, more memorable items are retrieved faster, can lead to more intrusion errors, and exhibit faster dynamics of neural reinstatement in brain regions that are closely involved in semantic cognition such as the ATL<sup>22</sup>. Taken together, these converging findings advance our understandings about stimulus memorability and provide insights into how the human brain prioritizes certain information to facilitate memory retrieval.

Memorability as a mnemonic phenomenon has been studied throughout the history of psychology<sup>3,38</sup>, but only recently has research leveraged findings from stimulus memorability to understand the relationship between memory and perception<sup>7</sup>. Our results advance this line of research in several ways. First, most studies of memorability have focused on visual recognition. Here, we show that associative memory recall, which is normally dominated by a search process based on a retrieval cue, can also be modulated by the memorability of the target item. Second, moving beyond visual recognition, the current paired associates verbal memory task offers a clear separation between perceptual input and retrieved memory content. Our data suggest that memorability of a target item is relevant for memory retrieval even in the absence of any perceptual information regarding the target. Finally, extending previous observations in visual memorability, our data suggest that, like images, words can also be characterized by their memorability. This memorability property of words is governed by their semantic structure, such that memorable words tend to be more centrally located within a semantic space. Collectively, our results add to the growing literature on memorability<sup>7</sup>, which may subsequently have broader implications for developing better materials used for educational<sup>39</sup>, commercial<sup>40</sup>, or clinical<sup>41</sup> purposes.



Similar to previous studies in the verbal domain, we approximated the extent to which a retrieval target matches a cue by using the semantic similarity between them<sup>18</sup>. Based on this approximation, our results provide direct evidence that high dimensional patterns of verbal features present in the natural linguistic environment<sup>29</sup> can give rise to variations in memorability of words. Consequently, the memorability of a word is related to the semantic similarity between that word and all other words, as opposed to other lower-level verbal features of a single word such as concreteness or word frequency<sup>21,30,42,43</sup>. An important theoretical contribution of our work, then, is in providing a formal mathematical model that well characterizes the memorability of words that emerges from the natural linguistic environment. This model may further provide a stable base for reaching upward towards a broader understanding of memorability for other complex and highly associative natural stimuli, such as content-rich images and videos<sup>44</sup>.

An equally important contribution of our work is in providing insights into the process of memory search. Models of associative memory posit that retrieval of associative content occurs through a search process that is guided by the similarity between a target item and its retrieval cue, referred to as the matching strength between them<sup>16,18–20</sup>. We observed converging behavioural and neural evidence that stimulus memorability can influence this search process. Rather than using a random walk during the initial period to begin memory search as assumed by traditional models<sup>45</sup>, our findings suggest that memory search begins in regions of the search space where memorable items reside. In this manner, memorability may be an important phenomenon for supporting optimal foraging during memory retrieval<sup>18,46</sup>. This may be particularly relevant for recall when the to-be-retrieved content is not perceptually available, and for other higher cognitive functions that critically rely upon information sampling, such as fluid intelligence and memory-based decisions<sup>47</sup>.

Behaviourally, our mathematical model describing the memorability of words as their relative matching strength with any arbitrary retrieval cues provides two core predictions. First, if memory search indeed begins with memorable words, it should take a shorter time to recall memorable words as compared with forgettable ones<sup>48</sup>. Considering that both correctly recalled items and incorrect intrusions reflect the overall memory search process, this predicted relation between word memorability and response time may exist when correct and incorrect recall trials are combined. Our behavioural data support this prediction. Second, if memory search begins with memorable items, memorable items should also more easily intrude into the search process when retrieval fails. Our data showing that intruded items have a higher level of memorability support this prediction. These behavioural observations, therefore, suggest that memorable and forgettable words indeed differ in their retrieval profiles.

Neurally, using iEEG, our data further reveal the fine-scale neural dynamics underlying the retrieval search process of memorable words in the human brain. Successful memory retrieval entails reinstatement of neural patterns of activity that were present during encoding<sup>32,33,49</sup>. Our model provides a core prediction regarding this process of neural reinstatement. Namely, if memory search prioritizes memorable items, memorable target words should more rapidly trigger reinstatement of the original memory experience to support successful recall. The finding that successful retrieval memorable, relative to

forgettable, items show faster neural reinstatement support this prediction. Critically, we observed the relationship between target word memorability and the speed of neural reinstatement in correct trials only, thus separating stimulus-driven memorability properties from generic individual differences in overall memory ability<sup>5</sup>.

Furthermore, we also found that the observed relationship between neural reinstatement and target word memorability was primarily localized to the ATL but not to the PTL. This observation is consistent with the prediction from our model that memorability of associative verbal content is driven by the structural relationship among interconnected semantic representations. According to this prediction, temporal differences in neural reinstatement between the retrieval of memorable and forgettable words should be more pronounced in semantically imbued brain regions such as the ATL<sup>22</sup>. However, this does not necessarily imply that memorability effects for all associative task content are specific to the ATL. For example, our current observation may simply reflect the dominant role of the ATL in semantic cognition in general<sup>22</sup>. Meanwhile, the PTL may be more sensitive to the memorability of associative visual content, given the critical role of the ventral stream in visual processing<sup>5</sup>. Alternatively, the ATL may serve a generic role for linking associative task content independent of the modality of a stimulus<sup>50</sup>. In this case, while the PTL may be involved in semantic representation at a single item level<sup>51,52</sup>, the ATL may play a larger role in associative representations. Future research can articulate these issues to better understand how the human brain processes memorability of relational information from different sensory inputs.

Although our data demonstrate that stimulus memorability indeed can modulate associative memory retrieval, it should be noted that the focus of our analyses has been on the memorability of target words in arbitrary verbal associations. This is fundamentally different from the memorability of the arbitrary verbal associations themselves. While the former addresses the degree to which certain items can be more successfully remembered in associative memory, the latter is directly related to the degree to which pairs of items are more memorable. Future research utilizing a smaller, exhaustive set of associative pairs may be able to reveal further memorability phenomena in associative memory. It should also be noted that the current study has collapsed data across left and right ATLs in the analyses, given that two ATLs may act as a coupled bilateral system to support semantic memory<sup>53</sup>. However, this does not preclude the possibility of hemisphere preference in the representation of memorability for certain verbal content. Future research with a larger sample that has bilateral ATL coverage will be more suitable to address this issue.

In conclusion, our data combine converging evidence from laboratory, online crowd-sourced, computational modelling, and iEEG methods to provide a mechanistic explanation of how stimulus memorability affects associative memory retrieval. We find that memorability of associative verbal content is grounded in the semantic similarity structure of words, and that memorable words can modulate the retrieval process in the ATL. By revealing how our internal experiences are profoundly shaped by intrinsic properties of stimuli from the external world, our results shed light on how rich relational information in everyday life can be efficiently remembered in the human brain.

## Methods

### Participants

Thirty participants (10 females;  $32.73 \pm 1.93$  [mean  $\pm$  s.e.m.] years old; Wechsler Intelligence Quotient [IQ]:  $87.65 \pm 2.06$ , all  $>70$ ; see Supplementary Table 1) with drug resistant epilepsy underwent a surgical procedure at the Clinical Center of the National Institutes of Health (NIH, Bethesda, Maryland, USA). In this procedure, platinum recording contacts were implanted subdurally on the cortical surface and within the brain parenchyma to localize epileptogenic brain regions. In all cases, the clinical team determined the placement of the contacts to best localize epileptogenic regions. All participants completed the English version of a paired associates verbal memory task. Part of the data from this sample (less than 30% overlap samples) has been published in previous studies examining other memory related phenomena<sup>54</sup>. We used the behavioural data from all 30 participants for the group-level analysis of memorability estimates. Inclusion criteria for the analysis of neural data were more restrictive: 1) each participant should have a structural MRI (Magnetic Resonance Imaging) scan to allow proper electrode localization<sup>55</sup>, 2) each participant should have at least 3 electrode contacts in the brain regions of interest (anterior and posterior temporal lobes; ATL and PTL) after removal of noisy electrodes (see Pre-processing steps below); 3) each participant should complete more than 10 correct trials<sup>33</sup> of the paired associates memory task across experimental sessions to ensure reasonable signal stability for the neural reinstatement analysis (see Extended Data Fig. 3); and 4) each participant should have no prior resection of brain regions. In the end, data from 19 participants met these criteria for ATL analyses, among which 18 participants met the criteria for PTL analyses (highlighted by an asterisk in Supplementary Table 1). The Institutional Review Board at the NIH approved the current study. All participants and/or their guardians provided informed consent prior to data collection. Except where otherwise noted, all computational analyses were performed using custom written MATLAB codes (MathWorks, Natick, MA).

For the online replication study, we recruited 3620 participants (2074 female;  $35.68 \pm 0.25$  years old) from the online crowd-sourcing experimental platform Amazon Mechanical Turk, following guidelines set by the NIH Office of Human Subjects Research Protections. Only participants who indicated English as their native language and who provided responses on at least three trials were included in the analyses, resulting in a final sample size of 2623 (1556 female;  $36.27 \pm 0.30$  years old) for further analysis. All participants were compensated for their time. Note, no statistical methods were used to pre-determine sample sizes. That said, our sample size for the iEEG study is larger than previous iEEG studies using the same behavioural task<sup>32,49,54</sup> and our sample size for the online study is similar to previous online cloud-source studies on memorability<sup>4,14</sup>.

### Paired Associates Verbal Memory Task

**iEEG Participants**—Each participant performed the paired associates verbal memory task using stimuli randomly chosen from a pool of 300 common nouns<sup>23,24</sup>. This task was divided into a set of lists, which included a study phase and a test phase each. During the study phase of a list, participants sequentially saw 6 word pairs and were instructed to

remember the arbitrary associations between them. On each trial of the encoding period, each word pair was preceded by a short fixation period (250–300 ms) with a cross on the screen, followed by a blank inter-stimulus interval (ISI) between 500–750 ms. Word pairs were presented stacked in the center of the screen for 4000 ms followed by a blank ISI of 1000 ms. To interrupt participants' active maintenance of the word pairs prior to the test phase, participants were probed to complete an arithmetic distractor task of the form 'A + B + C = ?' for about 20 seconds after the study phase. In the subsequent test phase, participants were cued with one word from each pair selected in a random order. They were instructed to say aloud the associated word, with their responses recorded by a microphone. Each cue word was presented for 4000 ms followed by a blank ISI of 1000 ms, making a total of 5000 ms as the maximum recall time window. Participants could vocalize their response any time during the recall period after cue presentation. They were instructed to vocalize 'PASS' if they could not retrieve the target word. After the study, trained research assistants manually designated each recorded response as correct, intrusion, or pass. A response would be designated as 'PASS' when no vocalization was made or when the participant vocalized the word 'PASS.' We also recorded participants' response times as the time lapsed before they made the first vocalization following the probe onset. A single experimental session contained 60 to 150 total word pairs, or trials (i.e., 10 to 25 lists), depending on how many trials a participant was able to complete in a single recording session (no more than 1 hour per session). We included at most two unique sessions from each participant to minimize potential practice effects or the influence of over-familiarization of task content on memorability estimates as the number of tested sessions increased. Due to the large number of permuted combinations of word pairs, the likelihood that a cue-target pairing is exactly the same across at least two trials within a participant or the same across at least 2 participants is very small. In the scenario when there was within-subject repetition, we averaged the participant's performance in repeated trials for the target word. This only happened in less than 1% of all trials in the current study.

**Online Participants**—We conducted a replication of the verbal paired associates memory task online on the Amazon Mechanical Turk experimental platform with a large sample of participants in order to examine whether memorability estimates are consistent across different populations and experimental settings. Amazon Mechanical Turk has been shown to be a reliable method for collecting large-scale psychological data, usually with broader demographic reach and comparable effect size relative to in-lab samples<sup>56</sup>. The methods for the current experiment were nearly identical to the in-lab paired associates memory task described above. Participants studied a list of 6 word pairs, performed an arithmetic distractor task, and then underwent a test phase. Timing was identical to the experiment for the iEEG participants, although the arithmetic distractor task differed slightly in that it asked for the summation of sets of 2 numbers and lasted 16 seconds. For the test phase, participants saw a cue word and were given as much time as needed to type in the associated target word, to accommodate for individual differences in typing speed. For words they could not recall, participants were instructed to write 'PASS' in order to proceed to subsequent trials. Each participant was tested on 3 study-and-test lists each, resulting in an experiment time of approximately 5 min. Before beginning the actual study, they underwent training using example word pairs that were not used in the main experiment. All word pairs

were randomly sampled from the list of word pairs tested with the iEEG participants, so that there were 200 responses per target word, with approximately 10 different possible matching cue words.

### Intracranial EEG Recordings and Pre-processing

We recorded iEEG signals from subdural electrodes (PMT Corporation, Chanhassen, MN) sampled at 1000 Hz using a Nihon Kohden or a Blackrock Microsystems (Salt Lake City, UT) EEG data acquisition system. These subdural contacts were arranged in both grid and strip configurations with an inter-contact spacing of 5 or 10 mm. The raw EEG traces were initially referenced to a common subcutaneous contact or to the system reference. Localization of electrode contacts was achieved by co-registering postoperative CT and preoperative MRI images using established methods<sup>55</sup>. We then projected the resulting contact locations to the cortical surface of a reconstruction of each individual participant's brain and a normalized cortical surface in the MNI space for visualization.

We analyzed data from 880 subdural contacts (range from 10 to 116, on average  $46 \pm 6$  per participant, see Supplementary Table 4) in brain regions (ATL and PTL) applicable to our study (Fig. 3B). We selected electrode contact for inclusion in analysis through a combination of cortical parcellation using FreeSurfer labels and using a customized procedure detailed in a previous study<sup>57</sup>. Specifically, we first identified temporal electrodes using subject-specific FreeSurfer labels generated for each electrode location. Next, we identified ATL electrodes based on the average surgical cut from participants who had received an anterior temporal lobectomy for clinical indications. Hence, the ATL was surgically defined as shown in a previous study<sup>57</sup>. The remaining temporal lobe contacts were then designated as PTL electrodes.

For each participant, iEEG data were pre-processed separately for each session in two steps. In the first step, we rejected electrodes exhibiting abnormal signal amplitude or large line noise<sup>58,59</sup>. Specifically, we divided each session into one-second epochs and for each electrode calculated the amplitude and variance of the continuous time series over each epoch. Any electrode with a voltage trace whose average amplitude or variance across epochs was more than 3 standard deviations away from the mean across all electrodes was flagged for rejection. We iteratively repeated this procedure until no electrode was rejected. We then applied a local detrending procedure to remove slow fluctuations from the time series of each electrode and used a regression-based approach to remove line noise at 60 Hz and 120 Hz. To approximate a reference free montage, we subtracted the common average reference, calculated using the retained electrodes in each participant within each session, from the voltage trace of each electrode.

In the second step, we rejected additional electrodes and trials that exhibited excessive signal kurtosis or variance during epochs of individual trials using a procedure we adapted from FieldTrip<sup>60</sup>. For each electrode channel and trial, we computed the variance of the voltage trace during two 8000 ms epochs (-2000 ms to 6000 ms), time-locked separately to the onset of study pair and the onset of retrieval cue of the same words, with 1000 ms buffer data each included at the beginning and at the end of an epoch. This resulted in a two-dimensional matrix of variance measures, channels by trials for each epoch, from which we

identified the maximum variance for each trial and the maximum variance for each channel. We identified trial outliers by setting a threshold,  $Q3 + w \times (Q3 - Q1)$ , where  $Q1$  and  $Q3$  are the mean voltage boundaries of the first and third quartiles, respectively. We empirically determined the weight  $w$  to be 2.3<sup>54</sup>. These two steps aimed to remove iEEG signals contaminated by epileptic activity, physical movement of the participant, or external sources of transient electrical perturbations.

We quantified spectral power and phase in each temporal epoch (encoding and retrieval windows) by convolving the pre-processed iEEG signals with complex valued Morlet wavelets (wavelet number 6)<sup>54</sup>. Specifically, we calculated spectral power using 30 logarithmically spaced wavelets between 3 and 150 Hz. We then squared and log-transformed the continuous-time wavelet transform to generate a continuous measure of instantaneous power. Next, we  $z$ -scored the power values separately for each frequency and for each session using the mean and standard deviation of all respective values between  $-700$  ms and  $-500$  ms prior to probe onset from that session<sup>61</sup>. Finally, we removed the 1000 ms buffered data at the beginning and at the end of each epoch and extracted  $z$ -scored power traces for each frequency and each electrode during task-related periods for later analyses (see Data Analysis).

### Modelling Memorability of Associative Concepts based on Word Similarity

We simulated memory retrieval for associative memory based on principles that are in line with the Search for Associative Memory model<sup>16</sup> and the Adaptive Control of Thought-Rational model<sup>20</sup>. These models posit that memory recall is achieved by searching associative memory content. Hence, the search process is contingent upon the extent to which the search target is similar to the search cue. This can be quantified as the similarity, or matching strength, between a cue and its target. In the current cued-recall paired associates verbal memory paradigm, each target item  $I_i$  ( $i = 1$  to  $N$ ) in the memory search space may be probed, with equal probability, by a set of independent cue items,  $Q_j$  ( $j = 1$  to  $M$ ). Hence, the likelihood of retrieving  $I_i$  given a cue  $Q_j$  is related to the matching strength between them,  $S(I_i, Q_j)$ . We defined the relative matching strength between  $I_i$  and  $Q_j$  by comparing their matching strength with the sum of all matching strengths between the same cue,  $Q_j$  and all possible target items in the memory search space<sup>16</sup>, namely all possible words in the current word pool,  $k \in 1$  to  $N$ :

$$P(I_i | Q_j) = \frac{S(Q_j, I_i)^{\beta_j}}{\sum_k^N S(Q_j, I_k)^{\beta_j}} \quad (1)$$

where  $\beta_j$  represents the saliency (or attention weight) of a given cue. We set  $\beta$  as 1 for all cues, similar to an early memory search model<sup>16</sup>. We calculated the matching strength for retrieving target  $I_i$  using cue  $Q_j$ ,  $S(I_i, Q_j)$ , as the cosine similarity of the feature vectors that represent the target and cue items:

$$S(I_i, Q_j) = \frac{I_i \cdot Q_j}{\|I_i\| \|Q_j\|} \quad (2)$$

where we defined a feature vector to represent each word in the word pool by using the GloVe for word representations, trained on Wikipedia 2014 and Gigaword 5<sup>29</sup>.

As the selection of retrieval cues for each memory target is random within session and across participants, the probability of retrieving  $I_i$  across different retrieval cues can be deemed as independent from one another. Hence, an item's simulated memorability can be approximated as the weighted sum of that item's relative matching strength to all cues:

$$Mem(I_i) \approx \sum_j^M P(I_i|Q_j)P(Q_j) \quad (3)$$

We calculated the simulated memorability of each word as a retrieval target using the actual cue words in the current study. Because word pairs were randomly assigned in each experimental session using the same word pool, each retrieval target word did not necessarily have the same cue across participants. Hence, for each retrieval target word, the number of actual cue words used to estimate memorability varies from 7 to 24 (average 18 cue words for each target). We also calculated the hypothetical memorability of each word as a retrieval target in the chosen pool, probed by all other possible cue word ( $M = 299$  based on the current word pool). These two estimates are highly correlated ( $\rho = 0.68$  [0.61, 0.73],  $p < 0.0001$ , two-tailed), suggesting that the identities of randomly chosen cue words may not play a critical rule in the memorability estimates of target words. Hence, the overall matching strength structure among words can be approximated by using arbitrary cues from the word pool.

## Data Analysis

**Consistency Analysis for Observed Memorability of Associative Verbal Content Across Participants**—We calculated the memorability of each target item as the probability of successful retrieval across participants for each word when it was used as a target item in the task. To evaluate the reliability of this measure, we conducted a split-half consistency analysis. First, we split data into random halves that included non-overlapping participants. We then separately calculated the memorability of each target word in each random half. Finally, we calculated the rank-order Spearman correlation of memorability across words between the two different halves, and applied a Spearman-Brown correction to obtain the split-half correlation coefficient<sup>25,26</sup>. We repeated this procedure 5000 times and used the mean correlation coefficient of these iterations as the representative split-half reliability measure. Before aggregating these bounded correlation coefficients across different iterations, we standardized the rank-order correlation coefficients using Fisher's  $Z$  transform function<sup>36</sup>. In principle, this split-half reliability measure across iterations should be a positive value on average<sup>27</sup>. To generate a surrogate distribution for significance testing, for each iteration, we calculated an additional rank-order correlation coefficient after randomly shuffling the word memorability rank from one of the split-halves. We then compared the representative split-half reliability measure, namely the mean, against this surrogate distribution to obtain a one-tailed  $p$  value that reflects the probability of obtaining an average split-half correlation coefficient from the observed data that is greater than the null value if the null hypothesis is true (i.e., 0 correlation in the shuffled data).

**Metrics of Neural Reinstatement**—We quantified the extent to which patterns of neural activity were reinstated between encoding and successful memory retrieval using methods established in previous studies<sup>32,33,49</sup>. Briefly, we binned the continuous time  $z$ -scored power in each frequency into 200-ms epochs spaced every 20 ms (90% overlap) and averaged the instantaneous power over each epoch. For each temporal epoch, we averaged the  $z$ -scored power across five frequency bands: theta (4–8 Hz), alpha (8–16 Hz), beta (16–32 Hz), gamma (32–70 Hz), and high-frequency broad band (70–150 Hz). For every encoding ( $E_i$ ) and retrieval ( $R_j$ ) temporal epoch in each trial, we then constructed feature vectors using the average  $z$ -scored power values from the five frequency bands for every electrode within a chosen brain region:

$$E_i = [z_{1,1(i)} \dots z_{1,K(i)} \dots z_{K,F(i)}]$$

$$R_j = [z_{1,1(j)} \dots z_{1,F(j)} \dots z_{K,F(j)}]$$

Where  $z_{k,f(i)}$  is the  $z$ -scored power of electrode  $k = 1$  to  $K$  at frequency band  $f = 1$  to  $F$  in temporal epoch  $i$ . Hence, each temporal epoch contained  $K \times F$  features, which represent the distributed spectral power across all electrodes and across 5 frequency bands at a single moment in time. The reinstatement for a given study-and-test trial,  $C_n$ , for each encoding epoch  $I$  and retrieval epoch  $J$  can thus be quantified as the cosine similarity between  $E_i$  and  $R_j$ :

$$C_n(i,j) = \frac{E_i \cdot R_j}{\|E_i\| \|R_j\|}$$

Computing reinstatement for every combination of encoding and retrieval epoch generates a reinstatement map for each trial. We then computed the average reinstatement map in each participant separately for all correct and incorrect trials. We further compared these reinstatement maps across participants at each encoding and retrieval time point to identify temporal regions exhibiting a significant difference in neural reinstatement between correct and incorrect trials (see Statistical Analyses).

### **Neural Reinstatement Profile for Memorable versus Forgettable Retrieval**

**Targets**—To account for unbalanced counts of correct trials for memorable and forgettable target words, for each participant, we resampled the data with replacement over 5000 iterations from each trial category (memorable and forgettable targets) based on the same trial count that came from the category with a lesser trial count. We then took the average neural reinstatement profile across these 5000 iterations as the representative neural reinstatement of a given participant for a certain trial type. To obtain group-level measures of fractional area latency, we resampled the participants' data, with replacement, over 5000 iterations. For each iteration, we calculated the percentage of area under the neural reinstatement waveform across different response time points based on participants' average response time within that iteration. Because participants had different response times, we normalized the data as the percentage of average response time along the temporal



dimension before aggregating the data across participants. We then quantified the time it took for the average neural reinstatement profile to reach 50% of its area under the waveform from the cue onset until the average response time as the 50% fractional area latency. Complementary to this bootstrapping approach, to measure peak latency for neural reinstatement profile (Fig. 5), we used a leave-one-out jack-knife approach with proper statistical correction recommended by previous studies<sup>35,62</sup>.

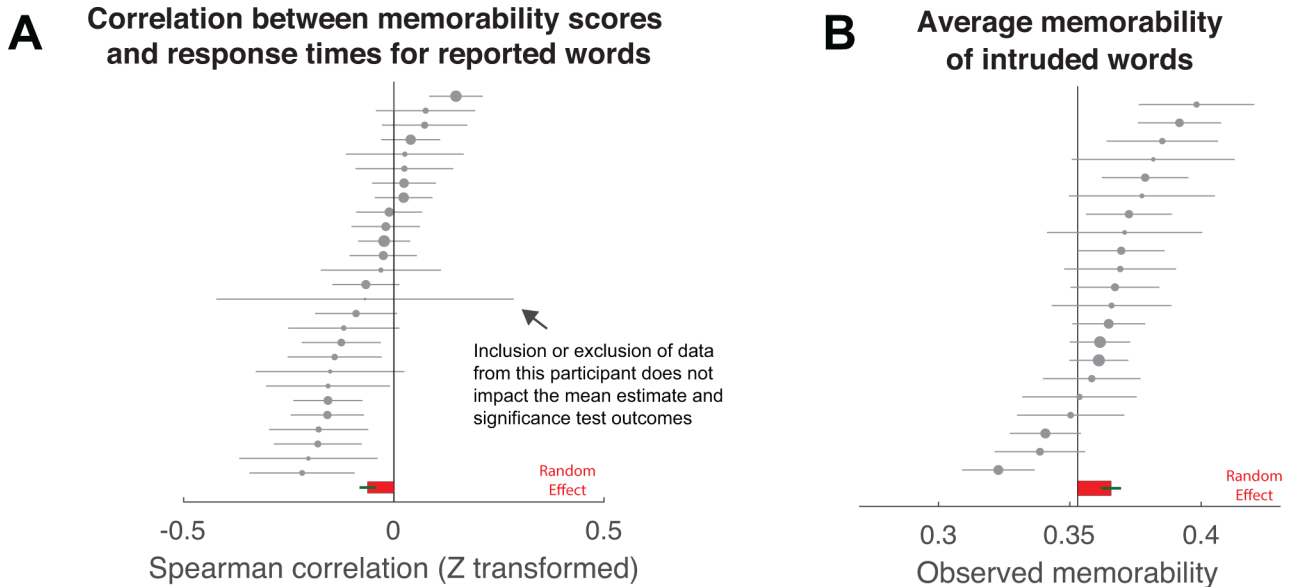
**Statistical Analyses**—Data collection and analysis were not performed blind to the conditions of the experiments. For both behavioural and neural data, we primarily used non-parametric resampling tests to obtain random-effect  $p$  values for effect size measures (e.g., correlation or mean difference) that were generalizable across participants. In brief, we resampled participants' data with replacement, for 5000 iterations, to obtain a sampling distribution of the estimated effect size for a certain measure along with its 95% confidence interval (CI) and  $p$  values. This procedure minimized the assumptions imposed on statistical tests. In general,  $p$  values reported here are two-tailed. When appropriate, we used a one-tailed  $p$  value to test directional significance (e.g., for split-half consistency analysis<sup>27</sup>, post-hoc tests, and contrast analysis<sup>36</sup>), and complemented non-parametric test results with parametric test results (e.g.,  $t$ -tests). In these parametric tests, data distribution was assumed to be normal, but this was not formally tested. Based on these parametric test results, we also reported a common scale effect size estimate,  $r_{\text{equivalent}}$  and its 95% CI<sup>36,63,64</sup>.

To identify temporal regions exhibiting a significant difference in reinstatement between correct and incorrect trials (Fig. 4), we used a combination of a permutation test and cluster-wise correction procedure. Specifically, for each pair of epochs in the reinstatement analysis, we computed the  $t$ -statistic and  $p$ -value across participants between correct and incorrect trials<sup>54</sup>. We then randomly permuted the participant-specific averages (correct vs incorrect) 5000 times to generate the empirical distribution for every epoch. We calculated  $t$ -statistics for each of the permuted epochs. To correct for multiple comparisons, we identified clusters containing epochs that were adjacent in time that exhibited a significant difference between trial types (in each epoch,  $p < 0.01$ ). For each cluster of significant epochs identified in the true and permuted cases, we defined a cluster statistic as the sum of the  $t$ -statistics within that temporal cluster. We retained the maximum cluster statistic during each of the 5000 permutations to create a distribution of maximum cluster statistics. We assigned  $p$ -values to each identified cluster of the true data by comparing its cluster statistic to the distribution of maximum cluster statistics from the permuted cases. Clusters were determined to be significant if the  $p$  value calculated at the cluster level was less than 0.05.

To evaluate specific spatial and temporal patterns of neural reinstatement profiles in relation to trial-by-trial memorability scores, we performed a focal contrast analysis<sup>36,37</sup>. In this analysis, we predicted that the correlation magnitude between memorability scores and neural reinstatement values in the early retrieval time window of the ATL should be higher than that in all the other conditions, namely the late retrieval time window of the ATL and both the early and late retrieval time windows of the PTL. We thus assigned the contrast weights, +3, -1, -1, and -1, respectively to these four conditions. This set of contrast weights should sum to be 0, reflecting a theoretical null result. However, if the data follow our prediction, then the weighted sum of observed data should on average yield a positive

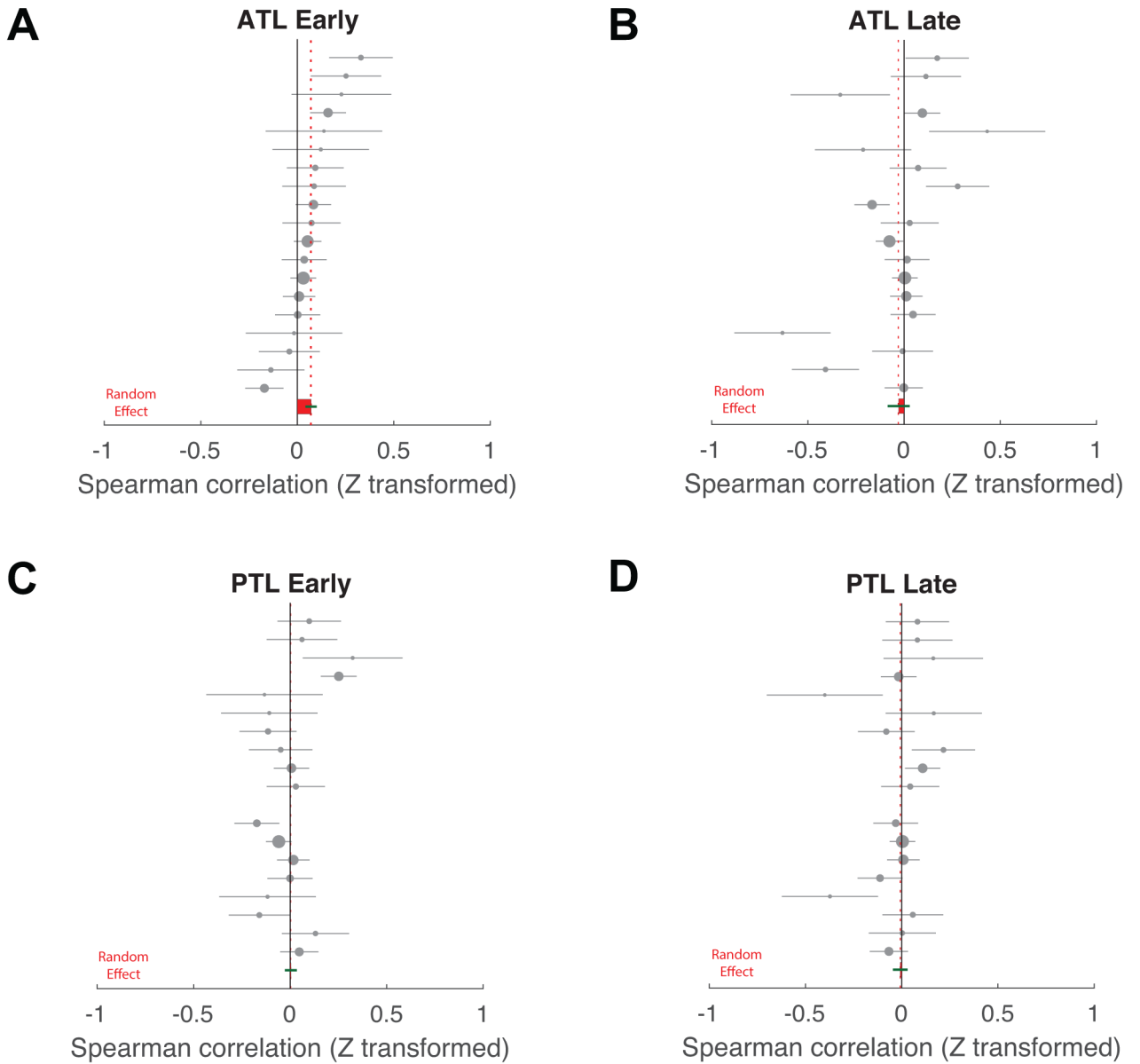
value across participants, and therefore can be tested using a one-tailed test using either a non-parametric bootstrap analysis or a parametric one-sample  $t$ -test<sup>36,37</sup>. This focal contrast analysis would thus allow us to draw a stronger inference on the particular spatial-temporal neural reinstatement pattern predicted by our hypothesis.

## Extended Data



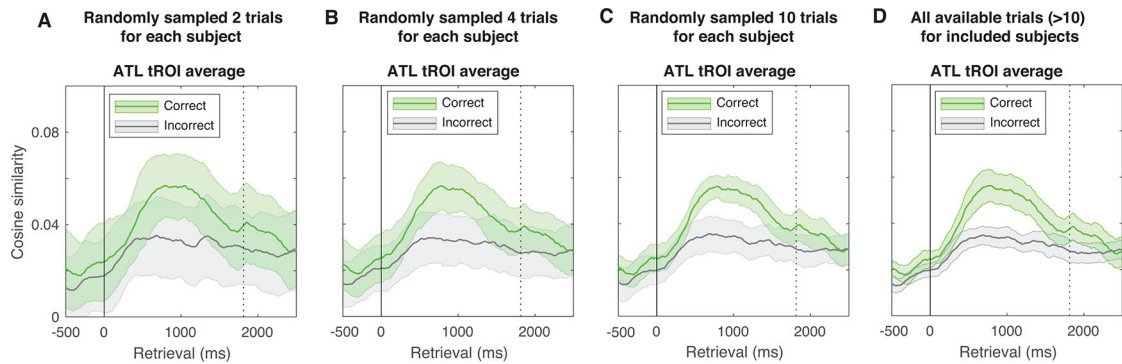
**Extended Data Fig. 1. Memorable words are retrieved more quickly but lead to more intrusion errors across individuals.**

(A) Participants' values for Spearman correlation (Fisher's  $z$  transformed) of the relationship between target words memorability and the response times of retrieved words and (B) average memorability of intruded words across participants in the iEEG sample. Each dot indicates a value from a single participant, with the whiskers indicating the within-participant standard error estimate across trials. The dot sizes are weighted by the overall within-participant standard error, with a larger size indicating smaller variability. The data are sorted by participant-specific estimates separately for (A) and (B). The random-effect mean estimates (in red) and their standard errors (in green) between participants are plotted at the bottom, which are identical to the bars shown in Fig. 3. Although there is a noisy estimate in (A) due to a low trial count (11 trials), inclusion or exclusion of this participant's data does not substantially impact the mean estimate and significant testing across participants.



**Extended Data Fig. 2. Correlation estimates (Fisher’s z transformed) for the association between trial-by-trial memorability of correctly retrieved items and neural reinstatement in the ATL and PTL.**

(A) Data across participants in the ALT during the early retrieval time window. (B) Data across participants in the ALT during the late retrieval time window. (C) Data across participants in the PLT during the early retrieval time window. (D) Data across participants in the PLT during the late retrieval time window. Each dot indicates a value from a single participant, with the whiskers indicating the within-participant standard error estimate across trials. The dot sizes are weighted by the overall within-participant standard error, with a larger size indicating smaller variability. All data are sorted by participant-specific correlation estimates based on (A). The random-effect mean estimates (in red) and their standard errors (in green) across participants are marked at the bottom of each plot, which are identical to the bars shown in Fig. 6C.



### Extended Data Fig. 3. Neural reinstatement effect stabilizes over around 10 trials.

(A) Resampling without replacement of the current dataset over 100 interactions with 2 trials per condition (i.e., 2 for correct and 2 for incorrect retrieval) per subject, (B) 4 trials per condition per subject, (C) 10 trials per condition per subject (10 trials), (D) and all available trials for included subjects. Intuitively, the more trials were included, the less noisy the data were. When the number of resampling trials reached to 10, the amount of variance in the estimate of mean neural reinstatement pattern for correct responses was similar to the data from all available trials from all included participants. This resampling analysis provides some analytical support for the trial count criterion we have imposed on the analysis.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Alex Martin, John H. Wittig Jr., Vishnu Sreekumar, Julio I. Chapeton, Christopher Zawora, and Alex Vaz for insightful comments on the project. This work was supported by Intramural Research Programs of the National Institute for Neurological Disorders and Stroke (ZIA-NS003144) and the National Institute for Mental Health (ZIA-MH002909). Weizhen Xie was funded by the National Institute for Neurological Disorders and Stroke Competitive Postdoctoral Fellowship Award. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We are indebted to all patients who have selflessly volunteered their time to participate in this study.

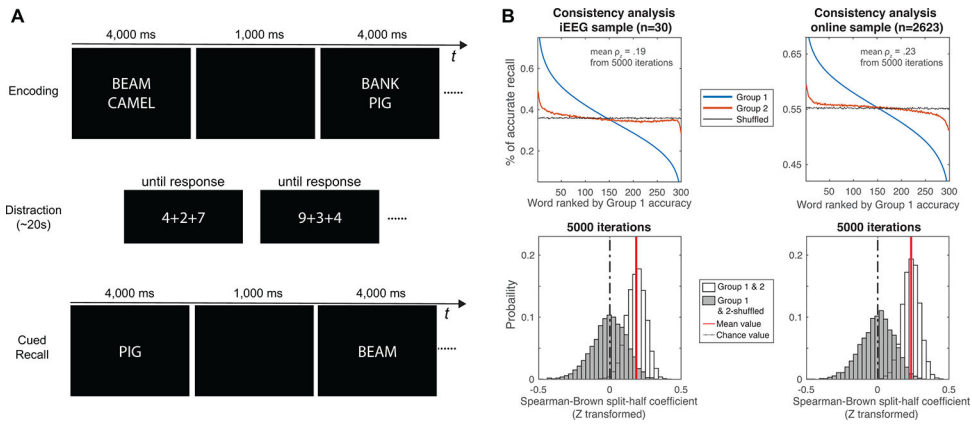
## References

1. Unsworth N Individual differences in long-term memory. *Psychol. Bull* 145, 79–139 (2019). [PubMed: 30596433]
2. Salthouse TA Item analyses of memory differences. *J. Clin. Exp. Neuropsychol* 39, 326–335 (2017). [PubMed: 27618285]
3. Rubin DC Memorability as a measure of processing: A unit analysis of prose and list learning. *J. Exp. Psychol. Gen* 114, 213–238 (1985). [PubMed: 3159830]
4. Bainbridge WA, Isola P & Oliva A The intrinsic memorability of face photographs. *J. Exp. Psychol. Gen* 142, 1323–1334 (2013). [PubMed: 24246059]
5. Bainbridge WA, Dilks DD & Oliva A Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *Neuroimage* 149, 141–152 (2017). [PubMed: 28132932]
6. Bylinskii Z, Isola P, Bainbridge C, Torralba A & Oliva A Intrinsic and extrinsic effects on image memorability. *Vision Res* 116, 165–178 (2015). [PubMed: 25796976]

7. Bainbridge WA Memorability: How what we see influences what we remember *Psychology of Learning and Motivation: Advances in Research and Theory* 70, (Elsevier Inc, 2019).
8. Bainbridge WA & Rissman J Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Sci. Rep* 8, 1–11 (2018). [PubMed: 29311619]
9. Mohsenzadeh Y, Mullin C, Oliva A & Pantazis D The perceptual neural trace of memorable unseen scenes. *Sci. Rep* 9, 6033 (2019). [PubMed: 30988333]
10. Smith SM Theoretical principles of context-dependent memory in *Theoretical Aspects of Memory* (eds. Gruneberg M & Morris PE) 167–194 (Routledge, 2006).
11. Xie W & Zhang W Mood-dependent retrieval in visual long-term memory: dissociable effects on retrieval probability and mnemonic precision. *Cogn. Emot* 32, (2018).
12. Xie W & Zhang W Negative emotion enhances mnemonic precision and subjective feelings of remembering in visual long-term memory. *Cognition* 166, 73–83 (2017). [PubMed: 28554087]
13. Xie W & Zhang W Negative emotion boosts quality of visual working memory representation. *Emotion* 16, 760–774 (2016). [PubMed: 27078744]
14. Bainbridge WA, Hall EH & Baker CI Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nat. Commun* 10, 5 (2019). [PubMed: 30602785]
15. Xie W & Zhang W Dissociations of the number and precision of visual short-term memory representations in change detection. *Mem. Cogn* 45, 1423–1437 (2017).
16. Raaijmakers JG & Shiffrin RM Search of associative memory. *Psychol. Rev* 88, 93–134 (1981).
17. Dehaene S et al. Imaging unconscious semantic priming. *Nature* 395, 597–600 (1998). [PubMed: 9783584]
18. Hills TT, Jones MN & Todd PM Optimal foraging in semantic memory. *Psychol. Rev* 119, 431–440 (2012). [PubMed: 22329683]
19. Clark SE & Gronlund SD Global matching models of recognition memory: How the models match the data. *Psychon. Bull. Rev* 3, 37–60 (1996). [PubMed: 24214802]
20. Anderson JR *Rules of the mind* (Lawrence Erlbaum Associates, Inc., 1993).
21. Griffiths TL, Steyvers M & Firl A Google and the mind: Predicting fluency with pagerank. *Psychol. Sci* 18, 1069–1076 (2007). [PubMed: 18031414]
22. Ralph MAL, Jefferies E, Patterson K & Rogers TT The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci* 18, 42–55 (2017). [PubMed: 27881854]
23. Burke JF et al. Synchronous and Asynchronous Theta and Gamma Activity during Episodic Memory Formation. *J. Neurosci* 33, 292–304 (2013). [PubMed: 23283342]
24. Long NM, Burke JF & Kahana MJ Subsequent memory effect in intracranial and scalp EEG. *Neuroimage* 84, 488–494 (2014). [PubMed: 24012858]
25. Spearman C Correlation Calculated From Faulty Data. *Br. J. Psychol* 1904–1920 3, 271–295 (1910).
26. Brown W Some Experimental Results in the Correlation of Mental Abilities. *Br. J. Psychol* 1904–1920 3, 296–322 (1910).
27. Kuder GF & Richardson MW The theory of the estimation of test reliability. *Psychometrika* 2, 151–160 (1937).
28. Lu H, Chen D & Holyoak KJ Bayesian analogy with relational transformations. *Psychol. Rev* 119, 617–648 (2012). [PubMed: 22775500]
29. Pennington J, Socher R & Manning CD GloVe: Global Vectors for Word Representation in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing 1532–1543 (Association for Computational Linguistics, 2014).
30. Brysbaert M, Warriner AB & Kuperman V Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res. Methods* 46, 904–911 (2014). [PubMed: 24142837]
31. Davies M & Gardner D *A frequency dictionary of contemporary American English: Word sketches, collocates and thematic lists* (Routledge, 2013).
32. Yaffe RB et al. Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proc. Natl. Acad. Sci. United States Am* 111, 18727–18732 (2014).

33. Jang AI, Wittig JH Jr, Inati SK & Zaghoul KA Human cortical neurons in the anterior temporal lobe reinstate spiking activity during verbal memory retrieval. *Curr. Biol* 27, 1700–1705.e5 (2017). [PubMed: 28552361]
34. Xie W & Zhang W Familiarity speeds up visual short-term memory consolidation: Electrophysiological evidence from contralateral delay activities. *J. Cogn. Neurosci* 30, 1–13 (2018). [PubMed: 28891784]
35. Kiesel A, Miller J, Jolicoeur P & Brisson B Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology* 45, 250–274 (2008). [PubMed: 17995913]
36. Rosenthal R, Rosnow RL & Rubin DB *Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach*. (Cambridge University Press, 2000).
37. Steiger JH Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychol. Methods* 9, 164–182 (2004). [PubMed: 15137887]
38. Bartlett FC *Remembering: A study in experimental and social psychology*. (Cambridge University Press, 1932).
39. Landrum RE & Gurung RAR The Memorability of Introductory Psychology Revisited. *Teach. Psychol* 40, 222–227 (2013).
40. Lowrey TM The relation between script complexity and commercial memorability. *J. Advert* 35, 7–15 (2006).
41. Bainbridge WA et al. Memorability of photographs in subjective cognitive decline and mild cognitive impairment: Implications for cognitive assessment. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit* 11, 610–618 (2019).
42. Rubin DC & Friendly M Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Mem. Cognit* 14, 79–94 (1986).
43. Brown J, Lewis VJ & Monk AF Memorability, word frequency and negative recognition. *Q. J. Exp. Psychol* 29, 461–473 (1977).
44. Sonkusare S, Breakspear M & Guo C Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends Cogn. Sci* 23, 699–714 (2019). [PubMed: 31257145]
45. Gillund G & Shiffrin RM A retrieval model for both recognition and recall. *Psychol. Rev* 91, 1–67 (1984). [PubMed: 6571421]
46. Hills TT, Todd PM & Jones MN Foraging in semantic fields: How we search through memory. *Top. Cogn. Sci* 7, 513–534 (2015). [PubMed: 26097107]
47. Stewart N, Chater N & Brown GDA Decision by sampling. *Cogn. Psychol* 53, 1–26 (2006). [PubMed: 16438947]
48. Sternberg S High-speed scanning in human memory. *Science* 153, 652–654 (1966). [PubMed: 5939936]
49. Yaffe RB, Shaikhouni A, Arai J, Inati SK & Zaghoul KA Cued memory retrieval exhibits reinstatement of high Gamma power on a faster timescale in the left temporal lobe and prefrontal cortex. *J. Neurosci* 37, 4472–4480 (2017). [PubMed: 28336569]
50. Machery E The amodal brain and the offloading hypothesis. *Psychon. Bull. Rev* 23, 1090–1095 (2016). [PubMed: 27294426]
51. Martin A The Representation of Object Concepts in the Brain. *Annu. Rev. Psychol* 58, 25–45 (2007). [PubMed: 16968210]
52. Martin A & Chao LL Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol* 11, 194–201 (2001). [PubMed: 11301239]
53. Rice GE, Caswell H, Moore P, Hoffman P & Lambon Ralph MA The roles of left versus right anterior temporal lobes in semantic memory: A neuropsychological comparison of postsurgical temporal lobe epilepsy patients. *Cereb. Cortex* 28, 1487–1501 (2018). [PubMed: 29351584]
54. Vaz AP, Inati SK, Brunel N & Zaghoul KA Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. *Science* 363, 975–978 (2019). [PubMed: 30819961]
55. Trotta MS et al. Surface based electrode localization and standardized regions of interest for intracranial EEG. *Hum. Brain Mapp* 39, 709–721 (2017). [PubMed: 29094783]

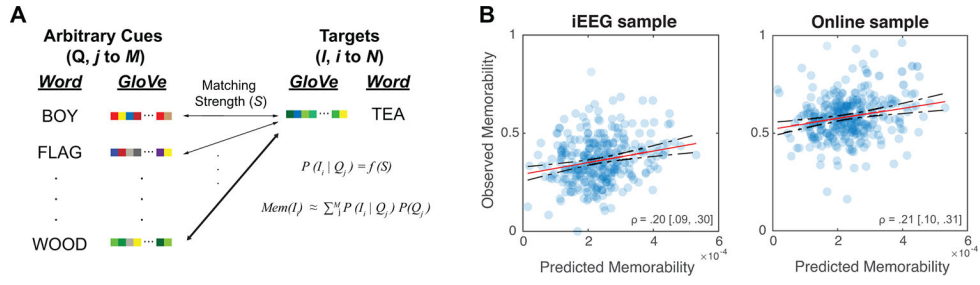
56. Buhrmester M, Kwang T & Gosling SD Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci* 6, 3–5 (2011). [PubMed: 26162106]
57. Wittig JH, Jang AI, Cocjin JB, Inati SK & Zaghoul KA Attention improves memory by suppressing spiking-neuron activity in the human anterior temporal lobe. *Nat. Neurosci* 21, 808–810 (2018). [PubMed: 29786083]
58. Chapeton JI, Haque R, Wittig JH, Inati SK & Zaghoul KA Large-Scale Communication in the Human Brain Is Rhythmically Modulated through Alpha Coherence. *Curr. Biol* 29, 2801–2811.e5 (2019). [PubMed: 31422882]
59. Bigdely-Shamlo N, Mullen T, Kothe C, Su KM & Robbins KA The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform* 9, 1–19 (2015). [PubMed: 25750622]
60. Joshua M, Elias S, Levine O & Bergman H Quantifying the isolation quality of extracellularly recorded action potentials. *J. Neurosci. Methods* 163, 267–282 (2007). [PubMed: 17477972]
61. Cohen MX *Analyzing Neural Time Series Data: Theory and Practice* (MIT Press, 2014).
62. Miller J, Ulrich R & Schwarz W Why jackknifing yields good latency estimates. *Psychophysiology* 46, 300–312 (2009). [PubMed: 19207197]
63. Rosenthal R & Rubin DB *r* equivalent: A simple effect size indicator. *Psychol. Methods* 8, 492–496 (2003). [PubMed: 14664684]
64. Xie W, Cappiello M, Meng M, Rosenthal R & Zhang W ADRA2B deletion variant and enhanced cognitive processing of emotional information: A meta-analytical review. *Neurosci. Biobehav. Rev* 92, 402–416 (2018). [PubMed: 29751052]



**Figure 1. Memorability of words is consistent across participants.**

(A) In the paired associates task, six pairs of words are sequentially presented on the screen. After a ~20s distraction period with simple addition math questions, one word from each pair is presented as a retrieval cue in a random order. Participants are instructed to vocalize (iEEG sample) or type (online sample) the associated word following the onset of the cue word. Each iEEG session consisted of up to 25 lists of this encoding-distractor-recall procedure, while the online experiment consisted of 3 lists. (B) Split-half analyses demonstrate that the memorability of each word is consistent in both the iEEG sample and online sample of participants. *top* The blue lines reflect word recall performance ranked for a random split half (Group 1), whereas the orange lines show recall performance of the remaining half (Group 2) ranked by data from Group 1 (both averaged across 5,000 iterations). The grey line shows an estimation of chance, by shuffling the rank orders of Group 2 (surrogate data). *bottom* The mean of the split-half correlation coefficients across 5000 iterations is significantly larger than the surrogate (null) distributions in both the iEEG sample (mean  $\rho_z = 0.19$ , bootstrapped 95% confidence interval, CI: [0.01, 0.40],  $p = 0.035$ , one-tailed) and in the online Sample (mean  $\rho_z = 0.23$ , bootstrapped 95% CI: [0.06, 0.44],  $p = 0.011$ , one-tailed). The red lines indicate the mean of the split-half correlation coefficients across 5000 iterations, whereas the black dashed lines indicate the mean of the surrogate data.





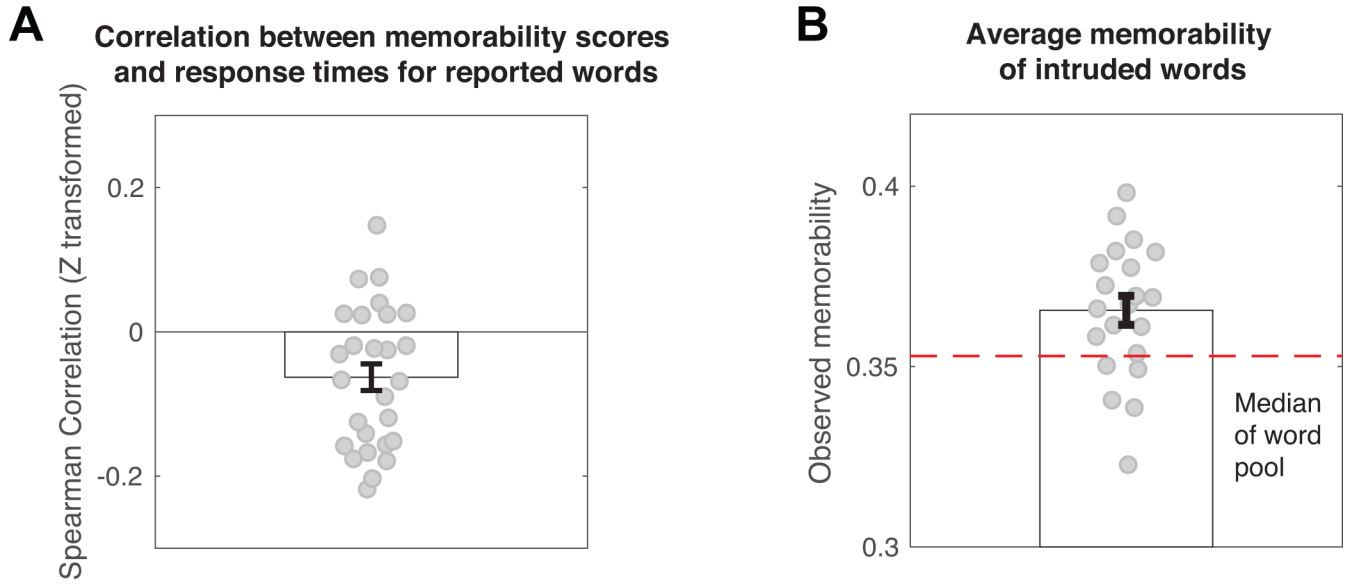
**Figure 2. Modelling memorability of target words in arbitrary verbal associations based on matching strength of words.**  
 (A) The likelihood that a cue word,  $Q_j$  ( $j = 1$  to  $M$ ), leads to the retrieval of a target word,  $I_i$  ( $I = 1$  to  $N$ ), is a function of the semantic similarity, or matching strength between  $Q_j$  and  $I_i$  (see Methods). We modelled the semantic similarity between words using the vectorized word features based on GloVe values<sup>29</sup>. Memorability of the target word  $I_i$ ,  $Mem(I_i)$ , can thus be approximated by the aggregated likelihood of successful retrieval cued by any arbitrary words in the available semantic space. (B) The predicted memorability estimates based on our computational model correlates with the observed memorability in both the iEEG sample (*left*;  $\rho = 0.20$ , 95% CI: [0.09, 0.30],  $p = 0.00057$ , two-tailed) and in the online sample (*right*;  $\rho = 0.21$ , 95% CI: [0.10, 0.31],  $p = 0.00029$ , two-tailed). Solid lines represent linear fits of the data, and the dashed lines represent 95% confidence interval of the linear fit.

Author Manuscript

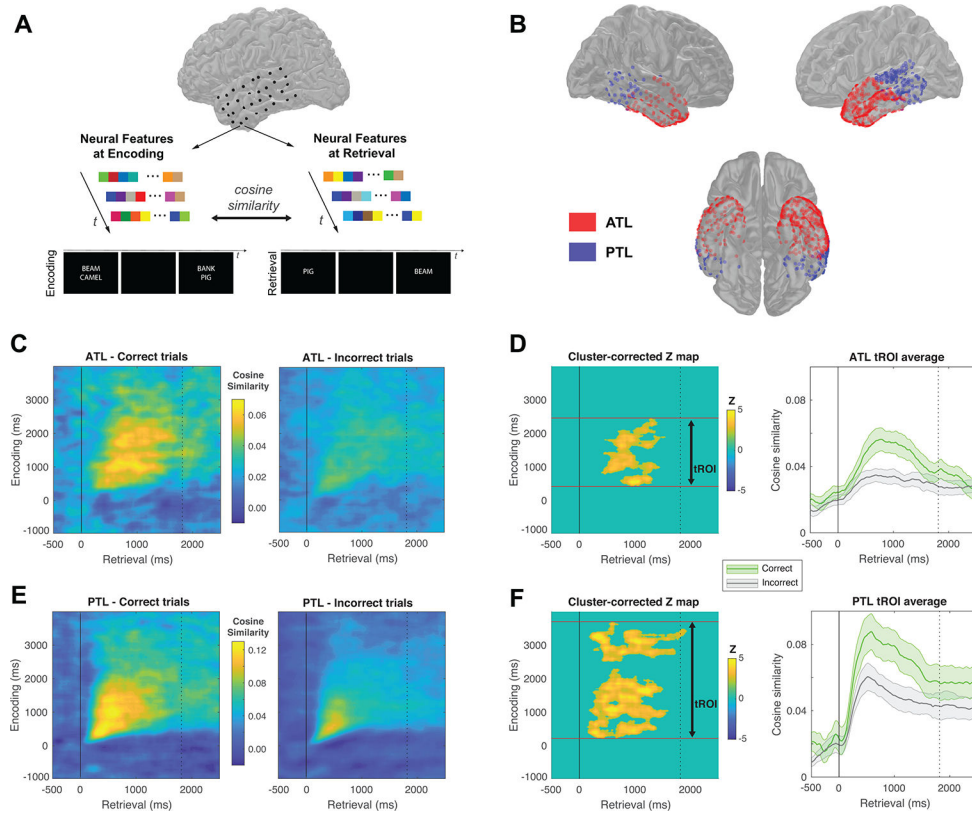
Author Manuscript

Author Manuscript

Author Manuscript

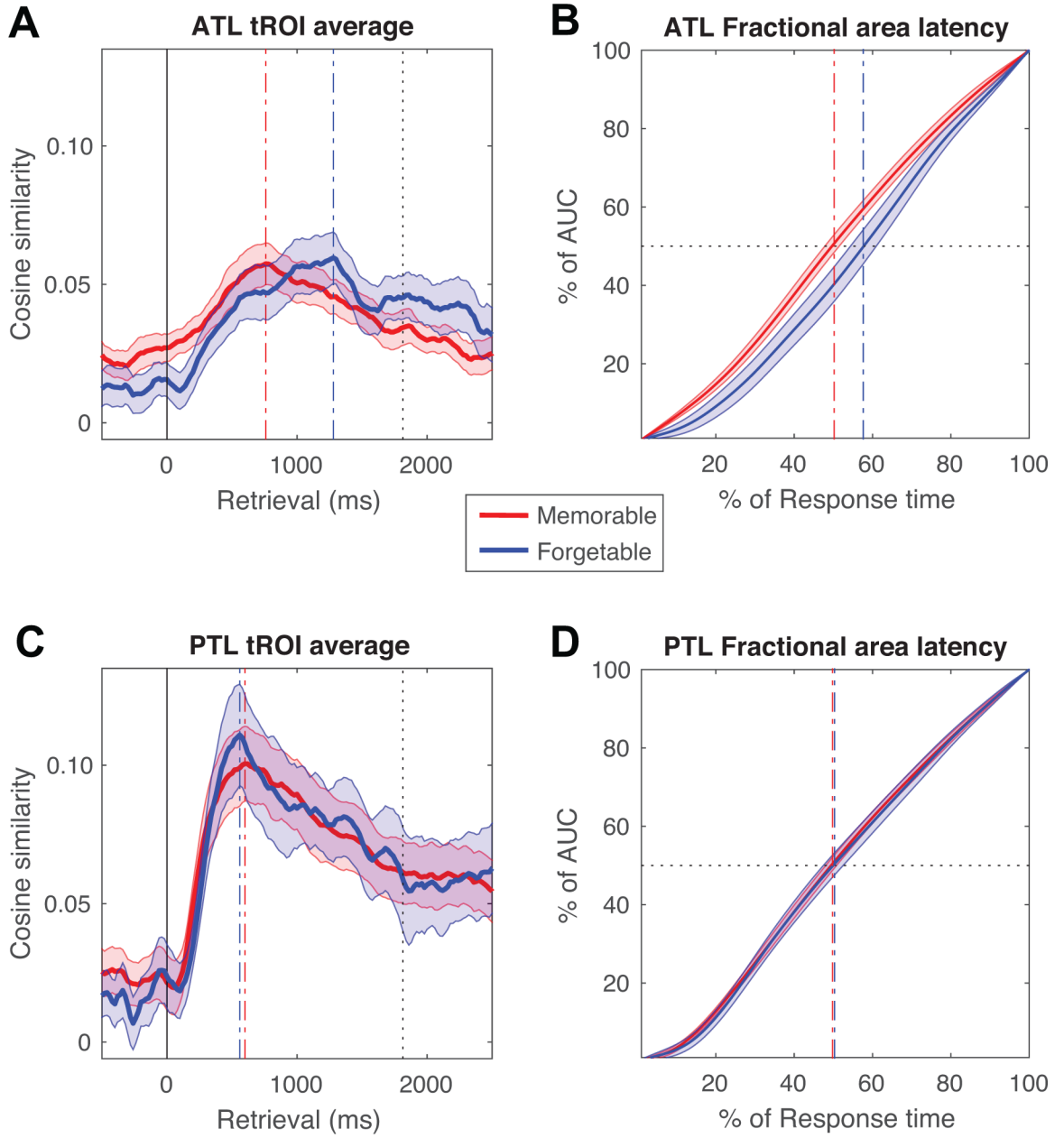


**Figure 3. Memorable words are retrieved more quickly but lead to more intrusion errors.** (A) Memorability of each responded word is significantly correlated with the response time required to vocalize the word across participants ( $\rho_z = -0.06$ , bootstrapped 95% CI:  $[-0.09, -0.03]$ ,  $p = 0.0004$ , two-tailed,  $n = 27$ ). Each point represents the Spearman correlation (Fisher's  $z$  transformed) observed in each iEEG participant. (B) The average memorability of intruded words across participants in the iEEG sample (0.366, bootstrapped 95% CI:  $[0.357, 0.373]$ ,  $n = 21$ ) is significantly larger than the median memorability of the entire word pool (0.353; bootstrapped  $p = 0.0016$ , two-tailed; complementary one-sample  $t$ -test:  $t(20) = 3.25$ ,  $p = 0.0040$ , two-tailed,  $r_{\text{equivalent}} = 0.59$   $[0.21, 0.81]$ ). Each dot indicates average memorability value from a single subject. In both panels, the bar and error bar represent bootstrapped mean and standard error across participants (random effect). Extended Data Fig. 1 summarizes the same data with within-participant variability shown in a forest plot.



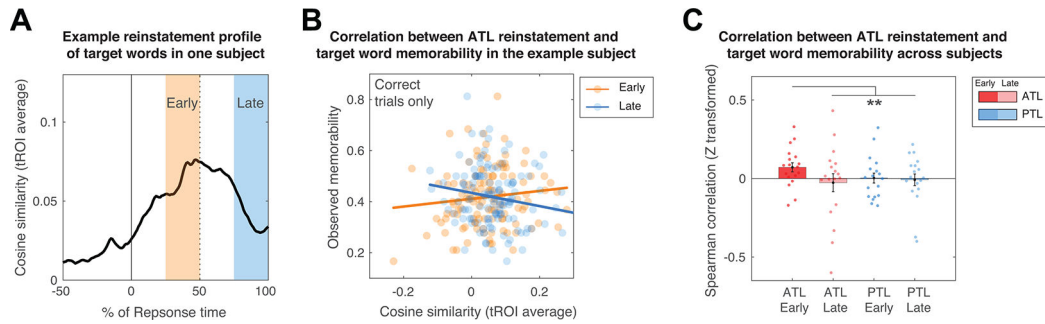
**Figure 4. Neural reinstatement during memory retrieval.**

(A) For every temporal window during encoding and retrieval, we construct a feature vector using the pattern of oscillatory power across electrodes. We calculate neural reinstatement as the cosine similarity between every brain state vector. (B) Electrode coverage in the ATL ( $n = 19$ ) and PTL ( $n = 18$ ) across participants who met inclusion criteria. (C) and (E) show reinstatement patterns for the ATL and PTL, respectively, for correct and incorrect recall trials time-locked to recall vocalization. Note the different colour scales used for plotting due to different reinstatement levels across brain regions. Significant differences in reinstatement between correct and incorrect trials in these regions were evaluated based on cluster-based permutation tests (see Methods). (D) and (F) show the average neural reinstatement during the temporal region of interest (tROI) in the encoding phase across different retrieval time points for the ATL and PTL, respectively. The response time averaged from participants' median response times across trials is plotted as a dashed line. Post-hoc comparisons confirmed that the neural reinstatement values averaged across encoding tROIs and the retrieval period from the probe onset to the median response time were significantly higher in the correct, relative to the incorrect, trials in both the ATL ( $t(18) = 3.39$ ,  $p = 0.0032$ , two-tailed,  $r_{\text{equivalent}} = 0.62$  [0.23, 0.84]) and the PTL ( $t(17) = 4.23$ ,  $p = 0.00056$ , two-tailed,  $r_{\text{equivalent}} = 0.72$  [0.38, 0.89]).



**Figure 5. Neural reinstatement is faster for memorable items in the anterior temporal lobe.** (A) Temporal profiles of ATL neural reinstatement during the temporal region of interest (tROI) in the encoding phase across different retrieval time points for correct memorable (red) and forgettable (blue) target words. The peak timepoints of the average reinstatement neural profiles are marked as a dashed vertical line in respective colours and the response time averaged from participants’ median response time across trials is plotted as a black dashed line. The solid lines at 0 indicate cue onset. (B) Fractional area measured as a function of response time (normalized by the average response time of participants) in the ATL with the 50% fractional area latency marked by the dashed horizontal grey line. (C) PTL neural reinstatement profiles for correct memorable (red) and forgettable (blue) target words, similar to (A). (D) Fractional area measured as a function of response time for neural

reinstatement profiles in the PTL, similar to (C). All error bars (areas) indicate bootstrapped standard errors across participants across 5000 iterations. Note, measures of fractional area latency and peak latency are complementary to each other. In general, memorable targets tend to be reinstated earlier during the retrieval time window in the ATL (bootstrapped means for 50% fractional area latency: 905 ms vs. 1050 ms,  $n = 19$ , bootstrapped 95% CI of latency difference:  $[-300 \text{ ms}, -20 \text{ ms}]$ ,  $p = 0.040$ , two-tailed; jack-knife peak latency<sup>35</sup>: 765 ms vs. 1264 ms,  $t(18) = 2.26$ ,  $p = 0.036$ , two-tailed,  $r_{\text{equivalent}} = 0.47 [0.02, 0.76]$ ).



**Figure 6. Target word memorability correlates with neural reinstatement in the anterior temporal lobe during early memory retrieval.**

(A) An example reinstatement temporal profile normalized by the trial-specific response time. Early and late time windows are respectively defined by the 25% to 50% and the 75% to 100% duration of the trial-specific response time. (B) Trial-level correlation between the neural reinstatement pattern and target word memorability from an example subject. Each dot represents trial level data, and the lines indicate linear fits of the data. (C) Across participants, the spatial and temporal profile of the correlation between the neural reinstatement pattern and target word memorability on average is captured by a contrast that predicts the strongest correlation in the ATL during the early time window as compared with other conditions (see Results and Methods for details). Each dot indicates the standardized correlation value from a single subject. Participants' data along with within-participant variability of the correlation measures can be seen in Extended Data Fig. 2. \*\*Bootstrapped  $p = 0.0016$ , one-tailed in a focal contrast analysis (complementary  $t$  contrast:  $t(17) = 2.88$ ,  $p = 0.0050$ , one-tailed,  $r_{\text{equivalent}} = 0.57$  [0.14, 0.82]) with a set of weights, +3, -1, -1, -1, for ATL-early, ATL-late, PTL-early, and PTL-late, respectively.

**Table 1.**

Multiple Regression Analyses on Observed Memorability Scores

	Observed memorability of 300 words			
	iEEG Sample (n = 30)		Online Sample (n = 2,623)	
	$\beta$ [95% CI]	<i>p</i> , two-tailed	$\beta$ [95% CI]	<i>p</i> , two-tailed
Predicted memorability	.19 [.06, .32]	.003	.18 [.05, .30]	.006
Word frequency <sup>a</sup>	-.09 [-.21, .03]	.16	-.03 [-.15, .10]	.68
Concreteness <sup>b</sup>	-.004 [-.12, .11]	.94	.09 [-.02, .21]	.11

Notes.

<sup>a</sup>Values ranked in the Corpus of Contemporary American English in ascending order. That is, a smaller value means higher word frequency<sup>31</sup>.

<sup>b</sup>Values from a large-scale word rating study, with a larger value indicate more concrete meaning<sup>30</sup>. A higher value reflects more concrete meaning. 95% CI = 95% confidence interval.