**Dissociating memory accessibility and precision in forgetting**

Sam C Berens1,\*, Blake A Richards 2,3,4, & Aidan J Horner1, 5,\*

1. Department of Psychology, University of York, UK.

2. Department of Biological Sciences, University of Toronto Scarborough, Canada.

3. Department of Cell and Systems Biology, University of Toronto, Canada.

4. Learning in Machines and Brains Program, Canadian Institute for Advanced Research, Canada.

5. York Biomedical Research Institute, University of York, UK.

\* Corresponding authors: sam.berens@york.ac.uk & aidan.horner@york.ac.uk

**Author ORCID identifies:**

Sam Berens: <https://orcid.org/0000-0001-8197-8745>

Blake Richards: https://orcid.org/0000-0001-9662-2151

Aidan Horner: <https://orcid.org/0000-0003-0882-9756>

Abstract

Forgetting involves the loss of information over time. While previous research has focussed on the rate of information loss, we know less about the form this loss takes. For example, are memory representations remembered with less precision, or do they instead become less accessible? Here we use ‘precision’ measures of memory, and probabilistic mixture models, to investigate the loss of both precision and accessibility of word-location associations over time. Importantly, we are able to directly compare these two measures by converting precision and accessibility into a common scale – information entropy. Using online testing, we will assess the extent to which forgetting is driven by a loss in either precision or accessibility. Further, we will assess how forgetting is modulated by shared content. Does learning multiple similar experiences decrease information loss for either precision or accessibility?

Forgetting is an inevitable consequence of remembering. We forget many of our everyday experiences over time, remembering only a small proportion of the large volume of information we process on a daily basis1. Psychologists have focussed on characterising the rate at which forgetting occurs – epitomised by Ebbinghaus’ forgetting curves2, and have asked why it occurs – for example via interference or decay3–6. This focus on both the when and why of forgetting has perhaps been at the expense of asking what is forgotten. When forgetting occurs, what type of information is lost? This question is critical given the proposal that forgetting is beneficial to decision-making processes7,8. If we are to understand how forgetting aids decision-making, we first need to reveal the form that such forgetting takes.

Here we outline two possible ways in which forgetting might occur – via decreases in memory accessibility or precision. Imagine being in a park and meeting a friend by a fountain in the north-east corner. Sometime in the future, you might want to remember the specific location where you met. A decrease in accessibility would decrease the probability of retrieving that specific memory, however if successfully retrieved you may remember the meeting location with the same accuracy as before. A decrease in precision would mean that the probability of successful retrieval does not decrease, but the spatial precision of retrieval does decrease. You might remember meeting your friend in the park, but not specifically by the fountain in the north-east corner. Both accessibility and precision can be defined as a loss of information, yet these two types of information loss should be behaviourally dissociable. Further, these potential two forms of forgetting might be underpinned by distinct mechanisms. For example, whereas accessibility might change as a function of the connection strength between a retrieval cue and its associated memory trace, precision might change as a function of noise in the underlying trace itself. Note, here we define ‘forgetting’ broadly in terms of a loss of information, as opposed to a more restrictive definition in relation to whether retrieval has been successful or not.

A number of theoretical accounts suggest that forgetting should involve different rates of decline for certain types of mnemonic information. In particular, Fuzzy-trace theory (FTT) posits that episodic memories are encoded by two independent traces that may be stored and retrieved in parallel9. One of these traces represents the fine-grained details of an event whereas the other encodes gist information in the form of semantic features. Relatedly, building on multiple trace theory10, the Trace Transformation Theory (TTT) proposes that the hippocampus supports the encoding and retention of episodic, context-rich, memories, while the neocortex transforms such representations into more semantic, gist-like, memories11,12. Empirical observations support these dissociations by showing that perceptual details may be lost faster than gist information13–16. However, this research focusses on loss of information for two distinct mnemonic representations, as opposed to losses in accessibility and precision for individual memory representations.

Recent research has shown that accessibility and precision are perhaps distinct components of an episodic representation. First, although accessibility and precision positively correlate across participants, they each have unique variance17. Participants can make accurate metacognitive judgements at retrieval related to this unique variance – they can subjectively report how accessible and precise memory retrieval is on a trial-by-trial basis18. Accessibility and precision have also been shown to be neurally dissociable. fMRI evidence has shown that trial-by-trial accessibility correlates with hippocampal activity, whereas trial-by-trial precision correlates with angular gyrus activity19, but also see20. Further, repetitive transcranial magnetic stimulation to the lateral parietal cortex produces improvements in precision, but not accessibility21. Thus, although there is evidence from working memory paradigms that accessibility and precision can be characterised using a single parameter model22, long-term memory studies have provided evidence that accessibility and precision are (at least partially) behaviourally and neurally dissociable.

One previous study has specifically focused on accessibility and precision in relation to forgetting in working memory23. Sun *et al* showed that encoding similar interfering material led to decreases in precision (referred to as ‘blurring’), whereas less dissimilar interfering material led to decreases in accessibility (referred to as ‘erasure’). In contrast to Sun *et al*, who focus on experimental interference in working memory, we focus on whether these plausibly distinct long-term memory processes can be dissociated via their forgetting rates over time. Assessing the temporal profile of forgetting is critical given that this reflects more naturalistic ‘everyday’ forgetting (i.e., participants are free to go about the daily lives in between encoding and retrieval). If forgetting does play a role in optimising decision-making processes, knowing what information is available to decision-making processes, and when it is available, is critical to the development of models of memory-guided decision-making. Additionally, understanding whether forgetting principally involves losses in precision or accessibility will inform theoretical accounts of long-term memory retention.

To date, research into forgetting has predominantly used binary measures of memory retrieval, where each retrieval trial can be classified as either correct or incorrect5. Forgetting under these experimental conditions is typically assessed by comparing accuracy (i.e., the proportion of correct responses) across experimental conditions. This general approach has been highly successful in delineating interference versus decay accounts of forgetting, and recently has shown that item-based familiarity is more susceptible to interference than decay, whereas recollection is more susceptible to decay than interference6. However, this experimental approach is not capable of dissociating between accessibility and precision. Note, there is no clear correspondence between familiarity vs recollection, and accessibility vs precision. Indeed, accessibility and precision may be independent components of recollection (dependent on the experimental task used). As such, we make no claims in relation to the debate surrounding possible dissociations between familiarity and recollection, instead focussing on potential dissociations between accessibility and precision.

As noted, ‘precision’ measures of memory have been used to study both working memory24,25 and long-term memory18–20. Here participants are required to remember a continuous perceptual detail of a stimulus, such as the colour of an object or its location on a circle. In the long-term memory literature, it is typical to pair a word with a location on a circle at encoding17,18,26. At retrieval, the word acts as the cue and participants have to move a cursor to the remembered location on the circle. Memory ‘precision’ is measured as the difference (angular error) between the correct and remembered location. Thus, memory performance is assessed with a continuous rather than binary measure.

Precision memory measures have also been combined with a statistical approach (mixture modelling) that allows for the characterisation of both memory accessibility and precision. Taking the angular error across all trials, mixture models allow one to fit a circular normal distribution (von Mises distribution) to the data. Once fit, the width of the von Mises distribution reflects the precision of memory retrieval. For example, if a participant is remembering circular locations very precisely, the distribution of angular errors will be narrow. Memory accessibility can also be estimated by considering the proportion of angular errors that were likely generated by the von Mises distribution, rather than being uniformly distributed around the circle (indicative of guessing). Importantly, these measures of accessibility and precision are independent of each other, such that if precision is high, accessibility can be either high or low (and vice versa). The combination of precision memory measures and mixture modelling therefore offers a unique opportunity to assess the extent to which forgetting decreases accessibility or precision.

Current measures of accessibility and precision are, to date, not directly comparable. Whereas the accessibility measure is related to ‘proportion correct’ in a more typical memory experiment, the precision measure relates to the width of the fitted von Mises distribution. To assess the extent to which forgetting is characterised by decreases in accessibility or precision, we need to develop a common metric. The concept of ‘information loss’ is related to entropy – which measures the lack of predictability in a given system27. As information is lost, the system behaves less predictably and so responses will become more variable. Here we use the entropy of behavioural responses28,29 to measure the amount of information loss across time. We introduce a common metric to directly compare information loss in terms of both accessibility and precision. Using this common metric, we will measure accessibility and precision across time, from immediate testing to 4 days after initial encoding. We will directly compare the pattern of decreases in accessibility and precision across time. Thus, we will be able to assess whether accessibility and precision decrease at differing rates.

Episodic memories are not encoded in isolation. We often experience events that are highly related, and can use that overlapping content to generalise across a set of events (referred to as schema30–32). Theories of consolidation, such as Standard Consolidation Theory33 (SCT) and TTT (introduced above11,12) propose that schematic representations, supported by the neocortex, are more stable and resilient to forgetting relative to more specific, hippocampal-based, episodic representations. Although existing schema can support the encoding of new item-based information34, the ability to generalise across related experiences might come at the expense of remembering individual events precisely35. Recent evidence suggests that participants use schema when making mnemonic decisions (which may be further modulated by systems consolidation36), and that this can result in systematic biases towards the ‘average’ representation across events when recalling individual events37. Thus, generalisation across a set of related experiences may result in a trade-off – decreasing total information loss over time at the expense of losing precise information related to specific events.

Here we ask whether similar events alter the rate of information loss for accessibility and precision over time. Word stimuli in the experiment will be grouped into two semantic categories, ‘manmade’ and ‘natural’. Participants will then associate these words with different locations around a circle. The circular locations for one group of words will be entirely random at encoding. However, locations for the other group of words will be spatially clustered (according to an underlying von Mises distribution with a fixed-width; conceptually similar to38). This clustering of locations for semantically similar words may allow participants to generalise across a set of related experiences (either at encoding or retrieval), potentially altering the rate of information loss for accessibility or precision (see hypotheses and supplementary information for pilot data). The present study will systematically characterise this potential differential loss of information, however future work will be needed to reveal whether this difference is driven by processes at encoding or retrieval, and what the nature of the underlying representations are in the clustered and non-clustered condition.

Using online testing, we will track rates of information loss in terms of accessibility and precision for word-location associations that are either randomly distributed around a circle (non-clustered) or spatially clustered. Our experimental approach focuses on memory for the word-location associations, rather than item memory for individual words (see planned exploratory analyses that differentiate item and associative memory).

Our preregistered analyses will assess five specific hypotheses (each has been assessed in our pilot data, providing evidence in favour of the alternative hypothesis; BFs>6; see Supplementary Figure 1 and Supplementary Table 1). Before decomposing into separate measures of accessibility and precision, we make two predictions in relation to the total amount of information (, see Methods). is a measure of the total amount of information in a given condition that takes into account both the level of accessibility and precision. First, we predict a decrease in total information across time, specifically for non-clustered words, consistent with the presence of forgetting (Hypothesis 1). Second, we predict that clustered words will confer an overall memory benefit relative to non-clustered words, consistent with a memory benefit when schema are formed (regardless of time; Hypothesis 2). These hypotheses act as positive controls, providing greater certainty for the validity of the more specific hypotheses below.

Of central theoretical interest is whether accessibility and precision differ in relation to forgetting, and how this further interacts with our manipulation of clustering. Here we decompose the measure of total information () into separate measures of accessibility () and precision () (the subscripts *p* and *k* refer to the corresponding parameters in the mixture model). First, we predict the temporal profile of forgetting, specifically for non-clustered words, will differ for accessibility and precision as these measures reflect different components of memory (Hypothesis 3). We remain agnostic as to whether this forgetting rate will be faster or slower for accessibility vs precision.

Our final two preregistered hypotheses relate to how clustering differentially affects accessibility and precision. As previously discussed, computational work has suggested a trade-off between generalisation and remembering individual events precisely35. Theories of consolidation also predict that gist-like, schematic, representations should be retained for longer periods of time, and that these representations might aid memory accessibility that the expense of precision11,12. We therefore predict that accessibility and precision will differ between the clustered and non-clustered condition (regardless of time; Hypothesis 4). In particular, this interaction is likely to present as increased accessibility, but decreased precision, in the clustered relative to non-clustered condition (see pilot data), however the statistical test is non-directional. Furthermore, this difference should be modulated by time, such that the rate of information loss for accessibility vs precision will differ dependent on whether words are clustered or non-clustered (Hypothesis 5). This three-way interaction is likely to present as a more rapid loss in accessibility in the non-clustered (relative to clustered) condition, in contrast to a more rapid loss in precision in the clustered (relative to non-clustered) condition (see pilot data), however the statistical test is non-directional.

As mentioned above, our principal hypotheses and preregistered analyses do not differentiate between failures to recognise individual cue words, and failures to recall specific locations when a cue word is remembered. However, potentially dissociating between these possibilities is also important. As such, at the end of each word-location retrieval trial, participants will be asked to provide subjective judgments regarding whether they remember both the cued word and its associated location (associative retrieval), the word alone (item recognition), or neither.

Planned exploratory analyses will then test for possible dissociations between item- vs associative-memory. These analyses may also shed light on potential differences between the clustered and non-clustered conditions. For instance, a performance advantage for clustered trials could result from either: (1) better memory for specific word-location associations within a spatial schema (enhanced retention), or (2) mnemonic generalisation involving the retrieval of representative locations when specific word-location associations have been forgotten (i.e., exemplar or prototype generalisation39). Higher proportions of associative retrieval judgments in the clustered condition would support an enhanced retention account whereas lower proportions would suggest the use of generalisation. Thus, our post-trial question will provide some measure of whether specific words or word-location associations are forgotten, depending on whether they are part of a semantic cluster.

To summarise, we will use online testing, precision memory measures, and mixture modelling to assess forgetting across time. Using a common metric (information), we will directly compare decreases in accessibility and precision over time, and how these decreases are further modulated by overlapping experience (i.e., clustered vs non-clustered words).

**Methods**

Participants

Participants (native English-speaking, aged between 18 and 35 years) will be recruited from Prolific (<https://prolific.ac/>). Prolific offers a web-based participant pool for behavioural scientists, manages participant payments, and ensures that individuals cannot participate in a given study more than once. All participants will have either normal or corrected-to-normal vision (by self-report) and will be compensated £7 for their time. The study was approved by a research ethics committee within the Department of Psychology at the University of York (ethical approval reference: 607).

Stimuli

A list of 200 common English nouns will be used as stimuli (<http://osf.io/8mzyc/>). These belong to one of 2 semantic categories; 100 manmade object nouns, and 100 natural object nouns. Words in each category were selected to be similar in length (mean difference: 0.020 characters; *d* = 0.011) and have a similar frequency in natural language (mean difference: 0.044; *d* = 0.063; as quantified by the Zipf scale in the Subtlex-UK database40). Additionally, we used a model of natural language word representations to ensure that the strength of semantic relationships between stimuli was similar in each category41. The word representations themselves were vectors in a 300-dimentional space and derived from a model that had been pre-trained on a set of web-based news articles containing approximately 100 billion words (see [https://code.google.com/archive/p/word2vec](https://code.google.com/archive/p/word2vec/)). We took the Euclidian distance between vectors as a measure of semantic relatedness. This showed that there was only a trivial difference between the manmade and natural categories in terms of the mean semantic similarity between words (*d* = 0.048). Nonetheless, a linear support vector machine was able to correctly classify 97% of the words as either manmade or natural using the vector representations alone. This suggests that the word categories were highly separable in semantic space. Finally, Kolmogorov–Smirnov tests showed that the distributions of word length, word frequency, and sematic relatedness did not substantially differ between the manmade and natural categories (each *D* ≤ 0.2).

Procedure

Participants recruited from Prolific will be directed to a secure website hosting the online experiment. An information sheet will be shown detailing what the study involves including a description of the data that will be collected and how it will be stored. At this time, participants will be randomly allocated to one of 7 conditions; an immediate retrieval condition, which directly follows an initial study phase, or a delayed retrieval condition (taking place either 3 hrs, 6 hrs, 12 hrs, 24 hrs, 48 hrs, or 96 hrs after the initial study phase). Before giving informed consent, participants will be made aware of which condition they have been allocated to. They will also be told to revisit the experiment website within ±1 hour of their scheduled retrieval session to complete the task and obtain a full payment. A unique participant identifier will then be provided by email which will be needed to start the retrieval session at the scheduled time. Participants will be prevented from running any phase of the experiment on mobile devices such as handheld smartphones or tablets. Additionally, the task prevents participants from using devises with a screen resolution less than 600 x 600 pixels.

Study phase

During the study phase, a circular dial will be visible in the centre of the screen. The task involves learning associations between different positions around this circle and specific words displayed on each trial (Figure 1A). All 200 word stimuli will be presented at least once during the study phase. Words belonging to either the manmade or natural sematic categories will be assigned to a ‘clustered’ condition. As such, they will be associated with similar locations around the circle - randomly sampled from a von Mises distributions with a fixed width (*k* = 2.0), and a fixed mean (randomly chosen for each participant). All words belonging to the other semantic category will be allocated to a ‘non-clustered’ condition. As such, they will be associated with circular locations that have no consistent mean angle (von Mises concentration parameter, *k* < 0.05). The assignment of manmade/natural words to the clustered/non-clustered conditions will be counterbalanced across participants.

Each study trial will start by indicating the circular position to be learned (location cue). A red cursor will be drawn at a particular location along the circles perimeter for 2 seconds (Figure 1A). Following this, the cursor will be removed, and a study word will be displayed onscreen for 4 seconds (word cue). Finally, with the word still visible, a red cursor will be redrawn at a random location. Using the mouse, participants will then verify that they have attended to the trial by repositioning the cursor at the cued location. This response window will last 6 seconds for each trial and will be followed by a 2-second inter-trail interval. If no response is made within the window, or if the response error is greater than 5°, the entire trial will be repeated. Pilot data indicated that participants rarely repeat a given encoding trial more than 5 times. Nonetheless, to limit trial-to-trial variability in the encoding procedure, word cues that are repeated more than 5 times will be excluded from the analyses. This study procedure is similar to that employed by previous investigations17,18. It is designed to ensure that participants attend to both the word and the location enabling an association to be learned between them.

Prior to starting the study phase, participants will watch a short video demonstrating how the session is to progress, including instructions on how to make responses (video transcript available at <http://osf.io/8mzyc/>). These instructions will emphasise that participants need to remember the word-location associations as they will be tested on them in the retrieval phase. As an aid to this, the video will ask participants to imagine an object that is related to the cue word appearing just beside the cued location before responding to each study trial. Following the study phase, participants in the immediate retrieval condition will complete the retrieval phase. Participants assigned to one of the delayed retrieval conditions will be reminded of when they need to revisit the experiment website.



**Figure 1. Schematic of the experimental procedure. A. Structure of a Study trial.** A location cursor is presented for 2 secs, followed by the word cue. The cursor then reappears in a randomly chosen location and the participant is required to move it back to the recently shown location (to within 5°). **B. Structure of a Test trial.** A location cursor is presented in a random location for 1 sec, followed by presentation of a word previously presented at study. The participant is required to move the cursor to the remembered location associated with the presented word (location judgement; 10 sec response window). Following this, participants are asked to indicate whether they remembered both the word and its associated location, the word alone, or neither of the two (subjective judgement; 5 sec response window).

Test phase

During test, participants will be tasked with recalling each of the 200 word-location associations. As with the study phase, a circular dial will be visible throughout. On each trial, a cue word will be presented onscreen and, following a 1 second delay, a red cursor will be drawn at a random location (Figure 1B). Participants will then move this cursor to the remembered location before making their response with a button press. Immediately after this, a prompt will be shown asking participants to indicate whether they: (1) remembered both the word and its associated location (‘Word + location’), (2) remembered the word but not its associated location (‘Word only’), or (3) had forgotten encountering the word (‘Neither’). Trails will be separated by a 2 seconds inter-trail interval and a response window will be imposed such that the next trail will begin automatically if both responses have not been entered within 15 seconds (10 sec response window for location judgement, 5 sec response window for subjective memory judgement). We will ask participants to be as accurate as possible, while ensuring a response is made on every trial. They will also be encouraged to make a best guess when entering a location response, even if they have no confidence in its accuracy.

As in the study phase, all participants will be shown a short video demonstrating how the retrieval session is to progress (video transcript available at <http://osf.io/8mzyc/>). After completing the retrieval phase, participants will be directed to a short questionnaire requesting a brief description of the strategy that they used when encoding and retrieving the word-locations associations. Participants will also be asked whether they had slept between the study and retrieval sessions and, if so, for how long. Following this, a debriefing sheet detailing the experimental hypotheses will be provided. If participants in one of the delayed retrieval conditions attempts to start the test session more than one hour before their scheduled time slot, they will be prevented from running the test and asked to return later. If participants miss their scheduled test session by over 1 hour, they will be directed to a dedicated debriefing sheet informing them that they are unable to participate further. This will further direct participants back to Prolific where they will be reimbursed for the time spent performing the study session (£3). Participants who return to the experiment website after completing the test session will be prevented from running the study and test phases a second time.

Recruitment protocol

An initial round of recruitment will run until we have collected 30 usable datasets per retention interval. At this point a statistical analysis of the data will be performed and recruitment will terminate if Bayes factors relating to each of our *a priori* hypotheses are either greater than 10 (strong evidence in favour of an effect) or less than 0.1 (strong evidence in favour of no effect). If the Bayes factors do not show this level of sensitivity, data collection will proceed in batches that add 10 usable datasets per retention interval. This will continue until all Bayes factors have met the sensitivity threshold up to a maximum of 60 datasets per retention interval (420 complete datasets in total; maximum number dictated by resource constraints). Simulations based on pilot data (see supplementary information) predict that all Bayes factors are likely to reach the sensitivity threshold at a sample size of ~26 participants per retention interval.

Data analysis

Mixture model estimation

We will simultaneously estimate retrieval probability (accessibility) and retrieval precision for individual participants using a probabilistic mixture model. First, we compute the replacement error of each response. This is given by the angular difference between a words target location at study, and the retrieved location at test (see *Eq. S1*). For the mixture model, angular errors are assumed to be drawn from one of two distributions; (1) a circular uniform representing random guesses, and (2) a von Mises distribution representing the precision of memory retrieval. Each of these distributions has an associated prior probability; a statistic reflecting the overall proportion of responses belonging to that distribution. The prior for the von Mises distribution (donated ) encodes the rate of memory retrieval (i.e., retrieval probability; ‘accessibility’). The von Mises distribution has two further parameters; a mean, and a dispersion statistic (known as the ‘concentration’). We fix the value of to remain at zero, assuming that the average angular error of retrieved responses is zero. The concentration parameter is analogous to the reciprocal of the variance; higher values of indicating a narrower distribution. As such, reflects the level of retrieval ‘precision’ and increases with better performance.

The parameters (retrieval probability, ‘accessibility’) and (memory precision) will be estimated for clustered and non-clustered trials (separately) using an expectation-maximization (EM) algorithm (detailed in the supplementary methods, *Eq. S2–S6*; MATLAB functions available at <http://osf.io/8mzyc/>). This attempts to identify values of and that maximise the likelihood of the observed data. The fit of the resulting mixture model is then compared to a reduced model that describes all angular errors with a single uniform distribution (i.e., no mnemonic components). This comparison is made by calculating the difference in Bayesian information criterion statistics between models (, see *Eq. S7*). If the mixture model fits the data substantially better than the reduced model ( < -10), the parameters returned by the EM algorithm are accepted.

When is > -10 (i.e., the mixture model provides a poor fit to the data relative to the reduced model) we will rely on an alternative fitting procedure (see supplementary methods for details). The EM algorithm often fails to achieve a good fit when accessibility is low ( ≲ 0.2; see supplementary methods). It is important to find a valid model fit to these datasets since merely excluding them will result in a survivorship bias - overestimating a population’s average performance because only the best performing individuals are included. Here, the parameter is systematically varied over a number of steps and is estimated from the corresponding proportion of responses with the smallest angular error. This procedure can identify valid model fits as local minimum values of the likelihood function that are missed by the EM algorithm. If this produces a fit that is substantially better than the reduced model (as above;  < -10), the parameters returned are accepted. However, if the alternative fitting procedure fails to return reliable estimates of both and for either the clustered or non-clustered condition, the participant’s entire data set will be excluded (exclusion criteria 6; see below).

Measures of memory-related information

While the model parameters and both reflect components of memory performance, these fundamentally different measures are not directly comparable. For instance, equivalent reductions in the values of and due to forgetting does not imply similar levels of forgetting in the form of accessibility and precision. We therefore use the differential entropy of angular errors to quantify the amount of mnemonic information that relates to each of these components. Entropy, denoted, describes the uncertainty associated with observing a set of responses (i.e., angular errors) from a given distribution. If responses are highly uncertain (i.e., angular errors are widely dispersed around zero), entropy will be high. This indicates that the distribution generating responses (i.e., the word-location memories) conveys little positional information. The entropy of a von Mises distribution reflecting recollected responses is defined as follows:

|  |  |  |
| --- | --- | --- |
|  |  | *Eq. 1* |

Which simplifies to:

|  |  |  |
| --- | --- | --- |
|  |  | *Eq. 2* |

The term, denotes the probability density function for a von Mises distribution at angle, with a mean of 0 and concentration of. The terms and refer to the modified Bessel function of the first kind with orders 0 and 1 (respectively), each evaluated at the point. When is zero, entropy is at a maxim () and corresponds to that of the circular uniform distribution. This would imply that memory provides no positional information at all. In contrast, when is large, (~17.5), entropy is near zero. This would suggest that responses are highly consistent with the learnt locations implying a large amount of mnemonic information. Given this, we subtract the entropy of recollected responses () from the maximum possible entropy () to produce a measure of mnemonic information, denoted:

|  |  |  |
| --- | --- | --- |
|  |  | *Eq. 3* |

This metric is 0 when precision is at a minimum, and increases monotonically with more precise memories. However, increasing values of to arbitrarily high levels will result in only marginal increases in . This reflects the fact that, beyond a certain point, increases in reflect only a small reduction in the angular span of the von Mises distribution.

Importantly, is unweighted by the retrieval probability () and so does not consider the proportion of word-location pairs that are recalled. We therefore define a similar measure of information related to retrieval probability, . As above, this is taken as the entropy (or uncertainty) associated with a given retrieval probability subtracted from. The act of retrieving a word-location association rules out random guesses (which are uniformly distributed). As such, the entropy associated with retrieval probability is taken as the uncertainty of random guessing () multiplied by the proportion of items that are not retrieved (). Subtracting this quantity from yields a measure of mnemonic information () that is 0 when retrieval probability is minimal, and increases linearly to a value of when retrieval probability is at a maximum:

|  |  |  |
| --- | --- | --- |
|  |  | *Eq. 4* |

As well as estimating the degree of mnemonic information associated with and separately, we will also use a combined measure of mnemonic information to assess overall memory performance. This measure, denoted, reflects the total amount of information retained in memory given how many word-location pairs are retrieved, and the precision of the retrieved responses. It is computed by taking a sum of the entropies associated with memory recall and random guessing, weighted by retrieval probability, and subtracting the result from:

|  |  |  |
| --- | --- | --- |
|  |  | *Eq. 5* |

This measure also relates to and in the following way:

|  |  |  |
| --- | --- | --- |
|  |  | *Eq. 6* |

Statistical modelling

Data from each participant will be included in the analyses provided five criteria are met: (1) the participant successfully completed both study and test phases, (2) less than 20% of study trials were repeated more than 5 times (due to missed responses or poor replacement accuracy), (3) the number of retrieval trials that timed out did not exceed 30 within each condition, (4) the strategy description provided by participants at the end of testing does not suggest cheating or a lack of understanding regarding the task, (5) the dataset is uncorrupted and free of technical errors, and (6) a mixture model can be satisfactorily fit to the participants data as discussed in the methods and supplementary information. With regards to criterion 4, three independent raters (lab-members, including the 1st and 3rd authors), blind to the experimental conditions, will review the strategy descriptions and determine whether each participant has followed the task instructions appropriately. Individual participants will be excluded if at least 2 of the 3 reviewers suspects cheating or a misunderstanding of the task.

Total information content of memory (It)

Hypotheses 1 and 2 concern the overall rate of forgetting (i.e., the loss mnemonic information measured by), and whether clustering of locations for semantically-related words improves overall memory performance (i.e., clustered vs non-clustered word-location pairs). To test these hypotheses, we will specify a generalised-linear mixed-effects regression model (GLMM) to predict within a 2x7 factorial structure (factor 1: clustering; factor 2: retention interval). Six binary coded predictors will model the effect of each delayed retention interval (3 hrs, 6 hrs, 12 hrs, 24 hrs, 48 hrs, or 96 hrs) by contrasting them to the intercept term (representing immediate retrieval). Another binary predictor will specify the effect of clustering by contrasting clustered vs non-clustered responses. Six further predictors will code the interaction between clustering and the delayed retention conditions.

In addition to the fixed effects predictors, a set of random effects parameters (2 per participant) will be included to allow the intercept and clustering terms to freely vary across participants. All elements of the associated random effects covariance matrix will be fully specified from the data. The model itself will use a log-link function and will be estimated via the maximum pseudolikelihood fitting method implemented in the MATLAB Statistics and Machine Learning Toolbox (MathWorks). Given that is bounded by zero, the dispersion of responses will be parametrised within the model using the gamma distribution. Pilot data (see supplementary information) revealed that this distribution provides a reasonable fit to the data and is better than all other commonly used distributions within the exponential family.

Table 1 lists each fixed-effects predictor and details the parameter contrast matrices that will be used to test hypotheses 1 and 2. Hypothesis 1 examines whether there is a monotonic change in the total information metric across the 7 non-clustered retention intervals. To implement this, we will run a linear contrast that compares estimates of across the intervals, weighted by the time difference between intervals. This requires a contrast vector that, when multiplied with the delayed retention parameters (D1-D6), yields an effect size representing linear changes in these estimates over time (as in table 1). Notably however, given that the GLMM uses a log link function, each parameter estimate reflects the log of . This means that the linear contrast actually tests for exponential changes in with respect to time. Exponential forgetting curves are known to provide a good fit to behaviour in both short-term and long-term memory experiments42, as well as our pilot data (discussed in the supplementary information). Hypothesis 2 tests the main effect of clustering; i.e., whether there are overall differences in the total information metric between the clustered and non-clustered conditions. As such, this involves specifying a contrast vector that takes a weighted average across the 7 clustering predictors. Further details of how these contrast vectors are computed and applied to test our hypotheses are outlined in the supplementary methods.

 **Table 1. Parameter contrast matrices for hypotheses 1 and 2 tested by the GLMM of total information ().**

|  |  |
| --- | --- |
|  | **Regressor name** |
|

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Clustered (C)** | **Delay 1 (D1)** | **Delay 2 (D2)** | **Delay 3 (D3)** | **Delay 4 (D4)** | **Delay 5 (D5)** | **Delay 6 (D6)** |  **C \* D1** |  **C \* D2** |  **C \* D3** |  **C \* D4** |  **C \* D5** |  **C \* D6** |

 |
|  |  |
| **Hypothesis 1**Change in total information across delay in the non-clustered condition |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **.299** | **.261** | **.187** | **.037** | **-.261** | **-.859** | **0** | **0** | **0** | **0** | **0** | **0** |

 |
| **Hypothesis 2**Difference in total information between clustered and non-clustered condition |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **.944** | **0** | **0** | **0** | **0** | **0** | **0** | **.135** | **.135** | **.135** | **.135** | **.135** | **.135** |

 |

Specific information content of memory (Ip and Ik)

Hypotheses 3, 4 and 5 concern differential rates of forgetting for clustered and non-clustered locations as measured by the two specific types of mnemonic information; (accessibility) and (precision). As above, we will test these hypotheses using a generalised-linear mixed-effects regression model (GLMM). The measures of mnemonic information, and, will serve as outcomes within this model, and the predictors will constitute a 2x2x7 factorial structure (factor 1: memory type; factor 2: clustering; factor 3: retention interval).

As before, one binary predictor will model the effect of clustering while a set of 6 dummy coded predictors will specify the effect of each delayed retention interval. An additional binary predictor will represent the difference between information types ( vs). Finally, a set of 19 predictors will model all interactions in the 3 factor structure. The model will also include a set of random effects predictors (3 per participant) enabling the intercept, information type and clustering terms to freely vary across participants. All elements of the associated random effects covariance matrix will be fully specified from the data. The model itself will use a log link function, a gamma distribution to parameterise dispersion, and will be estimated via the maximum pseudolikelihood fitting method.

Table 2 details how the fixed effect parameters of interest will be contrasted in order to test hypotheses 3, 4, and 5. As in the previous GLMM, two of these involve testing for log-linear differences over time. Specifically, hypothesis 3 examines whether there is a two-way interaction between delay and information type (specifically in the non-clustered condition), while hypothesis 5 tests for a three-way interaction between delay, information type and clustering. As above, the contrast vectors for these hypotheses are designed to compare all parameter estimates of interest with each other, weighted by the time difference between retention intervals. Hypothesis 4 tests for an interaction between clustering and information type and therefore constitutes a simple weighted average across the parameters coding for this effect. Further details of how these contrast vectors are computed and applied to test our hypotheses are outlined in the supplementary methods.

**Table 2. Parameter contrast matrices for hypotheses 3-5 for the GLMM for specific information content (, ).** Note, not all model parameters are listed; the model additionally includes parameters accounting for the non-interacting effects of information-type, clustering and delay. T = information type [ vs ]; C = clustering; D = delay condition.

|  |  |
| --- | --- |
|  | **Regressor name** |
|

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Info-type (T) \* C** | **T \* D1** | **T \* D2** | **T \* D3** | **T \* D4** | **T \* D5** | **T \* D6** | **C \* T \* D1** | **C \* T \* D2** | **C \* T \* D3** | **C \* T \* D4** | **C \* T \* D5** | **C \* T \* D6** |

 |
|  |  |
| **Hypothesis 3** The effect of delay will differ between accessibility and precision in the non-clustered condition |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **.299** | **.261** | **.187** | **.037** | **-.261** | **-.859** | **0** | **0** | **0** | **0** | **0** | **0** |

 |
| **Hypothesis 4**Clustering differentially effects accessibility vs precision. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **.944** | **0** | **0** | **0** | **0** | **0** | **0** | **.135** | **.135** | **.135** | **.135** | **.135** | **.135** |

 |
| **Hypothesis 5**Clustering changes the difference between accessibility and precision as a function of delay  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **0** | **0** | **0** | **0** | **0** | **0** | **.299** | **.261** | **.187** | **.037** | **-.261** | **-.859** |

 |

Bayesian inference

Each of our 5 *a priori* hypotheses will be tested by computing Bayes factors in favour of a meaningful effect (denoted *BF10*). Bayes factors greater than 10 indicate that, according to the data, there is at least 10 times more evidence in favour of the alternative hypothesis vs the null. Conversely, Bayes factors less than 0.1, indicate there is 10 times more evidence in favour of the null hypothesis over an alternative. When computing these statistics, we will use a Cauchy distribution with a scale parameter of 0.555 to represent our prior uncertainty of standardised effect sizes (see *Eq. S10*). This scale factor is approximately the median effect size observed in our pilot study (see supplementary information). It has been chosen such that the interval between the expected effect size and zero receives a similar prior weight to the interval between the expected effect size and infinity. Full details of how these Bayes factors are computed are provided in the Bayesian inference section in the supplementary information. To complement each Bayes factor, standardised effect sizes will also be reported. For completeness, we will also report frequentist inferential statistics, although these will not be used to make inferences.

Exploratory analyses

As well as testing our pre-registered hypotheses, we will run two additional planned exploratory analyses relating to the subjective memory judgements at the end of each retrieval trial. Given our lack of pilot data in relation to this aspect of the experiment, these are labelled as ‘exploratory’. Continued data collection will not depend on the Bayes factors from these analyses as we have no *a priori* way of estimating how many participants will be required to achieve sensitivity. Nonetheless, we will report all BFs, standardised effect sizes, and frequentist inferential statistics related to these exploratory analyses.

First, we will assess whether the subjective memory judgments provided at the end of each test trial suggest differential rates of forgetting for individual words (i.e., item memory) versus forgetting of work-location associations (i.e., associative memory). We will specify a cumulative link mixed-effects regression model using the ‘Ordinal’ package in the *R* programing language. This will account for relative changes in the proportion of test trials that are assigned either a ‘Word + location’, ‘Word only’, or ‘Neither’ response as a function of clustering and retention interval. The analysis will therefore involve 2x7 factorial structure with 3 responses categories. Random effects will be modelled in the same way as in the total information GLMM discussed previously. The model will use a logit link function, and will be estimated via the Gauss-Hermite fitting method. As this analysis involves subjective report data, it is not known *a priori* whether metacognitive response biases (e.g., a liberal tendency to respond ‘Word only’) will limit data quality and the conclusions that can be drawn. Nonetheless, the model may allow us to assess whether changes in accessibility seen in the pre-registered analyses were primarily driven by forgetting of the individual word (item memory) versus remembering the word but forgetting its associated location (associative memory).

Second, we will assess the relationship between word recognition as measured by subjective report, and mixture model estimates of accessibility and precision. Of specific interest is the extent to which losses in memory accessibility reflect either: (1) reduced accessibility for the cue word *per se*, or(2) failures to maintain the word-location association (in the presence of item memory for the word). To examine this, we will re-run the mixture models and GLMMs described above, but only including test trails where participants provided either a ‘Word + location’ or ‘Word only’ response. Excluding ‘Neither’ responses will result in measures of memory accessibility () that reflect participants ability to remember the word-location association when the word cue itself was subjectively recognised. However, as this analysis will be contingent on the proportion of words that receive either a ‘Word + location’ or ‘Word only’ response, it is again possible that metacognitive response biases will limit data quality. For instance, if ‘Word + location’ or ‘Word only’ responses are only made when recognition strength is very high, only highly memorable trails will be included in the mixture model thereby potentially biasing estimates of accessibility and precision. Additionally, limiting the number of trails in the analysis is likely to reduce the reliability of mixture model estimates in a way that does not uniformly affect each experimental condition.

Data availability

All anonymised behavioural data collected via the online task will be made freely available on the Open Science Framework (OSF) website (<http://osf.io/8mzyc/>).

Code availability

All HTML, PHP, and MATLAB scripts used to run the experimental task and analyse the data, will be made freely available on the OSF website (<http://osf.io/8mzyc/>).

Protocol registration

The Stage 1 protocol for this Registered Report was accepted in principle on **[TBC]**. The protocol, as accepted by the journal, can be found at **[TBC]**.

References

1. Wagenaar, W. A. My memory: A study of autobiographical memory over six years. *Cogn. Psychol.* **18**, 225–252 (1986).

2. Ebbinghaus, H. Memory: A Contribution to Experimental Psychology. *Ann. Neurosci.* (1913). doi:10.5214/ans.0972.7531.200408

3. McGeoch, J. A. Forgetting and the law of disuse. *Psychol. Rev.* (1932).

4. Postman, L. Transfer, interference and forgetting. *Exp. Psychol.* 1019–1132 (1971).

5. Wixted, J. T. The Psychology and Neuroscience of Forgetting. *Annu. Rev. Psychol.* **55**, 235–269 (2004).

6. Sadeh, T., Ozubko, J. D., Winocur, G. & Moscovitch, M. How we forget may depend on how we remember. *Trends Cogn. Sci.* **18**, 26–36 (2014).

7. Richards, B. A. & Frankland, P. W. The Persistence and Transience of Memory. *Neuron* **94**, 1071–1084 (2017).

8. Hardt, O., Nader, K. & Nadel, L. Decay happens: The role of active forgetting in memory. *Trends Cogn. Sci.* **17**, 111–120 (2013).

9. Reyna, V. F. & Brainerd, C. J. Fuzzy-Trace Theory: Interim Theory Synthesis. *Learn. Individ. Differ.* **7**, 1–75 (1995).

10. Nadel, L. & Moscovitch, M. Memory consolidation , retrograde amnesia and the hippocampal complex. *Curr. Opin. Neurobiol.* **7**, 217–227 (1997).

11. Winocur, G. & Moscovitch, M. Memory transformation and systems consolidation. *J. Int. Neuropsychol. Soc.* **17**, 766–780 (2011).

12. Sekeres, M. J., Winocur, G. & Moscovitch, M. The hippocampus and related neocortical structures in memory transformation. *Neurosci. Lett.* **680**, 39–53 (2018).

13. Murphy, G. L. & Shapiro, A. M. Forgetting of verbatim information in discourse. **22**, 84–94 (1994).

14. Kintsch, W., Welsch, D., Schmalhofer, F. & Zimny, S. Sentence Memory: A Theoretical. **159**, 133–159 (1990).

15. Sekeres, M. J. *et al.* Recovering and preventing loss of detailed memory: Differential rates of forgetting for detail types in episodic memory. *Learn. Mem.* **23**, 72–82 (2016).

16. Furman, O., Hasson, U., Davachi, L., Dorfman, N. & Dudai, Y. They saw a movie: Long-term memory for an extended audiovisual narrative. *Learn. Mem.* **14**, 457–467 (2007).

17. Harlow, I. M. & Donaldson, D. I. Source accuracy data reveal the thresholded nature of human episodic memory. *Psychon. Bull. Rev.* **20**, 318–325 (2013).

18. Harlow, I. M. & Yonelinas, A. P. Distinguishing between the success and precision of recollection. *Memory* (2016).

19. Richter, F. R., Cooper, R. A., Bays, P. M. & Simons, J. S. Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *Elife* **5**, 1–18 (2016).

20. Nilakantan, A. S., Bridge, D. J., VanHaerents, S. & Voss, J. L. Distinguishing the precision of spatial recollection from its success: Evidence from healthy aging and unilateral mesial temporal lobe resection. *Neuropsychologia* **119**, 101–106 (2018).

21. Nilakantan, A. S., Bridge, D. J., Gagnon, E. P., VanHaerents, S. A. & Voss, J. L. Stimulation of the Posterior Cortical-Hippocampal Network Enhances Precision of Memory Recollection. *Curr. Biol.* **27**, 465–470 (2017).

22. Schurgin, M. W., Wixted, J. T. & Brady, T. F. Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength. (2018).

23. Sun, S. Z. *et al.* Erasing and blurring memories: The differential impact of interference on separate aspects of forgetting. *J. Exp. Psychol. Gen.* **146**, 1606–1630 (2017).

24. Luck, S., Vogel, J. & Edward, K. The Capacity of Visual Working Memory for Features and Conjuctions. *Nature* **390**, 279–281 (1997).

25. Bays, P. M., Catalao, R. F. G. & Husain, M. The precision of visual working memory is set by allocation of a shared resource. *J. Vis.* **9**, 7–7 (2009).

26. Murray, J. G., Howie, C. A. & Donaldson, D. I. The neural mechanism underlying recollection is sensitive to the quality of episodic memory: Event related potentials reveal a some-or-none threshold. *Neuroimage* **120**, 298–308 (2015).

27. Cai, D., Kleeman, R. & Majda, A. A Mathematical Framework for Quantifying Predictability Through Relative Entropy. *Methods Appl. Anal.* **9**, 425–444 (2002).

28. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).

29. Verdugo Lazo, A. C. G. & Rathie, P. N. On the Entropy of Continuous Probability Distributions. *IEEE Trans. Inf. Theory* **24**, 120–122 (1978).

30. Bartlett, F. F. C. Remembering: An experimental and social study. *Cambridge: Cambridge University* (1932).

31. Ghosh, V. E. & Gilboa, A. What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia* **53**, 104–114 (2014).

32. Van Kesteren, M. T. R., Ruiter, D. J., Fernández, G. & Henson, R. N. How schema and novelty augment memory formation. *Trends Neurosci.* **35**, 211–219 (2012).

33. Mcclelland, J. L., Mcnaughton, B. L. & Reilly, R. C. O. McClelland PsycRev 1995 HPC-Neocortex.pdf. **102**, 419–457 (1995).

34. Kan, I. P., Alexander, M. P. & Verfaellie, M. learning in amnesia. **21**, 938–944 (2009).

35. Arpit, D. *et al.* A Closer Look at Memorization in Deep Networks. (2017).

36. Richter, F. R., Bays, P. M., Jeyarathnarajah, P. & Simons, J. S. Flexible updating of dynamic knowledge structures. *Sci. Rep.* **9**, 2272 (2019).

37. Brady, T. F., Schacter, D. L. & Alvarez, G. A. The Adaptive Nature of False Memories is Revealed by Gist-based Distortion of True Memories. *PsyArXiv* (2018).

38. Richards, B. A. *et al.* Patterns across multiple memories are identified over time. *Nat. Neurosci.* **17**, 981–986 (2014).

39. Mack, M. L., Preston, A. R. & Love, B. C. Decoding the brain’s algorithm for categorization from its neural implementation. *Curr. Biol.* **23**, 2023–2027 (2013).

40. van Heuven, W. J. B., Mandera, P., Keuleers, E. & Brysbaert, M. SUBTLEX-UK: A new and improved word frequency database for British English. *Q. J. Exp. Psychol.* **67**, 1176–1190 (2014).

41. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient Estimation of Word Representations in Vector Space. 1–12 (2013).

42. Rubin, D. C. & Wenzel, A. E. One hundred years of forgetting: A quantitative description of retention. *Psychol. Rev.* **103**, 734–760 (1996).

Acknowledgments

We thank Sara Mod, Lamia Begic, Tabitha Houldridge, and Thomas Maltby for help with collecting the lab-based pilot data. AJH is funded by the Wellcome Trust (204277/Z/16/Z) and ESRC (ES/R007454/1). BR is funded by a Learning in Machines and Brains Fellowship from the Canadian Institute for Advanced Research and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2014-04947). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Contributions

All authors contributed to research design and commentated on the manuscript. SCB and AJH wrote the manuscript and developed the analysis pipeline. SCB coded the experimental tasks, derived the experimental metrics, and implemented the statistical analyses.