Differential effects of aging and epilepsy in discriminating and reactivating memories

Xin You Tai[1,2], Arjune Sen[1,2,3], Masud Husain[1,2,4], Sanjay Manohar[1,2,4]

[1] Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford UK

[2] Division of Clinical Neurology, John Radcliffe Hospital, Oxford University Hospitals Trust, Oxford, UK

[3] Oxford Epilepsy Research Group, NIHR Biomedical Research Centre, Nuffield Department Clinical Neurosciences, John Radcliffe Hospital, Oxford, UK

[4] Department of Experimental Psychology, University of Oxford, Oxford, UK

Corresponding author:	Xin You Tai
Corresponding author's address:	Level 6, West Wing
	John Radcliffe Hospital
	University of Oxford
Corresponding author's e-mail:	xin.tai@ndcn.ox.ac.uk

Number of figure and tables: 5

Word count: 3055 (main text) and 156 (abstract)

Abstract

Pattern separation and pattern completion are distinct neurocognitive processes involved in encoding and retrieval of memories. However, there is currently no robust behavioural task in humans to measure both processes within the same paradigm. We describe the Memory Pinhole task, a novel paradigm which offers a distinct measure of each process, applied to healthy young, healthy older and people with epilepsy. Both pattern separation and pattern completion are observed in healthy younger individuals. A pattern completion deficit is seen in people with epilepsy, while the healthy older cohort show a deficit in pattern separation. To understand the neural mechanisms, we simulated human performance using an auto-associative neural network. Modelling indicated that disruption in different neuronal populations could explain the distinct memory profiles observed in ageing and epilepsy. These results demonstrate that pattern separation and pattern completion are distinct processes that can be measured from a single behavioural task and are differentially affected by ageing and epilepsy.

Introduction

Pattern separation and pattern completion are two specific and complementary neurocomputational operations that facilitate storage and retrieval of information in our memory[1]. Pattern separation is the process whereby similar stimuli are stored as separate neural representations. Pattern completion is the process of restoring entire and accurate neural representations of such stimuli when presented with a partial or degraded cue. Both processes are crucial for our day-to-day memory.

Theoretical and computational models describe distinct hippocampal circuitry implementing each of these operations, so that they can be performed independently[1–4], and experimental animal models have provided some corroborative evidence[5,6]. However, a fundamental question that remains is how these two theoretically distinct processes map onto human behaviour. Current human behavioural tasks tend to focus on one process or the other which makes it difficult to disambiguate between the two. Uncertainty partly arises because most paradigms have one behavioural outcome, wherein the participant is either correct or incorrect, but try to extract two distinct measures from this performance.

Pattern separation is commonly operationalized using memory tests where the test stimulus is similar -- but not identical -- to an originally remembered stimulus. A popular pattern separation paradigm, named the 'Mnemonic Similarity Task' (MST), is a change detection task whereby participants identify whether the probe is the same (*repeat*), similar (*lure)* or different (*foil/new*) to the encoding set[7]. There may be different levels of subjective similarity between original objects and their lures. A 'lure discrimination index' is calculated based on accuracy in correctly identifying *lures*[8] with most studies correcting for a bias in responding 'similar' by subtracting the rate that new objects are identified as 'similar'[9,10]. Other studies, however, calculate this pattern separation index from the *lure* accuracy in relation to false alarm rate (where similar items are reported as 'old')[11,12] and one experiment accounted for errors in both (responding 'old' or 'new' to similar lures)[13]. This inconsistency is further compounded as some authors consider the false alarm rate as a metric for erroneous pattern completion activity, or 'over-activity'[14–17]. However, there may be other reasons for this error including poor encoding of the original stimulus or forgetting due to intervening memory interference. Lure accuracy is directly related to false alarm rate and, therefore, this measure of pattern completion is likely to be an over-interpretation of the behavioural response.

Pattern completion, by contrast, has been measured using memory recall cued by a degraded or masked stimulus. For example, the 'Memory Image Completion' (MIC) task uses a visual delayed

recognition paradigm where participants learn a set of scenes and are later asked to identify the learned scenes, with new scenes mixed in, which are masked to different degrees[18]. Accuracy in identifying learned scenes decreases when more of the scenes were masked. A pattern completion metric termed as a 'bias towards completion' was derived by subtracting the accuracy of identifying a new scene (not present during the training phase) from the accuracy of identifying previously learned scenes. While this pattern completion measure could arise from incorrectly completing a previously stored scene, it may also be confounded by incorrectly discriminating a degraded new scene and thus reflect both processes of pattern separation and completion. Therefore, current tasks in the literature do not cleanly assess both cognitive processes.

Pattern separation and pattern completion offers an intriguing framework to understand encoding and retrieval of information in health as well as disease[19]. Studies in older individuals and people with Alzheimer's disease report reduced pattern separation which has been explained by 'over-pattern completion' [10,14,20]. As explained above, this conclusion may not be justified because in this task, the measure of pattern separation is directly related to pattern completion. A few behavioural studies in people with epilepsy show reduced pattern separation[21,22], while other studies examining individuals with hippocampal damage using two sequential tasks of pattern separation and pattern completion describe a deficit in both processes compared to controls[23]. Using sequential tasks does not overcome the issue of pattern completion or pattern separation being inadequately controlled for in each task. A final issue is that most pattern separation and completion metrics are dependent on overall task accuracy which may be affected by attention or executive function. These non-specific cognitive factors may be generally lower in patients with neurological conditions, such as epilepsy and dementia, and harder to interpret in the context of healthy controls.

A further attraction of pattern separation and completion is that they offer a computational framework that maps to neurobiology [3,24–26]. Computational models of memory provide controlled environments for studying the complex and interactive nature of information encoding and retrieval in the brain and allow testing of predictions in a parameter space beyond what is feasible or easily achieved with human or animal subjects. Importantly, computational models can move outside of physiology to simulate disease and "the patient" which is a considerably scarcer resource. Equipping recurrent artificial neural networks with associative plasticity gives a content-addressable memory system[27] that can emulate the functional properties of human memory [28]. This could be used to generate neural and behavioural hypotheses about memory and cognition.

Here, we describe the novel Pinhole Memory task which assesses both pattern separation and pattern completion within the same paradigm. Our metrics of each process correct for differences in

overall memory accuracy, and both are derived based on retrieving information about different dimensions of the same probe. The Pinhole Memory task was tested in a cohort of healthy young, healthy older and individuals with epilepsy to identify whether overall memory performance can be explained by specific deficits in these two cognitive processes. Further, we apply computational modelling of our behavioural results using a modified auto-associative Hopfield network to better understand potential underlying mechanisms of memory encoding and retrieval.

Results

Novel cognitive task with distinct measures of pattern separation and pattern completion

In the Pinhole Memory task, each working memory trial began with an encoding phase, where participants were presented with a shape to remember (**Figure 1A**). The shape had four arms which could be either long or short. After a delay, there was a retrieval phase, where they saw a probe which was either the same as the stimulus at encoding (50% of trials), similar with one arm different (more similar, 25%) or two arms different (less similar, 25%, **Figure 1B**). The probe was either more covered (two arms obscured) or less covered (one arm obscured). Participants were asked whether the probe was either the *same* or *different*. This design provided distinct assessment of pattern separation, based on two similarity levels, and pattern completion based on two levels of probe cover. Crucially, these two manipulations (similarity and covering) were performed on different dimensions of the same stimulus. In a second probe phase, participants were asked to indicate whether a covered arm was *long* or *short* as a measure of active pattern completion. There was a longer term memory probe every fourth trial to ascertain whether participants retained information over longer periods.

The Pinhole Memory task was first tested in a healthy young cohort (N=48) to establish baseline performance. Following signal detection theory, this control group demonstrated a high ability to discriminate between same vs different probes ($d' = 3.58$) as well as low bias ($c=0.05$, not significantly different to zero) across all trials. For non-match trials, as expected, accuracy for less similar probes was greater than more similar probes [$F(1, 94) = 11.30$, $p=0.001$] which measured pattern separation (**Figure 2a**). For match trials, accuracy was higher for probes which were less covered [$t=-3.15$, $p=0.028$], indexing pattern completion (**Figure 2a**). A small difference in accuracy between more similar and less similar conditions indicates better pattern separation. A small difference between more covered and less covered probes indicates better pattern completion, i.e. that covering an arm is less detrimental to performance.

There was a weak negative correlation between pattern separation and pattern completion indices (r^2= 0.08, p = 0.02), calculated by the difference in accuracy based on probe similarity and probe completeness respectively. This indicated that the pattern separation and pattern completion indices were two distinct but related metrics (**Figure 2b**). Active arm completion probe showed a higher accuracy with stimuli which were less covered (t = 2.17, p = 0.026), consistent with a pattern completion process. Overall accuracy for longer-term probes was 0.62 (SE = 0.05) with five individuals scoring below chance level (0.5) (Figure S1).

Dissociable effects of aging and epilepsy on pattern separation and pattern completion

The Pinhole Memory task was then tested in cohorts of healthy older individuals (N=28) and people with epilepsy (N=34). Across all working memory trials, both healthy younger and older individuals showed *d'* >4 and bias close to 0.05 (not significantly different from zero). People with epilepsy had a *d'*=2.44 and bias = -0.13 (significantly different from zero, t=-3.12, p = 0.028) indicating a bias towards reporting a stimulus was the same. Accordingly, the epilepsy cohort had a significantly greater mean accuracy on match trials compared with non-match trials (accuracy of 0.87 vs 0.80 respectively, t=2.87, p = 0.007 when testing the difference in means).

Two-factor ANOVA was performed for each cohort with main effects of similarity level and probe completeness. Performance on match trials (**Figure 2a**) was more accurate for less covered probes compared to more covered probes and was statistically significant in healthy young [F(1, 54) = 11.30, p=0.001] and healthy older individuals [F(1, 54) = 6.25, p=0.012] but not in people with epilepsy [F(1, 64) = 2.58, p=0.078]. For non-match trials (**Figure 2b**), accuracy was higher when probes were less similar to the original, compared to more similar, in healthy younger [F(1, 94) = 5.43, p=0.02] and people with epilepsy [F(1, 64) = 4.83, p=0.042] but not in healthy older individuals [F(1, 54) = 0.45, p=0.50].

Next, we asked whether separation and completion were independent. Pattern separation index was weakly correlated with pattern completion index for healthy young individuals but not for healthy older individuals or people with epilepsy (**Figure 2c**). The Bayes Factor suggested weak evidence towards the null hypothesis for each of the latter two cohorts (BF01 = 5.36 and BF01 = 4.87, respectively). To account for possible ceiling effects, an arcsine transform was applied to the data but this did not change the associations between pattern separation index and pattern completion index in any of the cohort groups.

To compare the groups directly, we used a trial-level generalised logistic mixed effect model, predicting accuracy from match (vs. non-match) trials, probe similarity (less vs. more), probe cover

(less vs. more), epilepsy patient group (vs. controls), age and sex along with interaction terms of patient group*match trials, patient group*probe cover and patient group*probe similarity (Figure S2). This model produced corroborative results with a significant interaction between match trials and being in the epilepsy group ($t= -6.83$, $p<0.001$), consistent with people with epilepsy having a bias to choose 'same' noted above. Ther was also a significant interaction between probes being less covered and being in a control group leading to higher accuracy ($t= 2.20$, $p=0.023$) which reflects that healthy individuals were better when the probe was less covered, but this was not seen in people with epilepsy. There was no significant interaction between probe similarity level and being a person with epilepsy ($t= -0.88$, $p=0.375$) reflecting reduced pattern separation in the epilepsy cohort. Main effects also demonstrated higher accuracy with match trials ($t=3.13$, $p=0.002$), probes being less covered ($t= -4.68$, $p<0.001$), probes being less similar ($t= 5.82$, $p<0.001$), being a healthy control ($t= -6.08$, $p<0.001$) and of older age ($t= 2.31$, $p=0.02$).

Active completion, longer term memory prove and reaction time

In the healthy older cohort, accuracy was higher in the active arm completion probe when stimuli were less covered ($t = 2.35$, $p = 0.026$), consistent with a pattern completion process. For the epilepsy cohort, there was no significant difference between active arm completion accuracy for more or less covered probes ($t = 1.33$, $p = 0.192$).

Performance was uniformly poor for longer term memory probes. The healthy older cohort had an overall accuracy of 0.61 (SE = 0.02) with one individual scoring below chance level (0.5) while the epilepsy cohort had an overall accuracy 0.54 (SE = 0.02) with six individual scoring below chance level (0.5).

Average time taken to give a response for match and non-match trials were similar within each cohort group (no significant difference) as younger individuals had the fastest average reaction time while people with epilepsy took the longest to respond (Figure S3). When plotting reaction time against accuracy at a trial level for each cohort, a speed-accuracy trade-off was observed with the healthy older cohort demonstrating a longer reaction time corresponding to higher accuracy (Figure S3). Importantly there were no RT differences between the level of similarity or between the level of completeness ($p>0.05$).

A modified Hopfield network accurately modelled the behavioural responses to pattern similarity and probe completeness

We used a Hopfield network to encode and retrieve the features present in our stimuli, encoded as a binary vector. Stimuli were encoded by activating the pattern of neurons corresponding to the features present in the visual stimulus, then updating the weights between the neurons according to a Hebbian rule to store the pattern auto-associatively. Visual features were either 'task' neurons for features to be reported, i.e. corresponding to arm length in the behavioural task, or 'context' neurons, corresponding to irrelevant features which differ for each new stimulus such as colour or central shape. The response of the network to a probe was calculated by activating the neurons corresponding to the features of the probe that were uncovered, and reading out the resulting activation pattern. We derived a *model certainty* measure indicating how strongly the network recognised the retrieval probe to be the same as a previous stimulus. This was based on the strength with which each pattern component was active or inactive (Figure S4).

Closely matched experimental parameters were used to model the Pinhole memory task. Specifically, 196 stimuli were encoded with corresponding match probes, more similar non-match probes (25% of the pattern flipped) and less similar non-match probes (50% flipped). All retrieval probes were either more covered (50%) or less covered (25%) and conditions were counterbalanced.

The key output was whether the model considered the probe as being 'same' or 'different' compared to encoding stimuli. The threshold for choosing "same" or "different" for a probe set as the median firing rate reflecting the assumption that participants would generally choose 50% same and 50% different in a standard behavioural task. As a baseline model architecture, we used N = 100 neurons with an equal number of context neurons. Remarkably the model has no other free parameters.

The model output showed worse performance in match trials for more covered probes and was therefore more 'certain' of a probe being a match when more information was present. This output matched the behavioural results of the Pinhole memory task and represented pattern completion (**Figure 3a**). For non-match trials, model performance was better for probes which were less similar to the encoding stimulus which also matched behavioural results and represented pattern separation (**Figure 3b**). Interestingly, the model predicted better performance at choosing "different" for non-match probes when they were more covered, which may reflect a lower likelihood of completing the pattern accurately when the probe is more obscured.

Fewer context neurons model behaviour from the healthy older group while fewer task neurons simulated performance from the epilepsy group

Reducing the number of context neurons was the model parameter manipulation that best modelled the behavioural results of the healthy older cohort in the Memory Pinhole task. With half the context neurons available, the model still showed better performance in match trials for less covered probes compared to more covered probes (pattern completion effect). This was observed in the behavioural results (**Figure 4a**). For non-match trials, however, there was no clear difference in performance for more or less similar probes suggesting no pattern separation effect. This also corresponded with the behavioural results (**Figure 4b**) suggesting different underlying biological computations with aging.

The behavioural results of people with epilepsy was simulated by lesioning the pool of task neurons. When considering a model with half the number of task neurons, the model maintained better performance for non-match trials with less similar probes representing a pattern separation effect that was observed in people with epilepsy (**Figure 4d**). For match trials, however, the effect of probe cover was smaller with reduced task neurons as the model was only slightly better at identifying less covered match probes (**Figure 4c**). Interestingly, people with epilepsy had a trend towards better performance with less covered match probes however this was not statistically significant. A larger sample size may be needed to detect this effect which was predicted by the model.

The simulations outlined above demonstrate that the auto-associative network can simulate the Memory Pinhole Task and that the dissociable results from healthy older people and people with epilepsy may be accounted for by lesioning different pools of neurons within the network. This may reflect underlying biological mechanisms in both cohorts

Discussion

The Pinhole Memory task provides a novel approach to obtaining distinct measures of pattern separation and pattern completion from a single memory paradigm by manipulating different dimensions of the same stimuli in a single task.

Healthy younger individuals demonstrated that probes with higher similarity and lower completeness were associated with worse performance, reflecting pattern separation and completion processes, respectively. These indices had a weak negative association which rebuts the idea that they lie on opposite ends of a unitary spectrum[14]. Comparing the performance with older

individuals and people with epilepsy provided insight on the effects of age and disease. While previous studies suggest that older individuals perform worse in change detection tasks due to 'under-separation' and 'over-completion' [14–17], results from the present experiment indicates a similar level of pattern completion to that seen in healthy younger individuals but less pattern separation occurred. The better average performance of the healthy older compared to younger cohort in our paradigm was likely due to a speed-accuracy trade-off.

People with epilepsy performed worse overall compared with healthy individuals but still showed better discrimination of probes which were less similar, reflecting a pattern separation process. By contrast, level of probe cover did not have a significant effect on performance which may reflect a deficit in pattern completion. This is the first report of a dissociable deficit between these two processes in older healthy individuals and those with epilepsy.

The Pinhole Memory task offers more specific and matched measures of pattern separation and pattern completion compared to previous tasks, and do not solely relate to overall memory accuracy. This is important as patients with neurological conditions may often perform worse than healthy controls which, in current tasks described in the literature, would usually be erroneously observed as deficits in pattern separation or pattern completion. Therefore, this improves upon several current tasks in the literature.

A modified Hopfield network was able to simulate results of the Pinhole Memory task, which characterises the separable computations underlying pattern separation and pattern completion. Compared to other autoassociative models of memory, we derived a same-different decision directly from the neurons in the network. Better performance identifying "same" probes was observed when match probes were less covered (pattern completion effect) while, in non-match trials, the model performed better in choosing "different" with less similar probes. Changing different parameters led to outputs that simulated behaviour from healthy older participants (decreasing context neurons) and people with epilepsy (decreasing task neurons). This corresponded with reduced pattern completion in healthy older participants and reduced pattern separation effect in people with epilepsy. These findings suggest that pattern separation and completion are distinct processes in which different neuronal populations are involved. Remarkably, this model performed these two separate processes using a single neuronal population. This may be surprising when considering that, neuroanatomically, the dentage gyrus is thought to perform pattern separation whilst CA3 is considered the site of pattern completion. However, this is consistent with the hippocampus performing multiple independent processes as a single unit and helps resolve the uncertainty of how neuronal signals move neuroanatomically from entorhinal cortex to dentate gyrus to CA3 for pattern

completion to occur without pattern separation or if there must be an additional perforant pathway between entorhinal cortex directly to CA3 without engaging dentate gyrus neurons.[29]

Interpreting the role or biological relevance of each neuronal pool is not straightforward. Task neurons are directly involved with retrieving stored representations within the model. Reducing or lesioning task neurons may be analogous to disrupting the memory retrieval process which may occur in the CA3 region of the hippocampus. CA3 has numerous recurrent synaptic collaterals [30] and is vulnerable to epileptic aberrant activity. However, task neurons are also required to encode stimuli in the first instance. Therefore, both encoding and retrieval are likely affected in people with epilepsy however a reduced pattern completion effect was more clearly observed in both behavioural results and predicted by the model output.

Changing context neurons modelled the behaviour of older participants. While not directly involved in the relevant task pattern, context neurons allow better encoding of each pattern by increasing the difference between other patterns in the model. This, in effect, describes a pattern separation effect which was reduced in the behavioural results of older participants. From a computational perspective, the context neurons allow patterns to be stored separately within more distinct basins of attraction within the network[6,31]. This reduces the chance that one pattern will be confused by another pattern at recall. This change in context neurons may reflect generalised atrophy observed in the ageing brain[32].

It is important to consider limitations to the current behavioural paradigm. While the stimuli were grounded in a mathematical framework that was designed to precisely manipulate similarity and probe completeness, some may argue that their abstract nature will make it difficult to employ in a more episodic paradigm with longer delay and more interference between encoding and retrieval. Adding more context and complexity to the probes will help with this. Furthermore, collecting baseline cognitive measures in healthy participants would have been preferred but this was not possible due to disruption to in-person testing during the study.

In conclusion, we demonstrated a novel paradigm to assess both pattern separation and pattern completion within a single task. Results suggested that these two cognitive processes do not lie on a unitary spectrum but are two related but distinct operations. This is consistent with theoretical models and helps to resolve an existing debate in the literature. Pattern separation and pattern completion may be a useful framework of memory encoding and retrieval to understand changes in health and disease with dissociable effects of age and epilepsy demonstrated here.

Methods

Ethics

Ethical approval was granted by the University of Oxford ethics committee (IRAS ID: 248379, Ethics Approval Reference: 18/SC/0448). All participants gave written informed consent prior to the start of the study.

Participants

There were three participant cohorts tested in this experiment. A group of healthy young individuals with forty-eight participants (34 female, mean age 27 years, range 19-34) were recruited via a local database or the online platform Prolific. Participants were paid £10-15 for the task which was performed on a desktop computer under supervision or via the online platform Pavlovia. A healthy older group of 28 volunteers (18 female, mean age 65 years, range 45-77) were recruited via a local database and the task was done at home via the online platform Pavlovia. Three older volunteers did not understand the task properly and were excluded from the study.

A group of 38 people with epilepsy (20 female, mean age 34 years, range 19-59) were recruited from the epilepsy video telemetry unit, John Radcliffe hospital, Oxford, United Kingdom. These participants were admitted to hospital for at least one week to have continuous monitoring of seizure activity. Most of this group had their anti-epileptic medications reduced during their hospital stay with the majority being tested off medication. Further demographics can be found in Table S1. Four people with epilepsy were excluded as they did not complete the task due to fatigue or a seizure occurring during the task. Addenbrooke Cognitive Examination (ACE-III) was performed in the epilepsy group but not the other cohorts due partly to logistical reasons as face-to-face testing was not carried out for a substantial proportion of the study due to Covid-19 lockdown protocol.

Pinhole Memory task

The 'Pinhole Memory task' was a novel pattern separation and pattern completion working memory paradigm which manipulated similarity between encoding stimuli and retrieval probe in addition to covering one or two dimensions of the probe. Memory stimuli followed a common base shape with four dimensions/ 'arms', either long or short (fixed lengths of 140 or 60 pixels), with a random angle of 15 degrees to either side of the base arm (**Figure 1a**). The crucial manipulation of similarity was achieved by altering arm lengths between encoding stimuli and retrieval probe on each trial. To create complex and visually rich stimuli, additional features were added including a coloured

background shape as well as end-arm, central and middle shapes which were held constant throughout the working memory trial of the task.

The Pinhole memory task had three different probe stages:

1) Working memory change detection (main probe)

During each trial, participants would see one stimulus at encoding and, after a delay, would be presented with a single probe which was either the same as the stimulus at encoding, similar with one arm different (more similar) or two arms different (less similar). The probe was either more covered (two arms obscured) or less covered (one arm obscured). Participants were asked whether the probe was either the same or different (**Figure 1b**). This probe stage was designed to assess pattern separation (two different similarity levels) and passive pattern completion (two different levels of probe cover). This working memory probe is the key condition described in this report. Each participant performed 196 trials of this main working memory change detection probe with counterbalanced conditions of more or less similar, and more or less covered.

2) Working memory arm completion

Following their first response, participants were asked to indicate whether one of the covered-up arms was either long or short as a test of active completion.

3) Longer-term memory

Finally, every four trials, a longer-term memory probe was shown and participants were asked if this stimulus was the same or different to a few trials ago. The full probe was shown (no arms were covered) and the arm lengths were either the all the same or all different. This was the largest contrast possible between stimuli and long term memory probe due to the difficult nature of this probe. An important reason to include these trials was to encourage participants to retain previous information in long term memory.

Analysis

Group accuracy was the dependent variable for rm-ANOVA or with a Student's T-test to compare means between two groups. Signal detection theory was applied to calculate discriminability and bias. When appropriate, Bayes factor was calculated to examine for evidence towards the null hypothesis using an unbiased prior distribution, such as theta = beta(1,1) or a stretched version, using JASP (Version 0.16.3).

A '*pattern separation index*' (PS index) was derived by taking the difference between mean accuracy of trials with less similar probes minus trials with more similar probes:

$$\boldsymbol{PS\ index = Accuracy\ of\ less\ similar\ probes - accuracy\ of\ more\ similar\ probes}$$

A '*pattern completion index*' (PC index) was derived by taking the difference in mean accuracy between trials with less covered probes minus trials with more covered probes on trials where the probe was the same as the encoding stimuli:

$$\boldsymbol{PC\ index = Accuracy\ of\ less\ covered\ probes - accuracy\ of\ more\ covered\ probes}$$

The advantage of using performance differences, based on similarity and probe cover, was to a measure of pattern separation and completion regardless of overall memory performance.

Additionally, a generalised logistic mixed-effect model analysed trial—level performance to specifically investigate higher-level interactions within the experiment. The model was specified based on prior hypothesis using the fitglme function within MATLAB, Mathworks. Match (vs non-match trials), similarity level, cover level, patient group (epilepsy vs. healthy), age and sex were fixed effects while a subject ID was entered as a random effect. An interaction term was included between patient variable and match, similarity and probe cover level. Reaction time was investigated separately as a factor that may explain overall performance and also whether non-match trials were slower.

Hopfield network model for memory encoding and retrieval

A Hopfield network of information encoding and retrieval has been previously used to model pattern completion that may occur in the hippocampus[33]. Memory representations are stored as orthogonal, or largely uncorrelated, patterns to increase network capacity[34]. For this study, we build upon previously described network architecture and propose modification to incorporate task and context

neurons, interpretation of non-binary neuron activity and a novel, unbiased model decision on "same" and "different" stimuli.

Mathematically, an external input e_i is applied to each neuron i of the network through unmodifiable synapses to produce output firing y_i. The output from each neuron is connected via recurrent collateral connections to the other neurons in the network and stored in a modifiable weights matrix w_{ij}. This organization allows the initial activity of these neurons to be associated and 'learned' to model information encoding.

A modified Hebbian local learning rule is applied to the recurrent synapses in the network:

$$\delta\, w_{ij} = \alpha(y_i - y)(y_j - y) \qquad \text{eq (1)}$$

Where α is a constant and y approximates the average value of y_i and y_j.

Subtracting the value y will prevent increasingly positive or negative, i.e. exploding, weight values. Once this stimuli information is stored, a subsequent recall probe can be presented to the network as a full or partial cue as an external input to produce neural output firing in the pattern of the original 'learned' input.

During recall, an external input e_i activates neurons, and this activity propagates through the recurrent collateral system to produce an activation output of each neuron. The activation vector h_i is the sum of activations produced in proportion to the firing rate of each axon y_i operating through each modified synapse w_{ij}.

$$h_i = \sum_j y_i w_{ij} \qquad \text{eq (2)}$$

Where the sum is over the C input axons to each neuron, indexed by j. The output firing y_i is a function of the activation produced by the recurrent collateral activity (internal recall) and the external input (e_i), where f is a non-linear (e.g. arc tan) activation function.

$$y_i = f(h_i + e_i) \qquad \text{eq (3)}$$

Model characteristics:

External inputs to the model

Following the classic Hopfield network literature, external input values fed into the network were generally constrained to either '1' or '0'. This often depicts an 'active' or 'inactive' neuron and could also represent 'long' or 'short' arms of our visual stimuli to be stored into memory. We extend this to non-binary values to reflect a specific experimental parameter. For example, an external input of '0.5' could represent a covered or uncertain arm length in the context of other inputs being either '1' or '0'.

Model sparseness

Sparseness level of model inputs is an important characteristic of an autoassociative network and has implications on the network storage capacity[35]. At a population level, sparseness S refers to the number of neurons which are firing at a given time. Therefore, S will range from 1/N to the S = 1 when all the neurons are firing, or 'active', and S = 0.5 if half the neurons are 'active' for a given neuronal pattern. To model the behavioural task, sparseness was fixed at S = 0.5 for this study.

Task and context neurons

This model had two main neuronal populations. 'Task' neurons which were directly involved with encoding a pattern into the network and retrieving the stored representation. For example, the activity of one neuron may correspond to one component or feature dimension of the stimulus. By contrast, 'context' neurons were a second neuron population with random activations for each stimulus, presented together with the pattern to be remembered (with sparse level S=0.5). Despite not being directly involved in encoding or retrieving a specific pattern, context neurons helped to provide a stable representation of each pattern within the network. Conceptually, they can be considered as providing some related information such as an environmental association or potential spatial or temporal context specific to that pattern. Binding of remembered information into a unique environmental context, such as time, place and person, are often considered an essential attribute of hippocampal memories[36]. For this study, a baseline model architecture comprised of 100 task neurons (N=100) with an equal number of context neurons unless otherwise specified, for example, when the model was specifically 'lesioned' to simulate pathology.

Modelling similarity

Pattern similarity can be manipulated in different ways. Following Euclidean geometry, the magnitude of the dot product between two vectors is a measurement of similarity whereby a higher value indicates greater similarity (this can also be represented by the cosine similarity) [37]. To manipulate similarity between two patterns, we start with one pattern and took the inverse state of one neuron component (dimension) of the pattern, which we refer to as 'flipping' the component. For example, two similar patterns can be created by taking pattern A and flipping one component neuron state from 'active' to 'inactive' to create pattern B and leaving all other components the same. Figure S5a shows how pattern A is changed to pattern B (more similar) by flipping one active and one inactive neuron or to pattern B' (less similar) by flipping all neuron states. As more components are flipped to create a new pattern B, the dot product between the patterns A and B decreases as they become less similar to each other. Generally, for the experiments outlined in this paper, a pair of inactive and active neurons were flipped at the same time to maintain the same model sparseness (S=0.5).

Similarity within a pattern can manipulated by repeating a series of component neurons within the pattern. For example, in Figure S5b, pattern A is comprised of 12 components with the first six components being the same as the second six components reflecting a high within-pattern similarity. Pattern A' has a lower within-pattern similarity as the second six components are randomly different from the first six components.

Model output: accuracy

The model output to a recall probe is a set of neural patterns that most closely matches the original input, based on stored model weights. These output patterns are generated from the firing rate of each neuron once the external input probe has passed through the recurrent collateral system. The representational space for the neuronal firing rate is continuous and will generally follow a gaussian distribution.

To convert the firing rates into binary patterns, the neuronal activity can be interpreted as an 'active' or 'inactive' neuronal state by applying a threshold activation value. Previous models will often use a threshold based on the sparseness value of the original stimuli[1]. For example, if original patterns had sparseness S = 0.5, then the threshold for being considered active was the median activation value of all neurons. Once each output neuron activity was interpreted as a 0 or 1, then this output pattern can be directly compared with the original learned patterns to calculate a ***model accuracy*** for a given set of probes i.e. how closely does the model output neural patterns match the learned input patterns for a given set of probes.

Our modified Hopfield network demonstrated several fundamental principles of memory encoding and retrieval including set size effect and retrieval probe completeness affecting model accuracy or other performance metrics (**Box 1**). Furthermore, it can also simulate results from popular pattern separation and pattern completion behavioural tasks in the literature (Box S1).

Model output: "Same" or "different" decision

To model a delayed recognition task, such as the Memory Pinhole task, a decision on whether the recalled memory or model output pattern is the "same" or "different" to the original learned stimuli is required. One straightforward way to do this was to set a minimum accuracy level for the model to consider the pattern as being "same". If the model output does not resemble the learned pattern, then the accuracy for that pattern would be low and would be considered as "different". This, however, is not biological plausible as it requires the model to know the ground truth, i.e. correct answer, to compare with the model output pattern before making a decision.

As an alternative, we used the raw firing output of each neuron as a *certainty* measurement for whether that component of the pattern was 'active' or 'inactive'. Based on the Hebbian learning principle, the stored synaptic values for active pattern components will be positive while the firing rate values for stored inactive pattern components will be negative. Therefore, the model will be more likely to recall a more positive firing rate neuron as 'active' and a more negative firing rate as 'inactive' whereas values in the middle may be considered either active or inactive by the model. This likelihood represents a *certainty* measure for the model. Since more positive and more negative values reflects greater certainty, we used the absolute value of each neuronal firing rate and applied a threshold to indicate whether the model is certain that the external input pattern matched a stored representation.

This threshold was set at the median firing rate to indicate an *unbiased prior* for choosing "same" or "different" and generally reflects having counterbalanced conditions within a psychological task. This threshold can be set differently according to task conditions. Furthermore, changing this threshold can reflect a bias which may arise with age or disease and this is explored later.

Data availability

Data will be made available on reasonable request.

Acknowledgments: This work was funded by the Wellcome Trust (Wellcome Trust PhD clinical fellowship to XYT and Wellcome Trust Principal Research Fellowship to MH) and MRC (Clinician Scientist Fellowship to S.M.) and by the Oxford NIHR Biomedical Research Centre.

Author contributions: Design, analysis, manuscript writing, critical revisions (XYT); critical revisions of the manuscript (AS, MH); design, statistical analysis and critical revisions of the manuscript (SM)

Competing Interests

The authors declare no competing interests

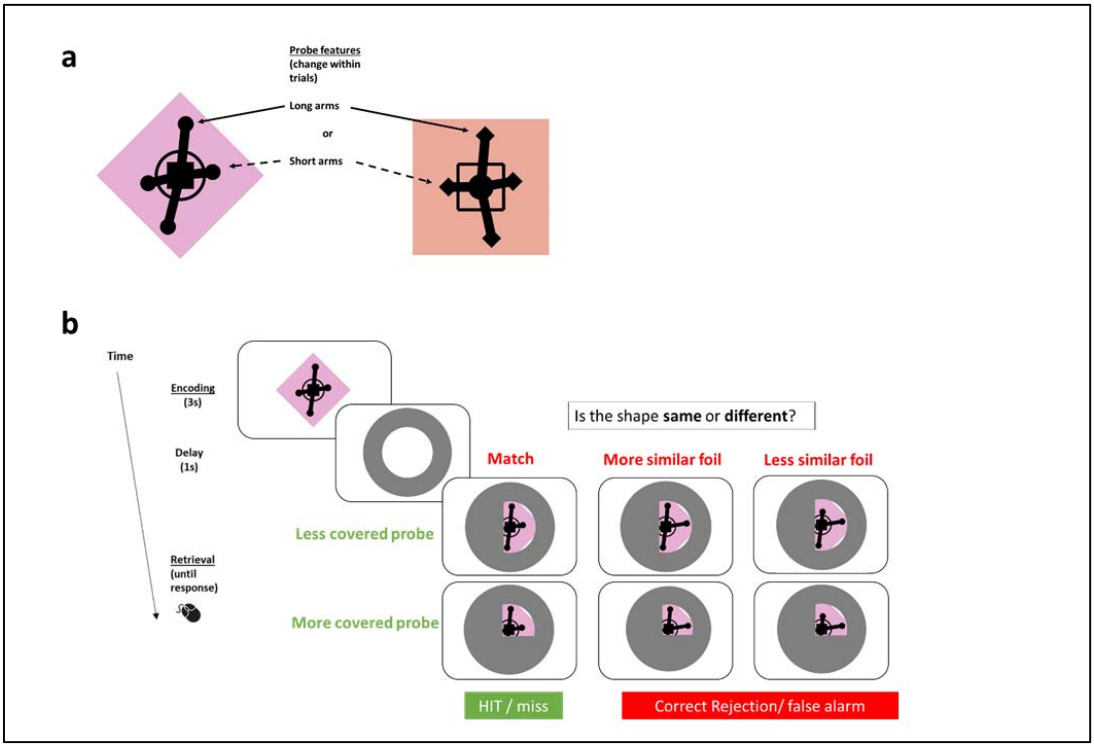


Figure 1. Pinhole memory task design

a Example of two stimuli used in the Pinhole memory task. Each stimuli had four arms which were either long or short which were used to manipulate similarity between encoding stimuli and probe. Other features included a coloured background shape as well as a middle, intermediate and end-arm shape which were not changed within a trial. **b** Main working memory task design showing a single encoding shape followed by a single probe. The probe would either be a match, more similar foil (different by one arm) or less similar foil (different by two arms). Probes were either more or less covered.

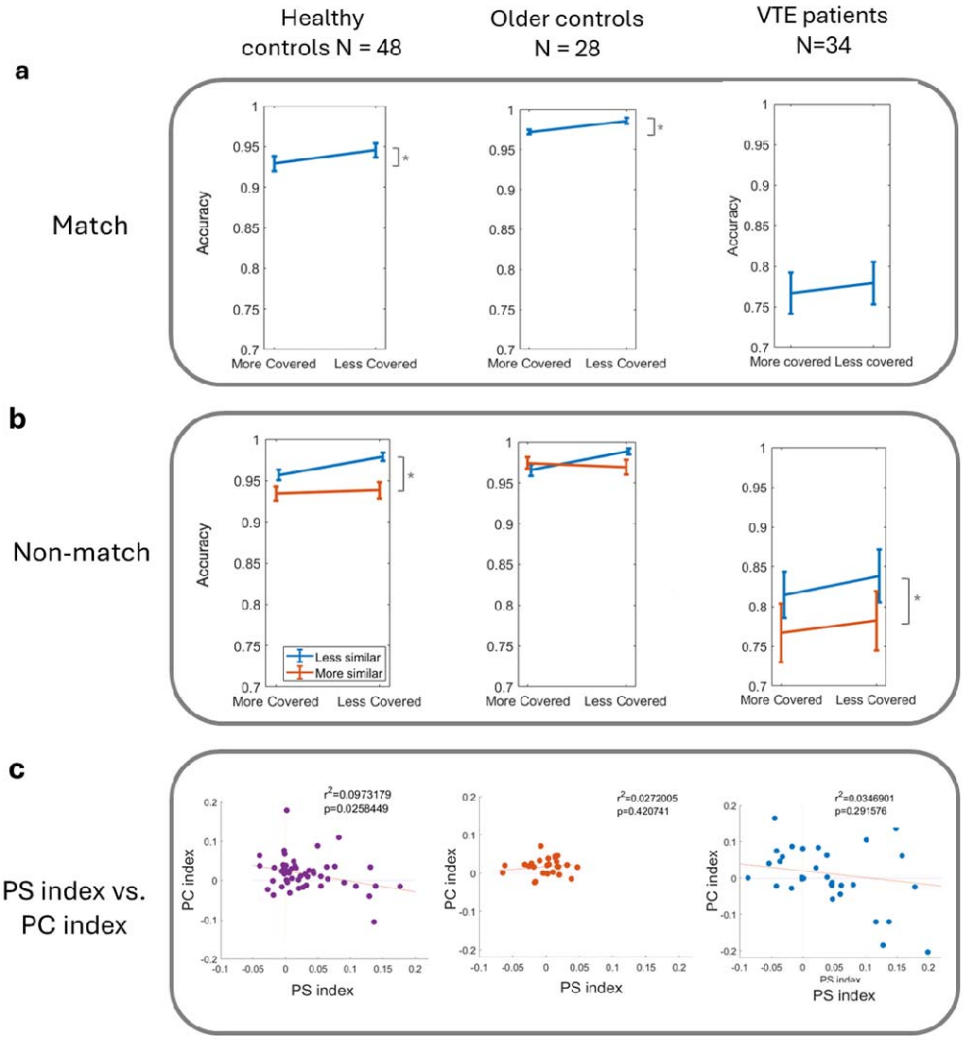


Figure 2. Pinhole memory task results across healthy young, older and people with epilepsy.

a Accuracy was significantly higher in match trials for probes which were less covered in the healthy young [F(1, 54) = 11.30, p=0.001] and healthy older cohorts [F(1, 54) = 6.25, p=0.012] but not people with epilepsy [F(1, 64) = 2.58, p=0.078]. **b** For non-match trials, healthy young [F(1, 94) = 5.43, p=0.02] and people with epilepsy [F(1, 64) = 4.83, p=0.042] had higher accuracy for probes which were less similar. **c** Correlation between PS index and PC index across the three cohorts showing a weak correlation for healthy young individuals only (r^2= 0.08, p = 0.02). *PS- pattern separation, PC- pattern completion.*

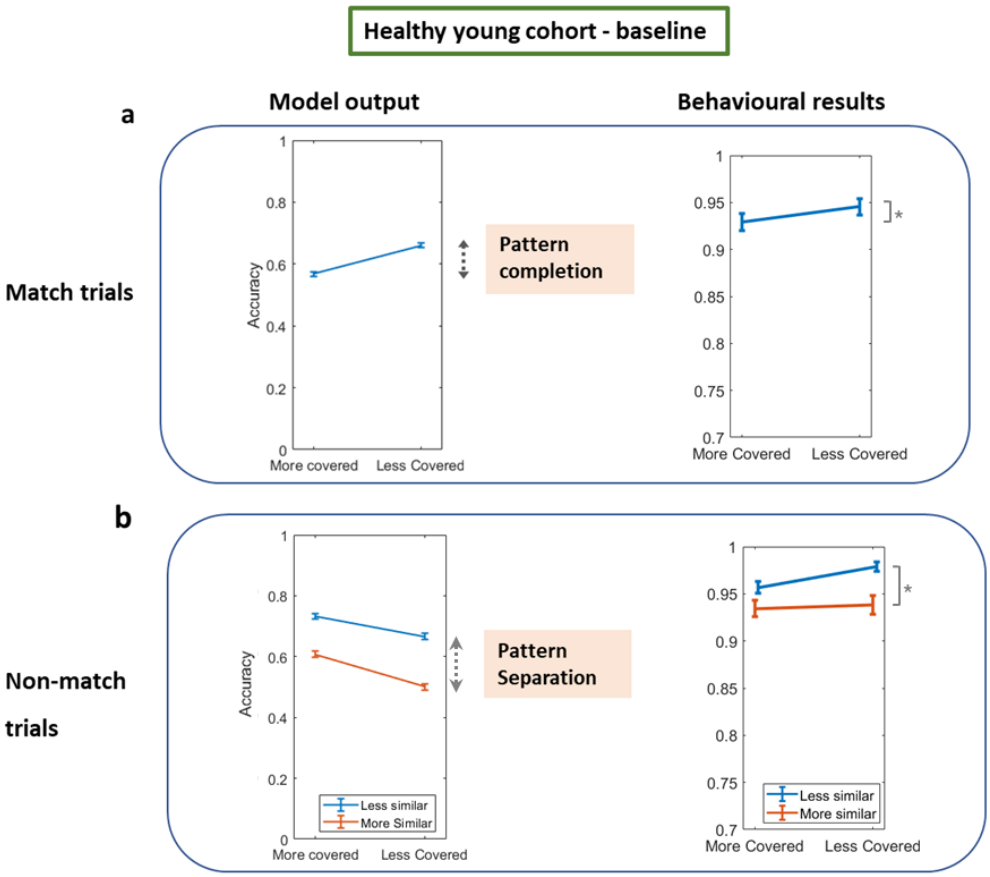


Figure 3. Modelling the Memory Pinhole Task in "healthy young" individuals

a Match trial model performance shows higher accuracy for less covered probes compared to more covered probes (pattern completion effect), consistent with behavioural results from the Memory Pinhole Task. **b** For non-match trials, the model performed better for less similar probes reflecting a pattern separation effect also seen in healthy younger individuals.

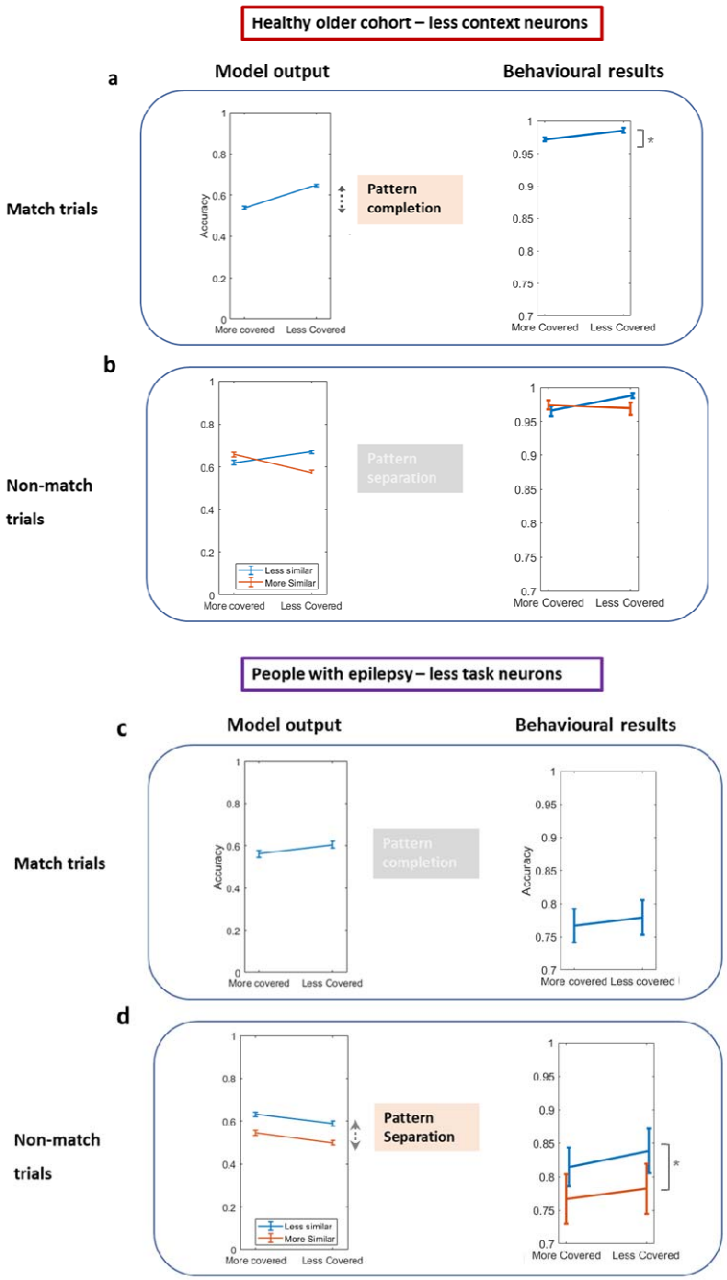


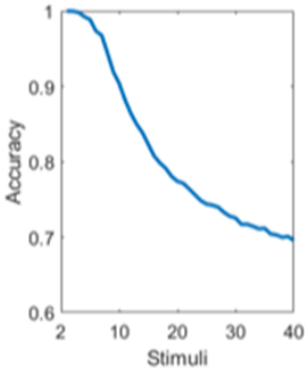
Figure 4. Modelling the Pinhole Memory task in the healthy older and epilepsy cohort.

The healthy older cohort behaviour was simulated by, **a**, reducing context neurons led to higher match trial model performance for less covered probes compared to more covered probes (pattern completion effect). **b** For non-match trials, different levels of probe similarity was not associated with an overall performance difference or clear pattern separation effect. Results from the epilepsy cohort was best simulated by reducing task neurons which, **c**, resulted in a smaller pattern completion effect (when performance is better with less covered probes). Correspondingly, behavioural results did not show a significant difference in accuracy between more and less covered match probes. **d** For non-match trials, better performance was associated with less similar compared to more similar probes.

Box 1. Modified Hopfield Network demonstrates fundamental principles of memory

A model of memory encoding and retrieval should be able to simulate key characteristics of both short and long-term memory. For example, the 'set-size' or 'list-length' effect is principally demonstrated by a reduction in performance with increasing number of items to be remembered. Behaviourally, this can be reflected by several metrics such as worse accuracy, lower memory precision or higher reaction times.

Our model demonstrates 'set-size' effect, impact of reducing model neurons as well as the effect of obscuring probe dimensions and altering similarity on model performance:

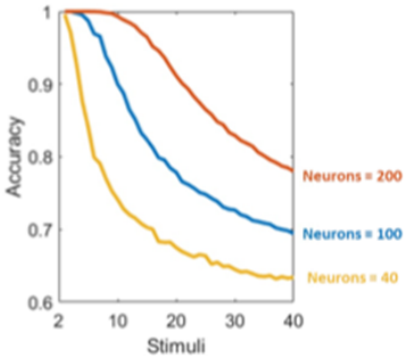


Set-size effect

Accuracy for recalling the correct probe components decreased with larger set size.

This simulation used a base model architecture with neurons N = 100 (plus 100 context neurons) and tested the network with sets increasing stimuli number from two to 40 stimuli.

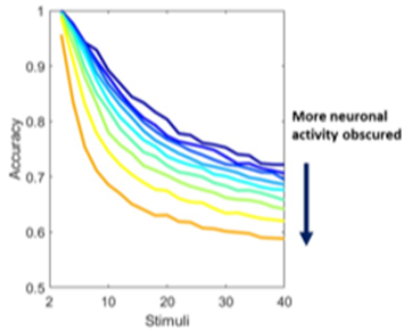
At recall, the model was given half covered probes and assessed on the accuracy for recreating the covered components.



Reducing neuron population

Three different network architectures with N = 200, N = 100 and N = 40 across increasing set sizes. Smaller network size was associated with a decrease in model accuracy for a given set size.

Retrieval errors increase when the number of stored patterns exceeds a proportion of the network size.



Obscuring neuron components

During recall, increasing the proportion of obscured neurons (5% - 50% obscured) is associated with reduced model accuracy.

This reflects the typical 'pattern completion' effect where performance decreases as less of the probe is available.

References

1. Rolls, E. T. A theory of hippocampal function in memory. *Hippocampus* **6**, 601–620 (1996).
2. Willshaw, D. J. & Buckingham, J. T. An assessment of Marr's theory of the hippocampus as a temporary memory store. *Philos Trans R Soc Lond B Biol Sci* **329**, 205–215 (1990).
3. O'Reilly, R. C. & McClelland, J. L. Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus* **4**, 661–682 (1994).
4. Marr, D. A theory for cerebral neocortex. *Proc R Soc Lond B Biol Sci* **176**, 161–234 (1970).
5. Neunuebel, J. P. & Knierim, J. J. CA3 retrieves coherent representations from degraded input: Direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron* **81**, 416–427 (2014).
6. Knierim, J. J. & Neunuebel, J. P. Tracking the Flow of Hippocampal Computation: Pattern Separation, Pattern Completion, and Attractor Dynamics. *Neurobiol Learn Mem* **129**, 38 (2016).
7. Yassa, M. A. & Stark, C. E. L. Pattern separation in the hippocampus. *Trends in Neurosciences* vol. 34 515–525 Preprint at https://doi.org/10.1016/j.tins.2011.06.006 (2011).
8. Kim, J. & Yassa, M. A. Assessing Recollection and Familiarity of Similar Lures in a Behavioral Pattern Separation Task. *Hippocampus* **23**, 287 (2013).
9. Leal, S. L. & Yassa, M. A. Integrating new findings and examining clinical applications of pattern separation. *Nat Neurosci* **21**, 163–173 (2018).
10. Stark, S. M., Yassa, M. A., Lacy, J. W. & Stark, C. E. L. A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia* **51**, 2442–2449 (2013).
11. Toner, C. K., Pirogovsky, E., Kirwan, C. B. & Gilbert, P. E. Visual object pattern separation deficits in nondemented older adults. *Learn Mem* **16**, 338–342 (2009).
12. Holden, H. M., Toner, C., Pirogovsky, E., Kirwan, C. B. & Gilbert, P. E. Visual object pattern separation varies in older adults. *Learning & Memory* **20**, 358 (2013).
13. Shelton, D. J. & Kirwan, C. B. A possible negative influence of depression on the ability to overcome memory interference. *Behavioural Brain Research* **256**, 20–26 (2013).
14. Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T. & Stark, C. E. L. Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using highresolution fMRI and variable mnemonic similarity. *Learning and Memory* **18**, 15–18 (2011).
15. Anderson, M. L., James, J. R. & Kirwan, C. B. An event-related potential investigation of pattern separation and pattern completion processes. *http://dx.doi.org/10.1080/17588928.2016.1195804* **8**, 9–23 (2016).
16. Bakker, A., Kirwan, C. B., Miller, M. & Stark, C. E. L. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science (1979)* **319**, 1640–1642 (2008).
17. Kirwan, C. B. & Stark, C. E. L. Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning and Memory* **14**, 625–633 (2007).

18. Vieweg, P., Riemer, M., Berron, D. & Wolbers, T. Memory Image Completion: Establishing a task to behaviorally assess pattern completion in humans. *Hippocampus* **29**, 340–351 (2019).
19. Leal, S. L. & Yassa, M. A. Integrating new findings and examining clinical applications of pattern separation. *Nat Neurosci* **21**, 163 (2018).
20. Ally, B. A., Hussey, E. P., Ko, P. C. & Molitor, R. J. Pattern separation and pattern completion in Alzheimer's disease: Evidence of rapid forgetting in amnestic mild cognitive impairment. *Hippocampus* **23**, 1246–1258 (2013).
21. Lalani, S. J. *et al.* Impaired Behavioral Pattern Separation in Refractory Temporal Lobe Epilepsy and Mild Cognitive Impairment. *Journal of the International Neuropsychological Society* **28**, 550–562 (2022).
22. Reyes, A. *et al.* Impaired spatial pattern separation performance in temporal lobe epilepsy is associated with visuospatial memory deficits and hippocampal volume loss. *Neuropsychologia* **111**, 209 (2018).
23. Baker, S. *et al.* The Human Dentate Gyrus Plays a Necessary Role in Discriminating New Memories. *Current Biology* **26**, 2629–2634 (2016).
24. Rolls, E. T. The mechanisms for pattern completion and pattern separation in the hippocampus. *Front Syst Neurosci* **7**, 74 (2013).
25. Rolls, E. T. & Treves, A. The neuronal encoding of information in the brain. *Prog Neurobiol* **95**, 448–490 (2011).
26. Rolls, E. T. A quantitative theory of the functions of the hippocampal CA3 network in memory. *Front Cell Neurosci* **7**, (2013).
27. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* **79**, 2554–2558 (1982).
28. Folli, V., Leonetti, M. & Ruocco, G. On the maximum storage capacity of the hopfield model. *Front Comput Neurosci* **10**, 144 (2017).
29. Aimone, J. B., Deng, W. & Gage, F. H. Resolving New Memories: A Critical Look at the Dentate Gyrus, Adult Neurogenesis, and Pattern Separation. *Neuron* vol. 70 589–596 Preprint at https://doi.org/10.1016/j.neuron.2011.05.010 (2011).
30. Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M. B. & Moser, E. I. Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science (1979)* **305**, 1295–1298 (2004).
31. Rolls, E. T., Treves, A., Foster, D. & Perez-Vicente, C. Simulation studies of the CA3 hippocampal subfield modelled as an attractor neural network. *Neural Networks* **10**, 1559–1569 (1997).
32. Markov, N. T. *et al.* Age-related brain atrophy is not a homogenous process: Different functional brain networks associate differentially with aging and blood factors. *Proc Natl Acad Sci U S A* **119**, e2207181119 (2022).
33. Rolls, E. T. & Kesner, R. P. Pattern separation and pattern completion in the hippocampal system. Introduction to the Special Issue. *Neurobiol Learn Mem* **129**, 1–3 (2016).
34. Rolls, E. T. Advantages of dilution in the connectivity of attractor networks in the brain. *Biologically Inspired Cognitive Architectures* **1**, 44–54 (2012).
35. Rolls, E. T. Pattern separation, completion, and categorisation in the hippocampus and neocortex. *Neurobiol Learn Mem* **129**, 4–28 (2016).
36. Smith, D. M. & Bulkin, D. A. The Form and Function of Hippocampal Context Representations. *Neurosci Biobehav Rev* **0**, 52 (2014).

37. Rodgers, J. L., Nicewander, W. A. & Toothaker, L. Linearly independent, orthogonal, and uncorrelated variables. *American Statistician* **38**, 133–134 (1984).