Investigating Pattern Separation in the Medial Temporal Lobe through the Parametric Manipulation of Item Similarity.

by

Corey Loo

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Department of Psychology
University of Toronto

© Copyright by Corey Loo 2018
Investigating Pattern Separation in the Medial Temporal Lobe through the Parametric Manipulation of Item Similarity.

Corey Loo

Master of Arts

Department of Psychology
University of Toronto

2018

Abstract

Pattern separation is the essential ability to create separate memory representations for overlapping events. This process is accomplished by magnifying the dissimilarity between memories. Computational models posit that the degree to which memories become separated varies as a function of the similarity between them. However, there is little human evidence to support this claim. Using fMRI, we recorded from participants as they performed an explicit memory task in which we manipulated the similarity between study and test items. Our analysis revealed that manipulating similarity did not reliably influence the degree to which memories were separated. We instead found that pattern separation was related to how confident participants were when making decisions related to memory. The implications of this finding, however, remain unclear with future studies needed to interpret them. These results suggest that trial-wise confidence ratings might be a truer measure of mnemonic “similarity” than our group-average similarity measures.
Table of Contents

Abstract .......................................................................................................................... ii
Table of Contents ........................................................................................................ iii
List of Tables .................................................................................................................. v
List of Figures ................................................................................................................ vi
1 Introduction ................................................................................................................ 1
2 Methods ..................................................................................................................... 7
  2.1 Participants ........................................................................................................... 7
    2.1.1 fMRI Task ................................................................................................... 7
  2.1.2 Similarity Rating Task .................................................................................... 8
  2.2 Stimuli .................................................................................................................. 8
  2.3 Procedure ............................................................................................................ 8
    2.3.1 fMRI Task ................................................................................................... 8
    2.3.2 Similarity Rating Task .................................................................................... 9
  2.4 Similarity Analysis .............................................................................................. 10
  2.5 fMRI Scanning Procedures ................................................................................. 11
  2.6 fMRI Data Preprocessing and Alignment ......................................................... 11
  2.7 fMRI Analysis .................................................................................................. 11
3 Results ...................................................................................................................... 12
  3.1 Behavioural Results ........................................................................................... 12
  3.2 fMRI Results ..................................................................................................... 12
4 Discussion ................................................................................................................ 14
  4.1 Future Direction .................................................................................................. 17
  4.2 Conclusion .......................................................................................................... 17
References ..................................................................................................................... 19
Tables........................................................................................................................................21
Figures......................................................................................................................................25
List of Tables

Table 1. Summary of linear mixed model results analyzing the effect of trial type on activity... 22

Table 2. Summary of linear mixed model results analyzing the effect of similarity on activity.. 23

Table 3. Summary of linear mixed model results analyzing the effect of confidence rating on activity during lure trials. ................................................................. 24
List of Figures

Figure 1. Depicting the nonlinear transfer function of DG and CA3. ........................................ 26

Figure 2. A depiction of a trial sequence of the fMRI task......................................................... 27

Figure 3. A depiction of a trial sequence from the subjective and behavioural similarity rating task. ................................................................................................................... 28

Figure 4. Behavioural performance during the fMRI task.......................................................... 29

Figure 5. Graphic depiction of the effect of Similarity on activity.............................................. 30

Figure 6. Graphic depiction of the effect of confidence ratings on activity during lure trials ..... 31

Figure 7. Graphic depiction of the effect of confidence ratings on activity during repeat trials .. 32

Figure 8. Graphic depiction of the effect of “Perceived Oldness” ............................................... 33
1 Introduction

Every day we encounter new situations that share many overlapping details with our previous experiences. For example, imagine being asked to remember where you parked your car today, yesterday, and the day before. Because these events are so similar, the memory of where you parked your car on one day can easily be confused with where you parked your car on another. Thus, there is an apparent need for the ability to keep these memories distinct and resistant to confusion.

“Pattern separation” is the process that enables the creation of distinct (orthogonal) complex memory representations of events with common elements (O’Reilly & McClelland, 1994; McClelland, McNaughton & O’Reilly, 1995; O’Reilly, Norman, & McClelland, 1998; Norman & O’Reilly, 2003). Neural representations are made orthogonal by reducing the degree of overlap in similar mnemonic representations through the magnification of the differences between them (O’Reilly & McClelland, 1994). Without the process of pattern separation, memory representations with too much overlap would trigger catastrophic interference, resulting in the generalization of two different experiences (Norman, & McClelland, 1998; Norman & O’Reilly, 2003).

The hippocampus, particularly the dentate gyrus (DG) and CA3 subregions (O’Reilly, & McClelland, 1994), is one of the structures that is believed to support pattern separation. It is thought that these regions facilitate pattern separation by compressing incoming signals from the entorhinal cortex (ERC). The entorhinal cortex receives input from neocortical areas and preserves their signal by representing them on a one-to-one neurological basis, with each neuron in ERC receiving input from one neocortical neuron (O’Reilly & McClelland, 1994). Neurons in the entorhinal cortex then project directly to the DG and CA3 subregions. Though DG neurons greatly outnumber ERC neurons, with a single ERC neuron projecting to multiple DG neurons, a single DG neuron will receive inputs from multiple ERC neurons. This compression of information is thought to contribute to the sparsity of firing in the DG, producing neurons in the DG that are highly selective for conjunctions of features (O’Reilly & McClelland, 1994). The selective firing of DG neurons leads to minimal overlap between hippocampal representations. This compression is preserved in the CA3 region as DG cells project to CA3 via the perforant
path (O’Reilly & McClelland, 1994). Thus pattern separation is achieved in the hippocampus by recruiting distinct populations of neurons that are highly selective for conjunctions of features or combinations of inputs that are active in a particular experience (Norman, & McClelland, 1998).

These predictions of computational models were first tested in the rodent literature, though the initial group of studies appeared to produce conflicting results. By recording from place cells in the hippocampus, these studies have been able to investigate how place cell ensembles respond to changes in environmental cues. In one of the first of these studies Lee, Rao, and Knierim (2004) implanted recording electrodes in the CA3 and CA1 subregions of rats that were trained to circle clockwise on a circular track in a stable environment. Then on subsequent mismatch trials, the environmental cues were altered by rotating the landmarks that were placed along the behavioural track relative to the landmarks that were placed along the walls of the course. By analysing the neural ensembles representing each environment, it was revealed that the place cell ensembles representing the original environment differed from the ensembles representing the cue-altered environment in the CA1 region. However, in CA3, it was observed that place cell ensembles were more consistent between environments. This indicated that the place cells within CA3 were tolerant of changes, suggesting that these cells support the process of pattern completion (a competing process with pattern separation whereby complete representations are activated by a partial or degraded input).

Shortly after the publication by Lee and colleagues (2004), another electrophysiological study published by Leutgeb, Leutgeb, Treves, Moser, and Moser (2004) also examined cell ensembles in CA3 and CA1 in response to rooms with common spatial elements. However, rather than altering the local cues in the testing environment on subsequent mismatch trials, Leutgeb et al. used distinct testing environments that varied in their level of similarity. These testing environments either differed in their size (high similarity condition) or in their shape (low similarity condition). Recordings from CA3 revealed that separate ensembles were activated for the different rooms, regardless of how similar they were, consistent with the prediction that new representations are formed for similar environments by pattern separation. Meanwhile, place cell ensembles in CA1 responded in a graded fashion, with the amount of overlap in activated ensembles varying with the level of similarity between testing environments.
Initially, these two studies appear to provide conflicting results, with CA3 demonstrating patterns of activity consistent with both pattern separation and pattern completion. These findings seem to be reconciled by the findings of an immediate-early gene (IEG) brain imaging study conducted by Vazdarjanova and Guzowski (2004). As in the previous studies, subjects were made to explore cue-altered environments. However, Vazdarjanova and Guzowski varied the similarity between testing environments in a more graded fashion by altering the configuration or identity of local cues in the environment, the experimental room containing the environment (distal cues), or exposing them to two completely different environments. Histological analysis revealed that neural ensembles in CA3 had greater overlap for variations of the first testing environment (changes in the identity or configuration of local cues or changing distal cues) than in CA1. This pattern of results suggest that these partially modified environments were treated as more similar to the original environment in CA3 than in CA1, an activity pattern that is consistent with pattern completion. However, when rats were exposed to completely different environments, the overlap of neural ensembles was lower in CA3 than in CA1. This indicated that the representations for the two environments were orthogonalized in CA3, providing evidence for the occurrence of pattern separation. Interpreting the results of Lee et al. (2004) and Leutgeb et al. (2004) together, Vazdarjanova and Guzowski hypothesized that CA3 is capable of supporting either pattern separation or pattern completion, depending on the degree of change in the input stimuli.

Though computational models implicate the DG as the primary facilitator of pattern separation, the above studies had only investigated the CA1 and CA3 subregions. Addressing this gap of knowledge in a continuation of their work, Leutgeb, Leutgeb, Moser, and Moser (2007) conducted another investigation of place cell ensembles, this time recording from DG in addition to CA3. In contrast to their previous study, Leutgeb and colleagues exerted more control over the similarity in their testing environments by gradually morphing the shape of the exploratory enclosure from a square to a circle or vice versa. It was reported that small changes in the shape of the enclosure caused place cells in both CA3 and DG to either alter their firing rates in their preferred locations or adopt a new preferred location entirely. However, small changes did not induce either region to recruit an independent ensemble of neurons. When examining neural responses to large differences in shape, it was observed that a new ensemble of CA3 neurons were recruited to encode the environment. The results suggested that the hippocampus possessed a dual mechanism of orthogonalization, utilizing place cell remapping in response to small
changes of similarity in DG and CA3, while processing larger changes through the recruitment of new neural populations in CA3.

Despite the continued accumulation of evidence for hippocampal pattern separation in the rodent literature, few studies have investigated the neurological constituents of pattern separation in human subjects. The first empirical evidence for pattern separation in the human hippocampus was reported in a high-resolution fMRI study by Bakker, Kirwan, Miller, and Stark (2008). Bakker and colleagues utilized an incidental encoding task in which a series of images were presented to participants while they were in a scanner. The images presented were either novel, a repetition of a previously shown image, or a slightly different version of a previously shown image (lure images). Bakker and colleagues obtained evidence for pattern separation by leveraging the repetition suppression effect, whereby levels of brain activity are reduced during a repeated presentation of a stimulus compared to its first presentation. They reasoned that regions performing pattern separation would treat a similar object more like a new stimulus rather than a repetition, and thus show activity similar to that of a first presentation. In line with the predictions advanced by computational models, the only interrogated region to demonstrate this pattern of activity was DG/CA3, thus providing evidence for the participation of the DG/CA3 region in the facilitation of pattern separation in the human hippocampus.

In an extension to this work, Lacy, Yassa, Stark, Muftuler, and Stark (2011) sought to define the transfer function of human hippocampal activity by parametrically varying the similarity of the images shown. Mnemonic similarity ratings of each image pair were obtained (mnemonic similarity was evaluated by taking the probability of responding to a lure stimuli as “old” during a behavioural recognition task) and activity in CA1 and CA3/DG was contrasted between high similarity items with low similarity items by performing a median split on item similarity. This contrast revealed that CA1 responded in a graded fashion, with activation in this area increasing with the degree of change between the lure and target items. In contrast, CA3/DG responded in a more stepwise fashion, experiencing a large change in activity in response to a small change in the input, indicating that orthogonal representations were being created for highly similar images. In order to assess how much change is needed to elicit this orthogonalization in DG/CA3, Lacy and colleagues further separated lures into smaller similarity bins. Though this procedure reduced the signal to noise ratio, due to the limited number of trials for each bin, a
large change in activity occurred for items in the bin with the highest similarity. Thus, it was concluded that CA3/DG is sensitive to even small amounts of change in input.

There have been a limited number of attempts to extend the research investigating the effect of parametric manipulation of similarity on pattern separation in the human hippocampus. To our knowledge, the only other study to further this research was conducted by Motley and Kirwan (2012) who investigated how top-down task demands influenced pattern separation. Instead of varying the similarity of their stimuli by using high similarity and low similarity lures, Motley and Kirwan directly controlled item similarity by using parametrically rotated objects in both an incidental encoding task and a continuous recognition task (intentional encoding). To delineate regions of interest, a contrast of activity for new objects compared to old objects was applied. This contrast identified the right and left hippocampus and the right and left parahippocampus as memory sensitive-regions. The authors next fit a power function to the activity patterns for each of these regions, using degree of object rotation as the predictor. Regions were identified as performing pattern separation if their pattern of activity was best defined by a power curve with decreasing slope (activity changes were sharp in response to small changes in rotation). Regions were identified as performing pattern completion if patterns of activity were best defined by either a linear function (activity changed proportional to the change in angle of rotation) or a power function with increasing slope (activity for the small rotation conditions was similar to the repeat condition but marked differences in activity were observed in large rotation conditions).

The resulting analysis revealed that the left hippocampus, as well as the right and left parahippocampal cortex, exhibited power functions consistent with pattern separation, whereas the right hippocampus exhibited a power function that was consistent with pattern completion. In the intentional condition, the right hippocampus and right parahippocampal cortex demonstrated power functions consistent with pattern completion whereas power functions exhibited by the left hippocampus and parahippocampal cortex were more consistent with pattern completion. These results were taken to indicate that task demands modulate the laterality of pattern separation.

Although these previous studies have attempted to define the transfer function of hippocampal activity to input similarity, each of them had limitations. Motley and Kirwan (2012) directly manipulated the perceptual similarity of their stimuli by parametrically rotating them. Their analysis of the imaging data, however, treated the entire hippocampus as a single ROI, without
sectioning it into its subregions. Thus, they could not examine how the transfer function may vary across the different regions of the hippocampus. On the other hand, Lacy et al. (2011) looked within the DG/CA3 and CA1, and though they did parametrically vary the mnemonic similarity of their stimuli, they were limited by the number of trials in each similarity bin, thus reducing the signal to noise ratio.

The goal of the current study was to examine the transfer function between probe similarity and activation in the subregions of the hippocampus. Additionally, we examined this transfer function in the surrounding medial temporal cortical areas. Specifically, we defined the CA3/DG region, CA1, entorhinal cortex, perirhinal cortex, subiculum, anterior hippocampus, and posterior hippocampus as our regions of interest. To accomplish this goal, we performed linear mixed-effect model regression on brain activity obtained at the presentation of lure stimuli using rating of similarity as a regressor.

The current research utilized data obtained from an experiment that we previously conducted. Participants were scanned using functional MRI (fMRI) while being probed for their memory of studied images. During the encoding phase of the task, participants were shown a continuous stream of images and had to indicate if each image had been seen on the previous trial. Each image was identified by an accompanying unique title. During the retrieval phase, participants were given the title, visualized these images and were subsequently probed for their memory of them. Although the visualization aspect of the experiment is not the focus of the proposed research, the paradigm was designed to examine how the vividness of an individual’s mental imagery influenced the processes of pattern completion and pattern separation. Each trial began with the presentation of a cue that identified an image from the encoding phase. Participants visualized the image that matched the title, and then rated the vividness of their visualization. They were then probed for their memory of this image, being shown the image that was previously associated with the cue or an image that was similar to it. Participants were required to indicate whether this image was old or new, and then give a confidence rating on this judgement.

In order to examine the transfer function between probe similarity and brain activity, a parametric measure of similarity of our stimuli is needed. Using a separate set of participants, we derived two measures of similarity: subjective ratings of similarity, behavioural rating of
similarity. These measures were later combined into a single measure using a principle component analysis. Subjective ratings of similarity were obtained by having participants view pairs of similar images and indicating how similar the two images were using a five-point Likert scale. Behavioural ratings of similarity were gauged using reactions times obtained from a short-term recognition memory task. Participants were shown two images separated by a one second delay. In this task, the second image presented was either identical to the first, or it was a visually similar image. Participants were required to make a speeded response, indicating whether or not the two images were identical. From this task, we used reaction times as a proxy for image similarity. We predicted that a highly similar image would take longer to reject than an image that was less similar. As a third measure of similarity, we incorporated accuracy data from this short-term memory task.

With these data, we used single trial regression to estimate fMRI activity for each trial of the experiment in each ROI and subject. This yielded a set of beta coefficients reflecting the estimated probe-related activity at each trial. This set of beta coefficient were further analyzed to see if they were modulated by our similarity measures. A separate linear mixed effects model was run for each ROI to test for this relationship.

Based on the predictions made by computational models (O’Reilly & McClelland, 1994), an ROI performing pattern separation should show a large change in activity in response to a small change in item similarity (see Figure 1). Therefore, we anticipated that this pattern of activity would be characterized by a power function with decreasing slope (Yassa and Stark, 2011). Our strongest predictions were for subregions DG/CA3 and CA1. Based on the results of Lacy et al. (2011), we hypothesize that activity in the DG/CA3 subregion will be best modeled by a power function with decreasing slope, whereas activity in the CA1 subregion will be best modeled by a linear function.

2 Methods

2.1 Participants

2.1.1 fMRI Task

Forty-three adults (19 males and 24 females) with normal or corrected-to-normal vision and no history of neurological or psychiatric disease were recruited through the Baycrest subject pool,
tested, and paid for their participation per a protocol approved by the Rotman Research Institute’s Ethics Board. Subjects were native or fluent English speakers and had no contraindications for MRI. Data from nine of these participants were excluded from analysis: three participants did not complete the memory task, one left the experiment before completion, and five participants had excessive head motion in the scanner. As a result, 34 participants were included in the study (15 males and 19 females).

2.1.2 Similarity Rating Task

We recruited a separate set of 35 participants through the Baycrest subject pool, tested, and paid for their participation in a protocol approved by the Rotman Research Institute’s Ethics Board.

2.2 Stimuli

One hundred and twenty pairs of visually similar images were gathered from online sources and resized to 757 by 522 pixels. Images contained a wide variety of subjects, including, people, scenes, animals, and objects. For the fMRI task, these image pairs were grouped into four sets of 30 images pairs. One of the four sets of image pairs was reserved for practice.

2.3 Procedure

2.3.1 fMRI Task

fMRI scanning consisted of three encoding runs and three retrieval runs, which were interleaved with one retrieval run following every encoding run (see Figure 2). Retrieval runs only tested images that were presented during the preceding encoding run. Encoding trials began with the presentation of a cross-hair (font size = 50) for 1680 ms. Following the fixation period, an image appeared in the center of the screen for 1820 ms. Images occupied 757 by 522 pixels of a 1024 by 768 pixel screen. Each image was accompanied by an image title (e.g. “monkey”, foosball table”) shown in the top portion of the screen. On each trial, participants indicated if the displayed image had been presented in the previous run of images. Each encoding run utilized a different set of thirty image pairs using only one image from each pair within a set. Each image was shown four times per run. One repetition occurred immediately after an image’s first presentation. Subsequent repetitions could only occur after at least five trials since the image’s
last presentation. In total, there were 120 trials in each encoding run, lasting 3500 ms each. Overall, a single encoding run lasted seven minutes.

There were two trial types during retrieval runs: repeat trials and lure trials. Each retrieval trial began with the presentation of an image title, shown in the center of the screen, for 1000 ms. For the next 6000 ms, the title was replaced by an empty rectangular box with edges corresponding to the edges of the stimulus images (757 by 522 pixels), displayed in the center of the screen. Participants were instructed to visualize the image that corresponded to the title as accurately and in as much detail as they could within the confines of the box. Once the box disappeared, participants had 2000 ms to give a vividness rating of their mental image on a 1-4 Likert scale using a four-button fiber optic response boxes placed in their right hand (1 = index finger; 4 = pinkie finger). Following this vividness rating, participants were presented with a memory probe that lasted 2000 ms. On repeat trials, this probe was identical to the image that was presented with the title in the preceding encoding run. On lure trials, this probe was the studied image’s corresponding visually similar image. Participants responded to this probe by indicating if it was identical to the image that corresponded to the image title, or if it was a new image, again using the fibre optic response box (1 = identical, 2 = new). Following the disappearance of the memory probe, participants had 2000 ms to give a confidence rating on this old/new judgement image on a 1-4 Likert scale, again using the fiber optic response boxes (1 = index finger; 4 = pinkie finger), with a rating of one indicating low confidence and a rating of four indicating high confidence. The intertrial interval was jittered at 1000-3000 ms with an average of 2000 ms. Each run consisted of 30 trials, each lasting 15 seconds. Thus, each retrieval run lasted 7.5 minutes. Retrieval runs consisted of 15 repeat trials and 15 lure trials. The order of presentation (which image was presented during the encoding run) was counter balanced across subjects so that each image of each pair was presented an equal number of times in each position.

### 2.3.2 Similarity Rating Task

On each trial of the subjective rating task, participants were presented with an image pair and were required to judge how similar the images were (see Figure 3). Each trial began with the presentation of an image for 500 ms, followed by a visual mask for 1000 ms. Following the disappearance of the visual mask, a second image was displayed. The second image was the corresponding visually similar image. Participants were prompted to rate the similarity of the two
images on a scale of 1-5 (with 1 indicating low similarity, 5 indicating high similarity) by pressing the corresponding keys on the keyboard. This display lasted until the participant responded. Each of the 180 images (90 image pairs) was shown as the first image in a trial. Thus, there were 180 trials in the subjective rating task.

In the behavioural rating task, participants performed a short-term change detection task (see Figure 3). Reaction times for this response were recorded as a measure of similarity. We expected that highly similar images would take longer to reject when presented as lures. Thus, reaction times should increase with the similarity of the images within a pair. Each trial began with the presentation of an image in the center of the screen for 500 ms followed by a visual mask for 1000 ms. Following the disappearance of the visual mask, a second image was displayed. On repeat trials, the second image was identical to the first image. On lure trials, the second image was the corresponding similar image. Participants were required to make a speeded response, indicating whether or not the two presented images were identical. Participants made their response using a keyboard and were instructed to press “1” if the images were identical or “2” if they were not. Participants were given 3000 ms to make this response before the trial timed out. Each trial was separated by a fixation period with a 1000 ms duration. Each of the 90 image pairs was presented in both repeat and lure trials. This was repeated again with the presentation order swapped so that each image in each pair was presented as the first image. Thus, there were 360 trials for in the behavioural rating task.

2.4 Similarity Analysis

The similarity rating tasks produced three measures of similarity: subjective ratings of similarity, short-term change detection reaction times, and short-term change detection accuracy. Similarity ratings for each image pair were obtained by averaging scores across all participants for each measure. In order to obtain a combined metric of similarity, we ran a principal component analysis using three measures of similarity. Eigen-values indicated that the first three factors explained 71%, 18%, and 10% of the variance. We took the first eigenvector values as our combined measure of similarity.
2.5 fMRI Scanning Procedures

Participants were scanned with a 3.0-T Siemens MAGNETOM Trio MRI scanner using a 12-channel head coil system. A high-resolution gradient-echo multi-slice T1-weighted scan coplanar with the echo-planar imaging scans (EPIs) was first acquired for localization. Functional images were acquired using a two-shot gradient-echo T2*-weighted EPI sequence sensitive to BOLD contrast (22.0 x 22.0 cm field of view with a 110 x 110 matrix size, resulting in an in-plane resolution of 2.00 x 2.00 mm for each of 63 2.00-mm axial slices; repetition time = 1.77s; echo time = 30ms; flip angle = 65 degrees). A high-resolution whole-brain magnetization prepared rapid gradient echo (MP-RAGE) 3-D T1 weighted scan (160 slices of 1mm thickness, 19.2 x 25.6 cm field of view) was also acquired for anatomical localization.

2.6 fMRI Data Preprocessing and Alignment

All statistical analyses were first conducted on realigned functional images in native EPI space. Functional images were converted into NIFTI-1 format, motion-corrected, and realigned to the average image of the first run with AFNI’s 3dvolreg program, and smoothed with a 4-mm FWHM Gaussian kernel. The maximum displacement for each EPI image relative to the reference image was recorded. Regions of interest (ROIs) were delineated by applying a probabilistic template (defined using a 50% overlap cutoff) derived from the manual tracing of the medial temporal lobes of an independent sample of 20 young adults. This resulted in the identification of eight ROIs in each hemisphere: CA3/DG region, CA1, entorhinal cortex, perirhinal cortex, the subiculum, anterior hippocampus, and posterior hippocampus. These ROIs were then overlaid with a global mask that demarcated all the voxels within each participant’s brain. This resulted in the exclusion of voxels that exhibited signal dropout from further analysis.

2.7 fMRI Analysis

fMRI activation was examined by extracting a time-series from each ROI for each trial and convoluting it with the SPM-canonical hemodynamic response function. Activity was averaged over all voxels within a region, yielding a single time-course per region for each subject. Signal drift was accounted for by fitting a locally linear smoothing spline to the time series data for each ROI. The resulting estimates were then subtracted from the time series data, centering each time series around its mean activity. A single trial regression was then applied to estimate fMRI
activity for each trial of the experiment in each ROI and subject. This regression analysis yielded a vector of beta coefficients for each ROI, reflecting the estimated probe-related activity at each trial in that region. The beta coefficients for each ROI were separately analyzed to investigate how they were related to our similarity measures by analyzing them with a linear mixed effects model. This model was fit to our composite measure of image similarity to determine how well the measure predicted changes in activity. From this analysis, ROIs performing pattern separation were defined as those exhibiting a transfer function with decreasing slope whereas ROIs performing pattern completion were defined as those exhibiting a transfer function with increasing slope (see Figure 1 for a depiction of these transfer functions).

3 Results

3.1 Behavioural Results

Figure 4 shows the mean accuracy for each trial type. Percentage correct was calculated as the proportion of trials for which participants correctly identified the object (i.e., the number of times the participant identified an old object as old divided by the total number of old stimuli). Participants were highly accurate at identifying repeated and similar stimuli (80% and 82% for repeat and lure trials, respectively).

3.2 fMRI Results

Our primary objective was to examine the response function of the medial temporal lobe to parametric manipulation of stimulus similarity during lure trials. Because this investigation leverages the repetition suppression effect (whereby activity is decreased upon the presentation of a previously seen item compared to a novel one), our first analysis sought to confirm that a repetition suppression effect was observed in the current study. To that end, we used R (R Core Team, 2012) and lme4 (Bates, Maechler & Bolker, 2012) to perform a linear mixed effects analysis collapsing across all ROIs to examine the relationship between activity and trial type, contrasting activity between repeat trials and lure trials. As a fixed effect, we entered trial type (lure, repeat) into the model. As random effects, we included intercepts for subjects and ROI, as well as by-subject and by-ROI random slopes for the effect of trial type. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained by likelihood ratio tests of the full model with the effect of trial type and the
model without the effect of trial type. This contrast revealed that activity was significantly greater on lure trials compared to repeat trials ($\chi^2(1) = 8.211, p < .01$). We then examined this effect in each ROI individually by conducting another set of mixed effect analyses. This analysis revealed that even though every ROI trended toward greater activity for lure than repeat trials, this effect was not significant in every ROI (the results of this analysis are summarized in Table 1). Namely, a significant repetition suppression effect was observed in right and left anterior hippocampus, right and left CA3/DG, left PRC, right CA1, right ERC, and right PHC. The results of these contrasts reflect what would be expected of a repetition suppression effect and reproduce the findings of previous studies (Bakker et al., 2008; Motely and Kirwan, 2012).

With support for the occurrence of a repetition suppression effect, our next step was to return to the main objective of the current research, to investigate the relationship between pattern separation and the degree of similarity during retrieval runs. We performed a linear mixed effects analysis collapsing across all ROIs to investigate of the relationship between activity and image pair similarity. We entered similarity into the model as a continuous covariate. Intercepts for subjects and ROI were included as random effects, as well as by-subject and by-ROI random slopes for the effect of similarity. P-values were obtained by likelihood ratio tests of the full model with the effect in question and the model without the effect in question. The effect of similarity on activity was non-significant ($\chi^2(1) = 0.097, p = 0.75$), with activity decreasing by $-0.005 \pm 0.016$ (standard errors) for each unit of similarity. We further investigated this relationship in each ROI individually by conducting another set of mixed effect analyses, this time omitting ROI as a random effect. These analyses revealed that no ROI exhibited a significant effect of similarity (results are summarized in Table 2. See Figure 5 for a graphical depiction of the trends). These results indicate that the repetition suppression effect did not vary with the degree of similarity, suggesting that the activation in MTL ROIs was insensitive to our manipulation of similarity.

We next sought to determine if any of our other measures could better model the observed patterns of activity during lure trials. We now added confidence ratings into the model as an additional continuous covariate. Intercepts for subjects and ROI were included as random effects, as well as by-subject and by-ROI random slopes for the effect of confidence ratings. The effect of confidence during lure trials was significant, ($\chi^2(1) = 5.792, p = .0161$), increasing activity by
$0.07586 \pm 0.03005$ (standard errors) for each rank of confidence. We then examined this effect in each ROI individually by conducting another set of mixed effect analyses, this time omitting ROI as a random effect. These analyses revealed that, although we observed a positive linear effect of confidence in every ROI, (see Table 3 for a summary of the results. See Figure 6 for a graphical depiction of the trends), significant effects were only observed in: right and left anterior hippocampus, right and left CA1, right and left CA3/DG, right and left ERC, and right PRC.

We sought to see if this effect of confidence rating was present during repeat trials, using the previous linear mixed effects model on repeat trial activity data. The effect of confidence was non-significant on repeat trials ($\chi^2(1) = 0.3071, p = .58$), increasing activity by $0.01131 \pm 0.02039$ (standard errors) for each rank of confidence (see Table 4 for a summary of results. See Figure 7 for a graphical representation).

4 Discussion

The present research was conducted in order to investigate the response gradient of the hippocampus as a function of image similarity. To that end, we recorded brain activity using fMRI as we had participants perform a short-term memory task in which we parametrically varied the similarity between study and test items. Although some ROIs exhibited activity trends that decreased linearly with image similarity, analysis of these trends revealed that none of them was significant. Subsequent analysis revealed that activity in some hippocampal regions was significantly modulated by confidence ratings. Thus while we fail to provide evidence for a response gradient of the hippocampus as a function of image similarity, we instead found evidence for a response gradient as a function of confidence ratings.

The fact that we observed a robust repetition suppression (or novelty) effect, but a null effect of similarity is puzzling. It simply could mean that the process of pattern separation is not a graded response, but rather occurs when a threshold of dissimilarity has been reached. This interpretation implies that new representations are formed when an event contains a minimum degree of mismatch to a stored memory. This interpretation, however, does not agree with the findings of previous research in both humans and rodents. The previous research investigating pattern separation have reported that the process occurs in a graded response (Lacy et al., 2011;
Motley and Kirwan, 2012). Thus there is reason to believe that our null finding is possibly due to task-specific reasons or lack of statistical power.

It is possible that our stimulus set and associated similarity metric did not sample a large enough range of stimulus pair similarity to detect an activation gradient in the hippocampus. Computational theories propose that the process of pattern separation can be modeled by a power function, depicting the relationship between the degree of difference in input and the degree of difference in the resulting mnemonic representations. Power functions depicting pattern separation are shown in the top half of Figure 1. If our range of similarity is too narrow and includes only a small portion of the input scale, we would only able to observe a small portion of the response curve. Thus, it may be possible we are not inducing a large enough change in input to observe a measureable change in output, given the relatively low signal-to-noise regime of fMRI.

Another potential issue is that our task was non-standard insofar as subjects were pre-cued with the image label and performed an imagery task preceding the presentation of the probe. The imagery task was included because the task was not designed to measure pattern separation per se, but rather the reactivation of episodic memories in the preceding visualization period. Nevertheless, it is not immediately clear why the imagery period would necessarily alter the response gradient of the hippocampus as a function of image similarity.

Our analysis is conducted through single trial regression, producing a beta coefficient for each trial. The influence of the imagery portion of each trial, however, was not examined for its influence on the probe response. Thus, the imagery portion in our task could affect the subsequent probe response in an unknown way, with signal related to imagery carrying over to the onset of the memory probe. At the current stage of the analysis, we are not yet able to disentangle the effects of imagery from those of the probe presentation. One potential way to assess the effect of the preceding delay mental imagery period is to use the acquired vividness ratings as a covariate to test whether the strength of the retrieved memory impacts the subsequent response to the probe image.

A further finding from our investigation was the modulation of brain activity by confidence ratings. This effect of confidence was observed bilaterally in ROIs encompassing subregions of
the hippocampus, bilateral ERC, and right PHC. This effect, however, was only observed on trials in which participants correctly identified the memory probe as a lure. There are a number of explanations that could account for this result.

One possible explanation that we considered is that this pattern of results reflects signals related to judgements of novelty, recognition, and recollection. Models of recollection and recognition have proposed that the level of activity in regions supporting these processes scales as a function of perceived newness for novelty signals, or oldness for recollection and recognition, with ratings of confidence reflecting the perceived oldness or newness (Yonelinas, 2002; Daselaar, Fleck, and Cabeza, 2006). The results we observed in the current study align with what would be expected of a novelty signal, whereby activity increases with confidence of perceived newness on trials in which a lure item is presented. Previous experiments, however, have shown that this trend should continue to decrease with the level of perceived oldness (Daselaar et al. 2006) (See Figure 8). It is possible that there is a dissociation among regions that support retrieval process (Daselaar et al. 2006), with different regions making separate contributions to each process. Although a triple dissociation within the MTL has been observed by Daselaar and colleagues, we only observed what would constitute to a novelty signal. In addition, we failed to provide evidence for recollection or recognition, even in regions where Daselaar et al. found evidence for signals related to judgements of familiarity or recollection.

Alternatively, the results of the current research can be interpreted through the lens of a novelty-encoding account (Tulving, Markowitsch, Craik, Habib, & Houle, 1996). Novelty encoding accounts posit that items that are deemed more novel are encoded more strongly, reflected in an increase in activity when participants study an new item. This interpretation would account for the lack of an effect of confidence in repeat trials. In such trials, the presented memory probe is identical to a studied item, thus there is nothing new to encode. This account predicts that items that are deemed more novel, and thus receive stronger encoding, will subsequently be more likely to be remembered in a secondary memory test. Testing this account would require a follow up study to the current research, one that implements a secondary memory test for items presented during lure trials. If lure items that induced a greater activation are in fact encoded more strongly, they should be more likely to be remembered during the secondary memory test.
Another potential limitation of our design is that it only employs an explicit memory task. This limitation has two implications. The first is that an explicit memory task provides subjects with a processing goal. Because of this, we would be unable to examine the transfer functions of our ROIs independent of the influence of top down processing. The second implication of this limitation is that our experiment will be unable to assess the effects that task demands may have on the process of pattern separation. Motley and Kirwan (2012) demonstrated that the laterality of pattern separation depended on the type of task being performed. A possible next step for this project would be to examine how the transfer function of each of our ROIs is modulated by task demands by incorporating additional memory tasks that emphasize different forms of memory as well as an incidental memory task to serve as a baseline.

4.1 Future Direction

The current study aimed to investigate the relationship between pattern separation and item similarity, hoping to observe a power function predicted by computational models. We were unsuccessful in observing a significant relationship between the image similarity and MTL activation. In order to rule out task-related issues, it would be best to replicate the current research while accounting for these potential issues. Our main concern for the cause of the null result is that it is due to our range of similarity being too narrow. To address this, a replication of the current research that uses an item set that contains a larger range of similarity between image pairs may be able to induce a large enough change in similarity to observe a change in representational output. If our replication study is successful in observing a relationship between similarity and pattern separation, a next step would be to investigate how this relationship is affected by aging. Studies investigating behavioural discrimination have found that older adults require a larger change in similarity before they can correctly identify a lure item as new/similar (Stark, Yassa, Lacy, & Stark, 2013). Perhaps this modulation can also be observed at the neurological level.

4.2 Conclusion

The current research attempted to further our understanding of how the human medial temporal lobe encodes new memories. Specifically, we investigated the prediction that the degree to which
two memories undergo the process of pattern separation varies with how similar the two memories are. To that end, we leveraged the repetition suppression effect, using it as a proxy measure of pattern separation. The current research replicated a repetition suppression effect, but activity did not vary with a parametric manipulation of similarity. Serendipitously, we found that the degree of repetition suppression was modulated by a self-reported rating of confidence of judgement, though in the current study we were unable to investigate the implications of this effect. Our findings do not support current theories put forth by the pattern separation literature. Going forward, we aim to attempt to replicate findings observed by other studies that parametrically manipulated similarity.
References


Table 1. Summary of linear mixed model results analyzing the effect of trial type on activity in each ROI. “*” indicates which effects were significant.

<table>
<thead>
<tr>
<th>ROI</th>
<th>Hemisphere</th>
<th>Estimate</th>
<th>SE</th>
<th>df</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior_Hippocampus</td>
<td>Left</td>
<td>0.085</td>
<td>0.031</td>
<td>1</td>
<td>6.188</td>
<td>0.013*</td>
</tr>
<tr>
<td>Anterior_Hippocampus</td>
<td>Right</td>
<td>0.077</td>
<td>0.034</td>
<td>1</td>
<td>4.694</td>
<td>0.030*</td>
</tr>
<tr>
<td>CA1</td>
<td>Left</td>
<td>0.054</td>
<td>0.031</td>
<td>1</td>
<td>2.869</td>
<td>0.090</td>
</tr>
<tr>
<td>CA1</td>
<td>Right</td>
<td>0.057</td>
<td>0.028</td>
<td>1</td>
<td>4.013</td>
<td>0.045*</td>
</tr>
<tr>
<td>CA3_DG</td>
<td>Left</td>
<td>0.110</td>
<td>0.034</td>
<td>1</td>
<td>9.111</td>
<td>0.003*</td>
</tr>
<tr>
<td>CA3_DG</td>
<td>Right</td>
<td>0.107</td>
<td>0.008</td>
<td>1</td>
<td>6.239</td>
<td>0.013*</td>
</tr>
<tr>
<td>ERC</td>
<td>Left</td>
<td>0.027</td>
<td>0.035</td>
<td>1</td>
<td>0.606</td>
<td>0.436</td>
</tr>
<tr>
<td>ERC</td>
<td>Right</td>
<td>0.067</td>
<td>0.030</td>
<td>1</td>
<td>4.933</td>
<td>0.026*</td>
</tr>
<tr>
<td>PHC</td>
<td>Left</td>
<td>0.058</td>
<td>0.031</td>
<td>1</td>
<td>3.489</td>
<td>0.062</td>
</tr>
<tr>
<td>PHC</td>
<td>Right</td>
<td>0.084</td>
<td>0.031</td>
<td>1</td>
<td>7.384</td>
<td>0.007*</td>
</tr>
<tr>
<td>Posterior_Hippocampus</td>
<td>Left</td>
<td>0.007</td>
<td>0.037</td>
<td>1</td>
<td>0.036</td>
<td>0.849</td>
</tr>
<tr>
<td>Posterior_Hippocampus</td>
<td>Right</td>
<td>0.055</td>
<td>0.030</td>
<td>1</td>
<td>3.187</td>
<td>0.074</td>
</tr>
<tr>
<td>PRC</td>
<td>Left</td>
<td>0.073</td>
<td>0.040</td>
<td>1</td>
<td>5.187</td>
<td>0.016*</td>
</tr>
<tr>
<td>PRC</td>
<td>Right</td>
<td>0.050</td>
<td>0.032</td>
<td>1</td>
<td>2.355</td>
<td>0.125</td>
</tr>
<tr>
<td>Sub</td>
<td>Left</td>
<td>0.036</td>
<td>0.029</td>
<td>1</td>
<td>1.476</td>
<td>0.225</td>
</tr>
<tr>
<td>Sub</td>
<td>Right</td>
<td>0.039</td>
<td>0.032</td>
<td>1</td>
<td>1.427</td>
<td>0.232</td>
</tr>
</tbody>
</table>
Table 2. Summary of linear mixed model results analyzing the effect of similarity on activity in each ROI. “*” indicates which effects were significant.

<table>
<thead>
<tr>
<th>ROI</th>
<th>Hemisphere</th>
<th>Estimate</th>
<th>SE</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior_Hippocampus</td>
<td>Left</td>
<td>-0.0307</td>
<td>0.02359</td>
<td>1</td>
<td>1.661</td>
<td>0.198</td>
</tr>
<tr>
<td>Anterior_Hippocampus</td>
<td>Right</td>
<td>-0.0044</td>
<td>0.02115</td>
<td>1</td>
<td>0.0432</td>
<td>0.8354</td>
</tr>
<tr>
<td>CA1</td>
<td>Left</td>
<td>-0.0102</td>
<td>0.01926</td>
<td>1</td>
<td>0.2794</td>
<td>0.5971</td>
</tr>
<tr>
<td>CA1</td>
<td>Right</td>
<td>-0.0105</td>
<td>0.0233</td>
<td>1</td>
<td>0.2029</td>
<td>0.6524</td>
</tr>
<tr>
<td>CA3_DG</td>
<td>Left</td>
<td>-0.0257</td>
<td>0.02185</td>
<td>1</td>
<td>1.3804</td>
<td>0.24</td>
</tr>
<tr>
<td>CA3_DG</td>
<td>Right</td>
<td>-0.0082</td>
<td>0.02238</td>
<td>1</td>
<td>0.1328</td>
<td>0.7156</td>
</tr>
<tr>
<td>ERC</td>
<td>Left</td>
<td>0.00761</td>
<td>0.02187</td>
<td>1</td>
<td>0.1211</td>
<td>0.7278</td>
</tr>
<tr>
<td>ERC</td>
<td>Right</td>
<td>0.01181</td>
<td>0.02142</td>
<td>1</td>
<td>0.3035</td>
<td>0.5817</td>
</tr>
<tr>
<td>PHC</td>
<td>Left</td>
<td>0.03144</td>
<td>0.02235</td>
<td>1</td>
<td>1.9672</td>
<td>0.1607</td>
</tr>
<tr>
<td>PHC</td>
<td>Right</td>
<td>0.0222</td>
<td>0.02411</td>
<td>1</td>
<td>0.8367</td>
<td>0.3604</td>
</tr>
<tr>
<td>Posterior_Hippocampus</td>
<td>Left</td>
<td>-0.0029</td>
<td>0.01848</td>
<td>1</td>
<td>0.0246</td>
<td>0.8755</td>
</tr>
<tr>
<td>Posterior_Hippocampus</td>
<td>Right</td>
<td>-0.0122</td>
<td>0.02096</td>
<td>1</td>
<td>0.3385</td>
<td>0.5607</td>
</tr>
<tr>
<td>PRC</td>
<td>Left</td>
<td>-0.0131</td>
<td>0.02708</td>
<td>1</td>
<td>0.2333</td>
<td>0.6291</td>
</tr>
<tr>
<td>PRC</td>
<td>Right</td>
<td>-0.0119</td>
<td>0.0251</td>
<td>1</td>
<td>0.2244</td>
<td>0.6357</td>
</tr>
<tr>
<td>Sub</td>
<td>Left</td>
<td>-0.0233</td>
<td>0.01967</td>
<td>1</td>
<td>1.4039</td>
<td>0.2361</td>
</tr>
<tr>
<td>Sub</td>
<td>Right</td>
<td>0.00545</td>
<td>0.02269</td>
<td>1</td>
<td>0.0574</td>
<td>0.8106</td>
</tr>
</tbody>
</table>
Table 3. Summary of linear mixed model results analyzing the effect of confidence rating on activity during lure trials in each ROI. “*” indicates which effects were significant.

<table>
<thead>
<tr>
<th>ROI</th>
<th>Hemisphere</th>
<th>Estimate</th>
<th>SE</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior_Hippocampus</td>
<td>Left</td>
<td>0.1319</td>
<td>0.0321</td>
<td>1</td>
<td>13.969</td>
<td>0.000*</td>
</tr>
<tr>
<td>Anterior_Hippocampus</td>
<td>Right</td>
<td>0.1519</td>
<td>0.0355</td>
<td>1</td>
<td>13.915</td>
<td>0.000*</td>
</tr>
<tr>
<td>CA1</td>
<td>Left</td>
<td>0.0696</td>
<td>0.0338</td>
<td>1</td>
<td>3.879</td>
<td>0.049*</td>
</tr>
<tr>
<td>CA1</td>
<td>Right</td>
<td>0.0970</td>
<td>0.0321</td>
<td>1</td>
<td>7.545</td>
<td>0.006*</td>
</tr>
<tr>
<td>CA3_DG</td>
<td>Left</td>
<td>0.1634</td>
<td>0.0332</td>
<td>1</td>
<td>15.423</td>
<td>0.000*</td>
</tr>
<tr>
<td>CA3_DG</td>
<td>Right</td>
<td>0.1663</td>
<td>0.0340</td>
<td>1</td>
<td>17.261</td>
<td>0.000*</td>
</tr>
<tr>
<td>ERC</td>
<td>Left</td>
<td>0.0768</td>
<td>0.0350</td>
<td>1</td>
<td>4.659</td>
<td>0.031*</td>
</tr>
<tr>
<td>ERC</td>
<td>Right</td>
<td>0.0697</td>
<td>0.0336</td>
<td>1</td>
<td>4.282</td>
<td>0.039*</td>
</tr>
<tr>
<td>PHC</td>
<td>Left</td>
<td>0.0421</td>
<td>0.0390</td>
<td>1</td>
<td>1.135</td>
<td>0.287</td>
</tr>
<tr>
<td>PHC</td>
<td>Right</td>
<td>0.0269</td>
<td>0.0371</td>
<td>1</td>
<td>0.525</td>
<td>0.469</td>
</tr>
<tr>
<td>Posterior_Hippocampus</td>
<td>Left</td>
<td>0.0232</td>
<td>0.0282</td>
<td>1</td>
<td>0.675</td>
<td>0.411</td>
</tr>
<tr>
<td>Posterior_Hippocampus</td>
<td>Right</td>
<td>0.0534</td>
<td>0.0307</td>
<td>1</td>
<td>2.992</td>
<td>0.084</td>
</tr>
<tr>
<td>PRC</td>
<td>Left</td>
<td>0.0634</td>
<td>0.0333</td>
<td>1</td>
<td>3.528</td>
<td>0.060</td>
</tr>
<tr>
<td>PRC</td>
<td>Right</td>
<td>0.0758</td>
<td>0.0341</td>
<td>1</td>
<td>4.243</td>
<td>0.039*</td>
</tr>
<tr>
<td>Sub</td>
<td>Left</td>
<td>0.0315</td>
<td>0.0317</td>
<td>1</td>
<td>0.956</td>
<td>0.328</td>
</tr>
<tr>
<td>Sub</td>
<td>Right</td>
<td>0.0525</td>
<td>0.0336</td>
<td>1</td>
<td>2.342</td>
<td>0.126</td>
</tr>
</tbody>
</table>
Figures
Figure 1. Depicting the nonlinear transfer function of DG and CA3. In the DG, pattern separation is defined by a power curve with increasing slope, whereby a small change in input produces a large change in neural output. The CA3 subregion initially supports pattern completion, defined by a power curve with increasing slope. This function adapts to a power curve with decreasing slope, consistent with pattern separation, once there is a sufficient change in input. Adapted from “Pattern separation in the hippocampus,” by M. A. Yassa and C. E. Stark (2011), Trends in neurosciences, 34(10), 515-525.
**Figure 2.** A depiction of a trial sequence from an encoding run (top) and a retrieval run (bottom) of the fMRI task. See Methods for an in-depth description of the task.

**Encoding**

1. +
2. Parking Lot
3. +
4. Anchor Tattoo
5. +
6. Chess Board

**Retrieval**

1. Anchor Tattoo
2. How Vivid? 1 – 8
3. 1: Old; 2: New
4. How Confident? 1 – 4
Figure 3. A depiction of a trial sequence from the subjective and behavioural similarity rating task. See Methods for an in-depth description of the task.
Figure 4. Behavioural performance during the fMRI task. Percentage correct was calculated as the proportion of trials for which participants correctly identified the object.
Figure 5. Graphic depiction of the effect of Similarity on activity collapsed across hemispheres. For graphing purposes, we grouped image pairs into bins based on their quantile ranking on the scale of our combined similarity metric. Bins increase in similarity with Bin 6 represents repeat trials (“Maximum similarity”). Error bars represent the 95% CI using the average Standard Error of each ROI.
Figure 6. Graphic depiction of the effect of confidence ratings on activity during lure trials, collapsed across hemispheres. Error bars represent the 95% CI using the average Standard Error of each ROI.
Figure 7. Graphic depiction of the effect of confidence ratings on activity during repeat trials, collapsed across hemispheres. Error bars represent the 95% CI using the average Standard Error of each ROI.
Figure 8. Graphic depiction of the effect of “Perceived Oldness” collapsed across hemispheres. The “Perceived Oldness” scale is formed by reverse coding confidence ratings on lure trials, and appending it with confidence ratings on repeat trials (1 = Confident new, 8 = Confident Old). Ratings of 4 and 5 on the “Perceived Oldness” scale signify the transition of confidence ratings from lure trials to repeat trials. Error bars represent the 95% CI using the average Standard Error of each ROI.