1

2

Confronting the elephant in the room – verbal paired associates and the hippocampus

5

6

7 **Ian A. Clark, Misun Kim, Eleanor A. Maguire**

8

9 Wellcome Centre for Human Neuroimaging, Institute of Neurology,

10 University College London, London, WC1N 3BG, UK

11

12

13 Abbreviated title: verbal paired associates and the hippocampus

14

15

16

17 Corresponding author: Eleanor A. Maguire, Wellcome Centre for Human Neuroimaging,
18 Institute of Neurology, University College London, 12 Queen Square, London, WC1N 3BG,
19 UK; e.maguire@ucl.ac.uk

20

21

22

23

24

25

30

31 The authors declare no competing financial interests and no conflicts of interest.

32

Abstract

It is widely agreed that patients with bilateral hippocampal damage are impaired at binding pairs of words together. Consequently, the verbal paired associates (VPA) task has become emblematic of hippocampal function. This VPA deficit is not well understood, and is particularly difficult for hippocampal theories with a visuospatial bias to explain (e.g., cognitive map and scene construction theories). Notably, however, VPA tasks typically use concrete imageable words and so conflate imagery and binding. To determine why VPA engages the hippocampus, we devised an fMRI encoding task involving pairs of scene words, pairs of object words and pairs of non-imageable abstract words which were closely matched on a wide range of features. Using univariate and multivariate methods we found that the anterior hippocampus in particular was engaged during the encoding of both scene and object word pairs. This was not the case for abstract word pairs, despite binding occurring here also, and even when only subsequently-remembered trials were considered. Moreover, for object word pairs, fMRI activity patterns in the anterior hippocampus were more similar to those for scene imagery than object imagery. This was especially evident in participants who were high imagery users, and not in mid and low imagery users. Overall, our results show that hippocampal engagement during VPA, even when object word pairs are involved, seems to be evoked by scene imagery rather than binding per se. This may help to resolve the issue that visuospatial hippocampal theories have in accounting for verbal memory.

Significance Statement

Binding pairs of words together is a task closely associated with the hippocampus. Therefore, explaining exactly how it achieves this could be important for leveraging a fuller understanding of hippocampal function. To date, word pair tasks have confounded the binding process with the use of words that are easy to imagine. We devised a task that separated binding from imagery. Combining this with functional MRI brain scanning we found that engagement of the anterior hippocampus was associated with imageability, and in particular the use of scene imagery, rather than binding. Moreover, this was especially true for participants who were high imagery users. We conclude that even apparently verbal tasks may be processed by the hippocampus by means of scene imagery.

1 Introduction

2   The field of hippocampal neuroscience is characterized by vigorous debates. But one point on

3   which there is wide agreement is that people with bilateral hippocampal damage and

4   concomitant amnesia (hippocampal amnesia) are significantly impaired on verbal paired

5   associates (VPA) tasks. For example, in tests like the widely-used Wechsler Memory Scale

6   (WMS-IV; Wechsler, 2009) the requirement is to encode pairs of words (e.g., bag–truck),

7   memory for which is then tested. The reliable deficit observed in hippocampal amnesia means

8   the VPA has become emblematic of hippocampal function.

9

10 One theory explains the VPA findings by positing that the hippocampus binds arbitrary

11 relations among individual elements (Cohen and Eichenbaum, 1993). However, this

12 associative perspective is at odds with other accounts. The cognitive map theory, for instance,

13 suggests that the hippocampus specifically supports flexible, allocentric representations of

14 spatial relationships (O'Keefe and Nadel, 1978). While the scene construction theory (see also

15 the emergent memory account; Graham et al., 2010) proposes that the anterior hippocampus

16 constructs models of the world in the form of spatially coherent scenes (Hassabis and Maguire,

17 2007; Maguire and Mullally, 2013; Zeidman and Maguire, 2016). These latter two theories do

18 not explain why learning of VPA is invariably compromised following hippocampal damage.

19 Indeed, for decades, this has been the elephant in the room of hippocampal theories with a

20 visuospatial bias.

21

22 Resolving the tension among hippocampal theories concerning the VPA could be important for

23 leveraging a fuller understanding of hippocampal function. In taking this issue forwards, it is

24 worthwhile first to step back. Examination of the words used in typical VPA tests shows the

25 vast majority are highly imageable. As it stands, therefore, when using VPA tests, associative

processes and imageability are conflated. One way to deal with this is to examine non-imageable abstract word pairs, which would assess binding without imageability, but these rarely feature in VPA tests used with patients or in neuroimaging. In addition, different types of imageable words are not distinguished in VPA tests. However, the scene construction theory links the anterior hippocampus specifically with scene imagery (Zeidman and Maguire, 2016; Dalton and Maguire, 2017), while the processing of single objects is usually associated with perirhinal and lateral occipital cortices (Malach et al., 1995; Murray et al., 2007). It could therefore be that a scene word in a pair engages the hippocampus and not binding per se. It has also been suggested that even where each word in a pair denotes an object, this might elicit imagery of both objects together in a scene and this is what recruits the hippocampus (Maguire and Mullally, 2013; Clark and Maguire, 2016).

To determine why VPA engages the hippocampus, we devised an fMRI encoding task with three types of word pairs: where both words in a pair denoted Scenes, where both words represented single Objects, and where both words were non-imageable Abstract words. This allowed us to separate imageability from binding, and to examine different types of imagery. Memory tests after scanning meant that we could also consider the effect of encoding success. Given that people vary in their use of mental imagery (Marks, 1973; Kosslyn et al., 1984; McAvinue and Robertson, 2007), we also tested groups of high, mid and low imagery users to assess whether this affected hippocampal engagement during VPA encoding.

We hypothesised that anterior hippocampal activity elicited during word pair encoding would be apparent for Scene words pairs and Object word pairs compared to Abstract word pairs. This would be best explained by the use of scene imagery, even for Object word pairs, and the effect would be most apparent in high imagery users. Furthermore, we predicted that neither

1 associative processing nor memory performance would account for the patterns of hippocampal

2 activity observed. As such, our main anatomical focus was the hippocampus, and of particular

3 interest were the Object word pairs and their relationship with scene imagery.

4
5
6 **Materials and Methods**

7 *Participants*

8 Forty five participants took part in the fMRI study. All were healthy, right-handed, proficient

9 in English and had normal or corrected to normal vision. Each participant gave written

10 informed consent. The study was approved by the University College London Research Ethics

11 Committee. Participants were recruited on the basis of their scores on the Vividness of Visual

12 Imagery Questionnaire (VVIQ; Marks, 1973). The VVIQ is a widely-used self-report

13 questionnaire which asks participants to bring images to mind and rate them on a 5 point scale

14 as to their vividness (anchored at 1: "perfectly clear and as vivid as normal vision", and 5: "No

15 image at all, you only 'know' that you are thinking of the object"). Therefore, a high score on

16 the VVIQ corresponds to low use of visual imagery. We required three groups for our fMRI

17 study (n=15 in each), low imagery users, mid imagery users and high imagery users. Initially,

18 184 people completed the VVIQ. Fifteen of the highest and 15 of the lowest scorers made up

19 the low and high imagery groups. A further 15 mid scorers served as the mid imagery group.

20 The groups did not differ significantly on age, gender, years of education and general intellect.

21 Table 1 provides details of the three groups.

22
23
24
25
26
27
28
29
30

Table 1. Characteristics of the participant groups.

	Imagery Group			p-value		
	Low	**Mid**	**High**	**Low vs Mid**	**Low vs High**	**Mid vs High**
Age	23.07 (2.31)	21.87 (2.20)	23.93 (5.26)	0.16	0.57	0.18
Number of males	6 (40.0%)	7 (46.67%)	8 (53.33%)	0.71	0.46	0.72
Years of education	16.0 (1.89)	15.8 (1.61)	16.0 (2.33)	0.76	1.0	0.79
Matrix Reasoning	12.47 (2.26)	11.47 (2.17)	12.07 (3.61)	0.23	0.72	0.59
TOPF	54.93 (5.13)	57.47 (5.49)	53.0 (9.47)	0.20	0.49	0.13
FSIQ	110.13 (5.48)	111.97 (6.13)	110.22 (5.99)	0.39	0.97	0.44
VCI	108.93 (5.32)	110.81 (6.18)	108.75 (6.0)	0.38	0.93	0.36
VVIQ mean score	3.08 (0.45)	2.15 (0.17)	1.51 (0.25)	< 0.001	< 0.001	< 0.001

Means (standard deviations). Two-tailed p-values for t-tests (χ^2 test for the number of males). General intellect was measured using the Matrix Reasoning subtest (scaled scores) of the Wechsler Adult Intelligence Scale-IV (WAIS-IV; Wechsler, 2008) and the Test of Premorbid Function (TOPF; Wechsler, 2011) providing an estimate of Full Scale IQ (FSIQ) and a Verbal Comprehension Index (VCI). VVIQ=Vividness of Visual Imagery Questionnaire.

Stimuli

To ensure that any fMRI differences were due to our imagery manipulation and not other word properties, the word conditions were highly matched. Six hundred and fifty four words were required for the study – 218 Scene words, 218 Object words and 218 Abstract words. Words were initially sourced from databases created by Brysbaert and colleagues, which provided ratings for concreteness, word frequency, age of acquisition, valence and arousal (Kuperman et al., 2012; Warriner et al., 2013; Brysbaert et al., 2014; van Heuven et al., 2014). It was important to control for valence and arousal given reports of higher emotional ratings for abstract words, which could influence fMRI activity (Kousta et al., 2011; Vigliocco et al., 2014). We also used data from the English Lexicon project (Balota et al., 2007) to provide lexical information about each word – word length, number of phonemes, number of syllables, number of orthographic neighbours and number of phonological and phonographic neighbours with and without homophones.

To verify that each word induced the expected imagery (i.e., scene imagery, object imagery or very little/no imagery for the abstract words), we collected two further ratings for each word.

First, a rating of imageability to ensure that Scene and Object words were not only concrete but also highly imageable (although concreteness and imageability are often interchanged, while they are highly related constructs, they are not the same; Paivio et al., 1968), and additionally that Abstract words were low on imageability. Second, a decision was elicited about the type of imagery the word brought to mind, i.e., was the imagery of a scene or an isolated object. These ratings were collected from 119 participants in total using Amazon Mechanical Turk's crowdsourcing website, following the procedures employed by Brysbaert and colleagues for the databases described above. Words were classified as a Scene or Object word when there was a minimum of 70% agreement on the type of imagery brought to mind, and the mean imageability rating was greater than 3.5 (out of 5). For Abstract words, the mean imageability had to be less than or equal to 2.

An overview of the word properties is shown in Table 2. This also includes summary comparison statistics. Scene, Object and Abstract words were matched on 13 out of the 16 measures. Scene and Object words were matched on all 16 measures, whereas Abstract words, as expected, were less concrete and less imageable than Scene and Object words and had a higher age of acquisition, as is normal for abstract words (Stadthagen-Gonzalez and Davis, 2006; Kuperman et al., 2012). As well as being matched at the overall word type level as shown on Table 2, within each word type, words were assigned to one of four lists (word pairs, single words, catch trials or post-scan memory test lures), and all lists were matched on all measures.

Table 2. Properties of each word type.

Word property	Type of word			p-value		
Lexical criteria	Scene	Object	Abstract	Scene vs Object	Scene vs Abstract	Object vs Abstract
Number of letters[a]	6.73 (1.99)	6.64 (1.91)	6.75 (1.97)	0.64	0.90	0.55
Number of phonemes[a]	5.61 (1.89)	5.41 (1.71)	5.66 (1.72)	0.25	0.77	0.13
Number of syllables[a]	2.09 (0.85)	2.06 (0.79)	2.16 (0.77)	0.68	0.38	0.18
Number of orthographic neighbours[a]	2.57 (4.97)	3.09 (4.96)	2.35 (4.33)	0.28	0.62	0.10
Number of phonological neighbours[a]	6.28 (11.71)	7.61 (11.80)	6.12 (11.31)	0.23	0.89	0.18
Number of phonological neighbours (inc homophones)[a]	6.88 (12.59)	8.13 (12.48)	6.66 (11.86)	0.30	0.85	0.21
Number of phonographic neighbours[a]	1.53 (3.51)	1.78 (3.48)	1.46 (3.21)	0.45	0.84	0.32
Number of phonographic neighbours (inc homophones)[a]	1.61 (3.67)	1.96 (3.63)	1.49 (3.36)	0.32	0.72	0.16
Word frequency: Zipf[b]	3.90 (0.71)	3.80 (0.61)	3.88 (0.82)	0.12	0.77	0.27
Age of acquisition[c]	7.69 (2.14)	7.40 (2.12)	9.78 (2.46)	0.15	< 0.001	< 0.001
Emotional constructs						
Valence[d]	5.68 (1.08)	5.63 (1.02)	5.58 (1.12)	0.63	0.34	0.61
Number of positive words[d,e]	171 (78.44%)	173 (79.6%)	167 (76.61%)	0.81	0.65	0.49
Hedonic valence[d,f]	1.07 (0.69)	0.98 (0.68)	1.04 (0.70)	0.18	0.69	0.34
Arousal[d]	4.07 (0.96)	3.99 (0.87)	4.04 (0.71)	0.34	0.73	0.46
Imagery						
Concreteness[g]	4.65 (0.22)	4.68 (0.22)	1.83 (0.29)	0.11	< 0.001	< 0.001
Imageability[h]	4.38 (0.29)	4.41 (0.32)	1.53 (0.20)	0.31	< 0.001	< 0.001

Means (standard deviations). Two-tailed p-values for t-tests (χ^2 test for the number of positive words). Note that each comparison was assessed separately in order to provide a greater opportunity for any differences between conditions to be identified.

[a]From the English Leixon Project (Balota et al., 2007: exlexicon.wustl.edu).

[b]From van Heuven et al. (2014). The Zipf scale is a standardised measure of word frequency using a logarithmic scale. Values go from 1 (low frequency words) to 6 (high frequency words).

[c]From Kuperman et al. (2012).

[d]From Warriner et al. (2013).

[e]Positive words were those that had a valence score greater than or equal to 5.

[f]Hedonic valence is the distance from neutrality (i.e., from 5), regardless of being positive or negative, as per Vigliocco et al. (2014).

[g]From Brysbaert et al. (2014).

[h]Collected for the current study as detailed in the Materials and Methods.

1 *Experimental design and task*

2 The fMRI task consisted of two elements, the encoding task and catch trials. The latter were

3 included to provide an active response element and to encourage concentration during the

4 experiment. To match the WMS-IV Verbal Paired Associate Test (Wechsler, 2009), each

5 stimulus was presented for 4 seconds. This was followed by a jittered baseline (a central

6 fixation cross) for between 2 and 5 seconds which aided concentration by reducing the

7 predictability of stimulus presentation (Figure 1D). The scanning session was split into four

8 runs of approximately equal length. Trials were presented randomly for each participant with

9 no restrictions on what could precede or follow each trial.

10

11 Unbeknownst to participants, there were six categories of stimuli – imageable Scene words,

12 imageable Object words and non-imageable Abstract words, shown either in pairs of the same

13 word type (Figure 1A) or as single words (Figure 1B). To equalise visual presentation between

14 the word pairs and the single words, the latter were presented with a random letter string that

15 did not follow the rules of the English language and did not resemble real words (Figure 1B).

16 The average, minimum and maximum length of the letter strings was matched to the real words.

17 Letter strings could either be presented at the top or the bottom of the screen. There were 45

18 trials of each condition, with each word shown only once to the participant. Our prime interest

19 was in the word pair conditions, and in particular the Object word pairs, as these related directly

20 to our research question. The single word conditions were included for the purposes of specific

21 analyses, which are detailed in the Results section.

22

23 Participants were asked to try and commit the words to memory for later memory tests, and

24 were specifically instructed that they would be asked to recall the word pairs as pairs. No

25 further instructions about how to memorise the stimuli were given (i.e., we did not tell

1 participants to use any particular strategy). Participants were told that occasionally there would

2 be catch trials where they had to indicate using a button press if they saw a real word presented

3 with a 'pseudoword' (Figure 1C). A pseudoword is a combination of letters that resembles a

4 real English word and follows the rules of the English language, but is not an actual real word.

5 Pseudowords were generated using the English Lexicon Project (Balota et al., 2007) and were

6 paired with Scene, Object or Abstract words. They were presented at either the top or the

7 bottom of the screen to ensure that participants attended to both. The number of letters and

8 orthographic neighbours of the pseudowords were matched to all of the real word conditions

9 and across the three pseudoword groups (all p's > 0.3). Additionally, across the pseudoword

10 groups we matched the accuracy of pseudoword identification (all p's > 0.6) as reported in the

11 English Lexicon Project (Balota et al., 2007). Forty eight catch trials were presented over the

12 course of the experiment, 16 trials with each of the word types, ranging between 10 and 15 in

13 each of the four runs. Catch trials were pseudo-randomly presented to ensure regular

14 presentation but not in a predictable manner. Feedback was provided at the end of each run as

15 to the number of correctly identified pseudowords and incorrectly identified real words.

16

17

18
19
20
21
22
23
24
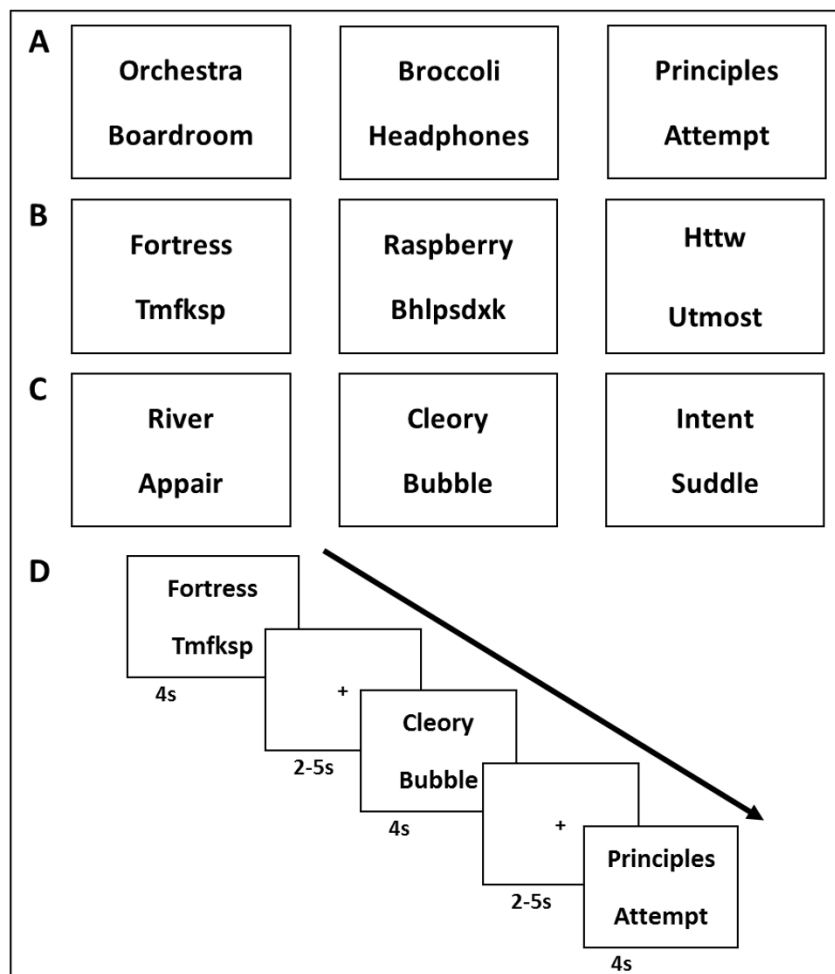
Figure 1. Example stimuli and trial timeline. **(A)** Examples of stimuli from each of the word types in the order of (from left to right) Scene word pair, Object word pair, Abstract word pair. **(B)** Examples of single word trials in the order of (from left to right) Scene single word, Object single word, Abstract single word. Single words were shown with random letter strings (which could be presented at either the top or the bottom) in order to be similar to the visual presentation of the word pairs. **(C).** Examples of catch trials, where a real word was presented with a pseudoword, which could be presented as either the top or bottom word. **(D).** Example timeline of several trials.

Post-scan recognition memory tests

Following scanning, participants had two recognition memory tests. The first was an item

recognition memory test for all 405 words presented during scanning (45 words for each of

three single word types, and 90 words for each of three paired word types) and a further 201

foils (67 of each word type). Each word was presented on its own in the centre of the screen

for up to 5 seconds. Words were presented randomly in a different order for each participant.

Participants had to indicate for each word whether they had seen it in the scanner (old) or not

11

1 (new). Following this, they rated their confidence in their answer on a 3 point scale – high

2 confidence, low confidence or guessing. Any trials where a participant correctly responded

3 "old" and then indicated they were guessing were excluded from subsequent analyses.

4

5 After the item memory test, memory for the pairs of words was examined. This associative

6 memory test presented all of the 135 word pairs shown to participants in the scanner and an

7 additional 66 lure pairs (22 of each type), one pair at a time, for up to 5 seconds. The word

8 pairs were presented in a different random order for each participant. The lure pairs were

9 constructed from the single words that were presented to the participants in the scanner.

10 Therefore, the participants had seen all of the words presented to them in the associative

11 recognition memory test, but not all were previously in pairs, specifically testing whether the

12 participants could remember the correct associations. Participants were asked to indicate

13 whether they saw that exact word pair presented to them in the scanner (old) or not (new). They

14 were explicitly told that some pairs would be constructed from the single words they had seen

15 during scanning and not to make judgements solely on individual words, but to consider the

16 pair itself. Confidence ratings were obtained in the same way as for the item memory test, and

17 trials where a participant correctly responded "old" and then indicated they were guessing were

18 excluded from subsequent analyses.

19

20 *Debriefing*

21 On completion of the memory tests, participants were asked about their strategies for

22 memorising the words while they were in the scanner. At this point, the participants were told

23 about the three different types of words presented to them – Scenes, Objects and Abstract. For

24 each word type, and separately for single words and word pairs, participants were presented

25 with reminders of the words, and were asked to choose from a list of options as to which

1   strategy best reflected how they attempted to memorise that word type. Options included: "I

2   had a visual image of a scene related to this type of single word" (scene imagery), "I had a

3   visual image of a single entity (e.g. one specific object) for a word with no other background

4   imagery" (object imagery), "I read each word without forming any visual imagery at all" (no

5   imagery).

6

7   *Statistical analyses of the behavioural data*

8   *Stimuli creation and participant group comparisons.* Comparisons between word conditions,

9   and between the participant groups, were performed using independent samples t-tests for

10   continuous variables and chi squared tests for categorical variables. An alpha level of $p > 0.05$

11   was used to determine that the stimuli/groups were matched. Note that each comparison was

12   assessed separately (using t-tests or chi squared tests) in order to provide a greater opportunity

13   for any differences between conditions to be identified.

14

15   *Main study.* Both within and between participants designs were used. The majority of analyses

16   followed a within-participants design, with all participants seeing all word conditions.

17   Additionally, participants were split into three groups dependent on their VVIQ score allowing

18   for between-participants analyses to be performed.

19

20   All data were assessed for outliers, defined as values that were at least 2 standard deviations

21   away from the mean. If an outlier was identified then the participant was removed from the

22   analysis in question (and this is explicitly noted in the Results section). Memory performance

23   for each word condition was compared to chance level (50%) using one sample t-tests. For all

24   within-participants analyses, when comparing across three conditions, repeated measures

25   ANOVAs with follow-up paired t-tests were employed, and for comparison across two

13

1 conditions paired t-tests were utilised. For between-participants analyses a one-way ANOVA

2 was performed with follow up independent samples t-tests.

3

4 All ANOVAs were subjected to Greenhouse-Geisser adjustment to the degrees of freedom if

5 Mauchly's sphericity test identified that sphericity had been violated. For all statistical tests

6 alpha was set at 0.05. Effect sizes are reported following significant results as Cohen's d for

7 one sample and independent sample t-tests, Eta squared for repeated measures ANOVA and

8 Cohen's d for repeated measures (d_{rm}) for paired samples t-tests (Lakens, 2013). All analyses

9 were performed in IBM SPSS statistics v22.

10

11 ***Scanning parameters and data pre-processing***

12 T2*-weighted echo planar images (EPI) were acquired using a 3T Siemens Trio scanner

13 (Siemens Healthcare, Erlangen, Germany) with a 32-channel head coil. fMRI data were

14 acquired over four scanning runs using scanning parameters optimised for reducing

15 susceptibility-induced signal loss in the medial temporal lobe: 48 transversal slices angled at

16 -30°, TR=3.36 s, TE=30 ms, resolution=3×3x3mm, matrix size=64x74, z-shim gradient

17 moment of -0.4mT/m ms (Weiskopf et al., 2006). Fieldmaps were acquired with a standard

18 manufacturer's double echo gradient echo field map sequence (short TE=10 ms, long

19 TE=12.46 ms, 64 axial slices with 2 mm thickness and 1 mm gap yielding whole brain

20 coverage; in-plane resolution 3 x 3 mm). After the functional scans, a 3D MDEFT structural

21 scan was obtained with 1mm isotropic resolution (Deichmann et al., 2004).

22

23 Preprocessing of data was performed using SPM12 (www.fil.ion.ucl.ac.uk/spm). Functional

24 images were co-registered to the structural image, and then realigned and unwarped using field

25 maps. The participant's structural image was segmented and spatially normalised to a standard

1   EPI template in MNI space with a voxel size of 3x3x3mm and the normalisation parameters

2   were then applied to the functional data. For the univariate analyses, the functional data were

3   smoothed using an 8mm full-width-half-maximum Gaussian kernel. The multivariate analyses

4   used unsmoothed data.

5

6   The multivariate analysis was performed on a region of interest (ROI) that encompassed the

7   anterior hippocampus bilaterally. This was delineated using an anatomical mask that was

8   defined in the coronal plane and went from the first slice where the hippocampus can be

9   observed in its most anterior extent (see Dalton et al., 2017 for more details) until the final slice

10   of the uncus.

11

12   *fMRI analysis: univariate*

13   The six experimental word conditions were Scene, Object and Abstract words, presented as

14   either word pairs or single words. As noted above, our prime interest was in the word pair

15   conditions, and in particular the Object word pairs, as these related directly to our research

16   question. We therefore directly contrasted fMRI BOLD responses between these conditions.

17   The single word conditions were included for the purposes of specific analyses, which are

18   detailed in the Results section. We performed two types of whole brain analysis, one using all

19   of the trials and another using only trials where the items were subsequently remembered.

20

21   For both analyses, the GLM consisted of the word condition regressors convolved with the

22   haemodynamic response function, in addition to participant-specific movement regressors and

23   physiological   noise   regressors.   The   Artifact   Detection   Toolbox

24   (http://www.nitrc.org/projects/artifact_detect/) was used to identify spikes in global brain

25   activation and these were entered as a separate regressor. Participant-specific parameter

estimates for each regressor of interest were calculated for each voxel. Second level random effects analyses were then performed using one sample t-tests on the parameter estimates. For comparison across VVIQ imagery groups, we performed an ANOVA with follow up independent sample t-tests. We report results at a peak-level threshold of p less than 0.001 whole-brain uncorrected for our a priori region of interest – the hippocampus – and p less than 0.05 family-wise error (FWE) corrected elsewhere.

fMRI analysis: multivariate

Multivoxel pattern analysis was used to test whether the neural representations of the Object word pairs were more similar to the Scene single words than the Object single words. For each participant, T-statistics for each voxel in the anatomically defined anterior hippocampal ROI were computed for each condition (Object word pair, Object single word, Scene single word) and in each scanning run. The Pearson correlation between each condition was then calculated as a similarity measure (Object word pair/Object word pair, Object word pair/Scene single word, Object word pair/Object single word). The similarity measure was cross-validated across the different scanning runs to guarantee the independence of each data set. Repeated measures ANOVA and paired t-tests were used to compare the similarity between conditions at the group level. This multivariate analysis was first applied to the data from all participants, and then to the three subsets of participants (low, mid and high imagery users). All data were assessed for outliers, defined as values that were at least 2 standard deviations away from the group mean. If an outlier was identified then the participant was removed from the analysis in question (and this is explicitly noted in the Results section).

Note that the absolute correlation of the similarity value is expected to be low due to inherent neural variability and the fact that unique set of words were presented for each scanning

session. As such, the important measure is the comparison of the similarity value between the conditions, not the absolute similarity value of a single condition. The range of similarity values that we found are entirely consistent with those in other studies employing a similar representational similarity approach in a variety of learning, memory and navigation tasks in a wide range of brain regions (Staresina et al., 2012; Hsieh et al., 2014; Chadwick et al., 2015; Hsieh and Ranganath, 2015; Milivojevic et al., 2015; Bellmund et al., 2016; Deuker et al., 2016; Schapiro et al., 2016; Schuck et al., 2016; Kim et al., 2017).

Results

Behavioural

On average, participants identified 85.56% (SD=11.52) of the pseudowords during catch trials, showing that they maintained concentration during the fMRI experiment. On the post-scan item memory test, Scene, Object and Abstract words were remembered above chance and there were no differences between the conditions (Table 3). Performance on the associative memory test also showed that Scene, Object and Abstract word pairs were remembered above chance (Table 4). Comparison of memory performance across the word types found differences in performance in line with the literature (Paivio, 1969). Both types of imageable word pairs (Scene and Object) were remembered better than Abstract word pairs (Figure 2; Table 4), while Object word pairs were remembered better than Scene word pairs. Overall these behavioural findings show that, despite the challenging nature of the experiment, with so many stimuli to encode, participants engaged with the task and committed a good deal of information to memory.

Table 3. Performance (% correct) on the post-scan item memory test (non-guessing trials).

	Scene single words	Object single words	Abstract single words
Mean	67.41	66.37	67.61
Standard Deviation	14.93	17.71	16.06
Comparison to chance (50%)			
	t_{44}	P	d
Scene single words	7.82	< 0.001	2.36
Object single words	6.20	< 0.001	1.87
Abstract single words	7.36	< 0.001	2.22
Comparison across the word types			
	$F_{1.76,77.51}$	P	
Main effect	0.28	0.73	

Table 4. Performance (% correct) on the post-scan associative memory test (non-guessing trials).

	Scene word pairs	Object word pairs	Abstract word pairs
Mean	69.98	76.74	60.94
Standard Deviation	13.88	12.97	19.27
Comparison to chance (50%)			
	t_{44}	P	d
Scene word pairs	9.65	< 0.001	2.91
Object word pairs	13.83	< 0.001	4.17
Abstract word pairs	3.81	< 0.001	1.15
Comparison across the word types			
	$F_{1.35,59.48}$	P	η^2
Main effect	24.21	< 0.001	0.36
	t_{44}	P	d_{rm}
Scene vs Object	5.25	< 0.001	0.50
Scene vs Abstract	3.58	0.001	0.52
Object vs Abstract	5.75	< 0.001	0.94
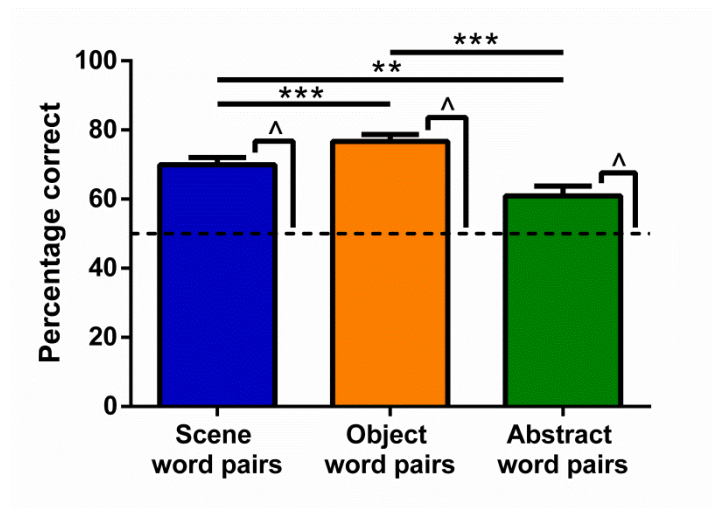
Figure 2. Memory performance on the associative memory test. Error bars are 1 standard error of the mean. ^ indicates a significant difference from chance (dashed line, 50%) at p < 0.001. Stars show the significant differences across the word pair types; **p < 0.01, *** p < 0.001.

fMRI

Univariate analysis. We performed two types of whole brain analysis, one using all of the trials and another using only trials where the items were subsequently remembered in the post-scan memory tests (the item memory test for the single word trials, the associative memory test for the word pairs, excluding trials where participants correctly responded "old" and then indicated they were guessing). The two analyses yielded very similar results across the whole brain. Given that our interest was in the point at which participants were encoding the information and potentially using mental imagery to do so, we present here the results of the analysis using all of the trials. For completeness, we also report the results of the analysis using just the remembered stimuli specifically for our brain region of primary interest, the hippocampus.

We first compared the imageable (Scene and Object) and non-imageable (Abstract) word pairs. All of the conditions involved associative memory, and so we reasoned that any differences we observed, particularly in hippocampal engagement, would be due to the imageability of the Scene and Object pairs. As predicted, Scene word pairs compared to Abstract word pairs

19

elicited greater bilateral anterior hippocampal activity (Figure 3A, full details in Table 5A). Of note, increased activity was also observed in bilateral parahippocampal, bilateral fusiform, retrosplenial and left ventromedial prefrontal cortices (vmPFC). The analysis using only the remembered stimuli showed very similar results, including for the anterior hippocampus (Table 5A). The reverse contrast identified no hippocampal engagement, but rather greater activity in middle temporal cortex (-58, -36, -2, T=6.58), temporal pole (-52, 10, -22, T=6.16) and inferior occipital cortex (40, -68, -14, T=5.59).
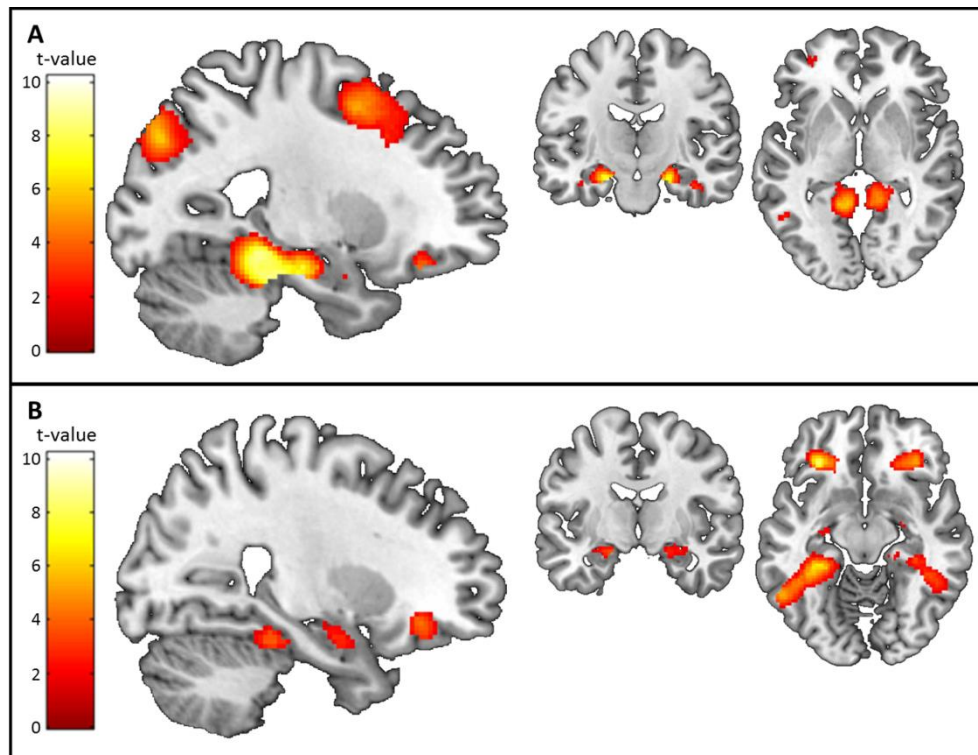
Object word pairs compared with the Abstract word pairs also showed greater bilateral anterior hippocampal activity, along with engagement of bilateral parahippocampal cortex, fusiform cortex and vmPFC, with increased anterior hippocampal activity also apparent when just the subsequently remembered stimuli were considered (Figure 3B, Table 5B). The reverse contrast identified no hippocampal engagement, but rather greater activity in middle temporal cortex (-62, -32, -2, T=8) and temporal pole (-54, 10, -18, T=7.12).

Increased anterior hippocampal activity was therefore observed for both Scene and Object word pairs compared to the non-imageable Abstract word pairs. As greater hippocampal engagement was apparent even when using just the remembered stimuli, it is unlikely that this result can be explained by better associative memory or successful encoding for the imageable word pairs. Rather the results suggest that the anterior hippocampal activity for word pair encoding may be related to the use of visual imagery.

1 **Table 5.** Imageable word pairs compared with Abstract word pairs.
2

A. Scene word pairs > Abstract word pairs

Region	Peak voxel coordinates	t-value
Left anterior hippocampus	-20, -16, -20	8.79
Right anterior hippocampus	20, -10, -20	7.58
Left anterior hippocampus (remembered stimuli only)	*-28, -22, -18*	*7.53*
Right anterior hippocampus (remembered stimuli only)	*24, -20, -18*	*5.09*
Left retrosplenial cortex	-10, -52, 4	9.15
Left fusiform cortex	-22, -34, -20	9.03
Right retrosplenial cortex	10, -48, 6	8.73
Left middle occipital cortex	-30, -74, 34	8.49
Right parahippocampal cortex	24, -34, -20	8.40
Left inferior temporal cortex	-56, -54, -10	8.03
Right fusiform cortex	32, -32, -14	7.72
Left ventrolateral prefrontal cortex	-30, 32, -16	6.51
Right middle occipital cortex	44, -70, 26	6.48
Left middle frontal cortex	-26, 6, 50	6.19
Left inferior frontal cortex	-42, 32, 12	5.74

B. Object word pairs > Abstract word pairs

Left anterior hippocampus	-20, -10, -18	4.45
Right anterior hippocampus	20, -10, -18	3.98
Left anterior hippocampus (remembered stimuli only)	*-32, -22, -12*	*5.05*
Left ventral medial prefrontal cortex	-32, 32, -14	9.45
Left fusiform cortex (extending to parahippocampal cortex)	-32, -34, -20	8.88
Left middle occipital cortex	-34, -80, 28	6.17
Right ventrolateral prefrontal cortex	34, 32, -12	6.05
Left inferior frontal cortex	-40, 28, 14	6.05
Right fusiform gyrus (extending to parahippocampal cortex)	34, -32, -18	5.72

3 P < 0.001 uncorrected for the hippocampus and p < 0.05 FWE for the rest of the brain.
4

5

1

Figure 3. Comparison of imageable Scene or Object word pairs with non-imageable Abstract word pairs. The sagittal slice is of the left hemisphere which is from the ch2better template brain in MRicron (Holmes et al., 1998; Rorden and Brett, 2000). The left of the image is the left side of the brain. The coloured bar indicates the t-value associated with each voxel. **(A)** Scene word pairs > Abstract word pairs. **(B)** Object word pairs > Abstract word pairs. Images are thresholded at p < 0.001 uncorrected for display purposes.

All of the above contrasts involved word pairs, suggesting that associative binding per se cannot explain the results. However, it could still be the case that binding Abstract word pairs does elicit increased hippocampal activity but at a lower level than Scene and Object word pairs. To address this point, we compared the Abstract word pairs with the Abstract single words, as this should reveal any hippocampal activity related to associative processing of the pairs. No hippocampal engagement was evident for the Abstract word pairs (Table 6), and this was also the case when just the remembered stimuli were considered. This lends support to the idea that the use of visual imagery might be important for inducing hippocampal responses to word pairs.

1 **Table 6**. Abstract word pairs compared with Abstract single words.
2

Abstract word pairs > Abstract single words		
Region	Peak voxel coordinates	t-value
Left middle temporal cortex	-64, -36, 2	8.39
Left temporal pole	-52, 12, -16	6.72
Left fusiform cortex	-38, -46, -20	6.64
Left inferior frontal cortex	-54, 24, 12	6.54
Left inferior occipital cortex	-42, -68, -12	6.52
Right inferior occipital cortex	36, -74, -12	6.11
Right lingual cortex	20, -82, -10	5.87
Left pre-central gyrus	-50, 0, 48	5.84

3 $P < 0.001$ uncorrected for the hippocampus (no activations found) and $p < 0.05$ FWE for the rest of the brain.
4

5 We also predicted that anterior hippocampal activity would be specifically influenced by the

6 use of scene imagery, as opposed to visual imagery per se. The inclusion of both Scene and

7 Object word pairs offered the opportunity to test this. Scene word pairs would be expected to

8 consistently evoke scene imagery (as both words in a pair represented scenes), while Object

9 word pairs could evoke both object and scene imagery (e.g., object imagery by imagining the

10 two objects without a background context, or scene imagery by creating a scene and placing

11 the two objects into it), thus potentially diluting the hippocampal scene effect. Scene word pairs

12 might therefore activate the anterior hippocampus to a greater extent that Object word pairs.

13

14 This is indeed what we found, with Scene word pairs evoking greater bilateral anterior

15 hippocampal activity than the Object word pairs (Figure 4, Table 7A). Analysis using the just

16 remembered stimuli gave similar results (Table 7A). Other areas that showed increased activity

17 for the Scene pairs included the retrosplenial and parahippocampal cortices. The reverse

18 contrast examining what was more activated for Object word pairs compared to Scene word

19 pairs found no evidence of hippocampal activity despite better subsequent memory

20 performance for the Object word pairs, and even when the just the remembered stimuli were

21 examined (Table 7B). It seems therefore, that the anterior hippocampus may be particularly

22 responsive to scene imagery.

Table 7. Scene word pairs compared with Object word pairs.

| A. Scene word pairs > Object word pairs		
Region	**Peak voxel coordinates**	**t-value**
Left anterior hippocampus	-22, -18, -20	5.55
Right anterior hippocampus	22, -20, -20	6.07
Left anterior hippocampus (remembered stimuli only)	-22, -20, -18	4.26
Right anterior hippocampus (remembered stimuli only)	24, -20, -20	5.18
Right retrosplenial cortex	16, -54, 20	7.35
Left retrosplenial cortex	-10, -50, 4	7.34
Left fusiform cortex (extending to parahippocampal cortex)	-28, -38, -12	7.25
Right fusiform cortex (extending to parahippocampal cortex)	28, -26, -20	6.87
Left middle temporal cortex	-58, -6, -14	5.77
B. Object word pairs > Scene word pairs		
Left inferior temporal cortex	-42, -48, -16	7.16

P < 0.001 uncorrected for the hippocampus and p < 0.05 FWE for the rest of the brain.
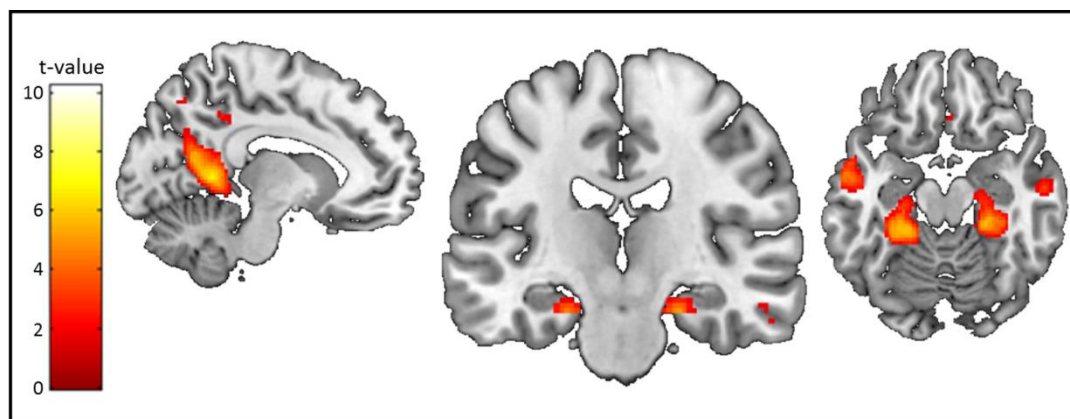
Figure 4. Brain areas more activated by Scene word pairs than Object word pairs. The sagittal slice is of the left hemisphere which is from the ch2better template brain in MRicron (Holmes et al., 1998; Rorden and Brett, 2000). The left of the image is the left side of the brain. The coloured bar indicates the t-value associated with each voxel. Images are thresholded at p < 0.001 uncorrected for display purposes.

1   To summarise, our univariate analyses found that Scene word pairs engaged the anterior

2   hippocampus the most, followed by the Object word pairs, with the Abstract word pairs not

3   eliciting any significant increase in activation (Figure 5). This is what we predicted, and may

4   be suggestive of particular responsivity of the anterior hippocampus to scenes.

5

6

7
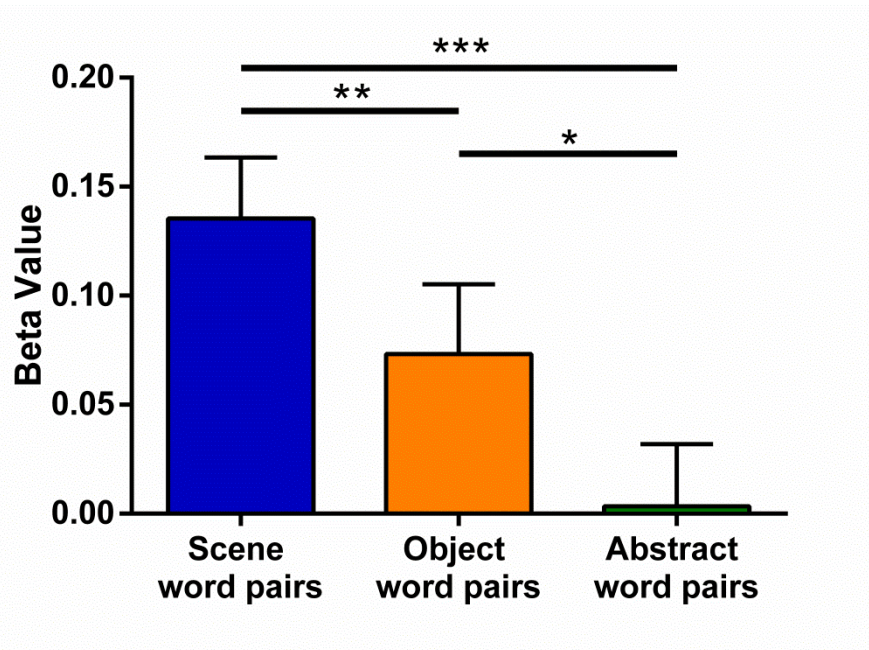
8   **Figure 5. Comparison of each word pair condition with the fixation cross baseline.** Mean beta values
9   extracted from a bilateral anatomical mask of the anterior hippocampus for each of the word pair conditions
10   compared to the central fixation cross baseline. Error bars are 1 standard error of the mean. A repeated measures
11   ANOVA showed significant differences between the conditions ($F_{1.69,74.51}=16.06$, $p <0.001$, $\eta^2=0.27$). Follow-up
12   paired t-tests revealed significant differences between the word pair conditions: Scene word pairs vs Object word
13   pairs $t_{44}=2.97$, $p=0.005$, $d_{rm}=0.30$; Scene word pairs vs Abstract word pairs $t_{44}=6.46$, $p<0.001$, $d_{rm}=0.70$; Object
14   word pairs vs Abstract word pairs $t_{44}= 2.51$, $p=0.016$, $d_{rm}=0.34$.
15   *$p < 0.05$, **$p < 0.01$, ***$p<0.001$.

16

17   *Multivariate analysis*. We next sought further, more direct, evidence that our condition of main

18   interest, Object word pairs, elicited hippocampal activity via scene imagery. Given our

19   univariate findings and the extant literature (e.g. Zeidman and Maguire, 2016), we focused on

20   an anatomically-defined bilateral anterior hippocampal ROI. We then used multivariate

21   representational similarity analysis (RSA; Kriegeskorte et al., 2008) to compare the neural

22   patterns of activity associated with encoding Object word pairs with Scene or Object single

1 words. The single words were chosen as comparators because they consistently elicit either

2 scene or object imagery (see Material and Methods).

3

4 Three similarity correlations were calculated. First, the similarity between Object word pairs

5 and themselves to provide a baseline measure of similarity (i.e., the correlation of Object word

6 pairs over the 4 runs of the scanning experiment). The two similarities of interest were the

7 similarity between Object word pairs and Scene single words, and the similarity between

8 Object word pairs and Object single words. Two participants showed similarity scores greater

9 than 2 standard deviations away from the mean and were removed from further analysis,

10 leaving a sample of 43 participants.

11

12 Repeated measures ANOVA found a significant difference between the three similarities

13 ($F_{2,84}=3.40$, p=0.038, $\eta^2=0.075$). As predicted, the neural representations in the anterior

14 hippocampus of Object word pairs were more similar to Scene single words (Figure 6, purple

15 bar) than to Object single words (Figure 6, light green bar; $t_{42}=2.09$, p=0.042, $d_{rm}=0.21$). In

16 fact, representations of Object word pairs were as similar to Scene single words as to

17 themselves (Figure 6, orange bar; $t_{42}=0.38$, p=0.71). Object word pairs were significantly less

18 similar to Object single words than to themselves ($t_{42}=2.54$, p=0.015, $d_{rm}=0.23$). Of note, these

19 results cannot be explained by subsequent memory performance because Scene single words

20 and Object single words were remembered equally well ($t_{42}=0.68$, p=0.50).

21

22 Overall, these multivariate results show that within the anterior hippocampus, Object word

23 pairs were represented in a similar manner to Scene single words, but not Object single words.

24 This provides further support for our hypothesis that Object word pairs evoke anterior

25 hippocampal activity when scene imagery is involved.
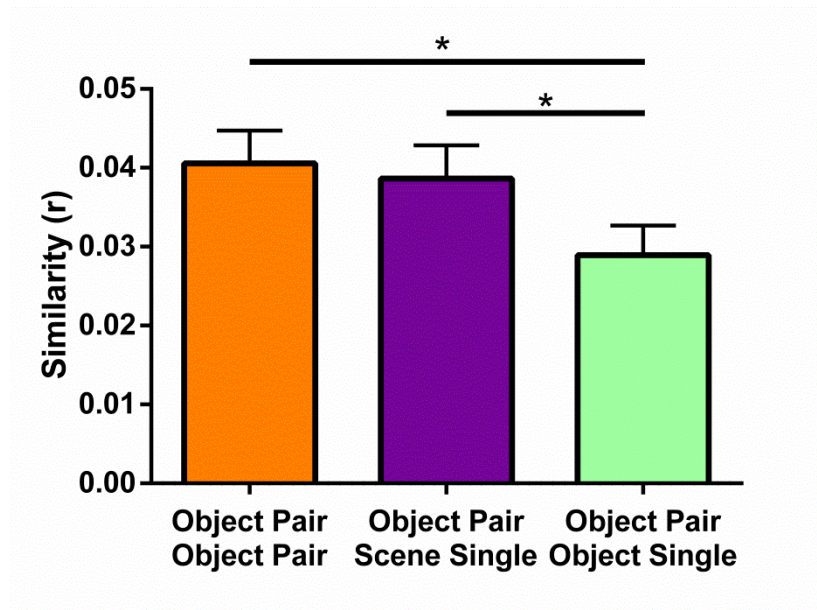
1
2 **Figure 6. The neural similarity of Object word pairs, Scene single words and Object single words.** Object
3 Pair Object Pair – the similarity between Object word pairs between runs. Object Pair Scene Single – the similarity
4 between Object word pairs and Scene single words. Object Pair Object Single – the similarity between Object
5 word pairs and Object single words. Error bars represent 1 standard error of the mean adjusted for repeated
6 measures (Morey, 2008). *p < 0.05.

7
8 *VVIQ and the use of imagery.* As well as examining participants in one large group, as above,

9 we also divided them into three groups based on whether they reported high, mid or low use of

10 imagery on the VVIQ. We found no differences in memory performance among the groups on

11 the word pair tasks (F<0.4 for all contrasts). Similarly, fMRI univariate analyses involving the

12 word pair conditions revealed no differences in hippocampal activity. Voxel based morphology

13 (VBM; Ashburner and Friston, 2000; Mechelli et al., 2005; Ashburner, 2009) showed no

14 structural differences between the groups anywhere in the brain, including in the hippocampus.

15
16 Interestingly however, the imagery groups did differ in one specific way – their strategy for

17 memorising the Object word pairs. While strategy use was similar across the imagery groups

18 for all other word conditions, for the Object word pairs, twice as many participants indicated

19 using a scene imagery strategy in the high imagery group (n=12/15; 80%) than in the mid or

20 low imagery groups (n=5/15; 33% and 6/15; 40% respectively). Comparison of scene strategy

1   use compared to other strategy use across the imagery groups revealed this to be a significant

2   difference (χ^2 (2) = 7.65, p = 0.022).

3

4   Given this clear difference in scene imagery use specifically for the Object word pairs, we

5   performed the anterior hippocampus RSA analysis separately for the three imagery groups. We

6   hypothesised that in the high imagery group, Object word pairs would be represented in the

7   hippocampus in a similar manner to scene single words (as with our whole group analyses)

8   whereas this would not be the case in the mid or low imagery groups. Participants with

9   similarity values greater than 2 standard deviations away from the mean were excluded,

10   resulting in one participant being removed from each group. Importantly, the pattern of scene

11   imagery strategy remained the same even after the removal of these few participants (high

12   imagery group, n=11/14; mid imagery group, n=5/14; low imagery group, n=5/14; χ^2 (2) =

13   6.86, p = 0.032).

14

15   As predicted, for the high imagery group, Object word pairs were more similar to Scene single

16   words than Object single words (Figure 7A; t_{13}=4.63, p<0.001, d=0.78). This was not the case

17   for the mid or low imagery groups (t_{13}=0.472, p=0.65; t_{13}=0.20, p=0.85, respectively). Of note,

18   the interaction between the imagery groups was significant (Figure 7B; $F_{2,39}$=3.53, p=0.039,

19   η^2=0.15). Independent samples t-tests showed that the difference between the similarities was

20   greater in the high imagery group than in the mid and low imagery groups (t_{26}=2.09, p=0.046,

21   d=0.79; t_{26}=2.72, p=0.011, d=1.03, respectively). As before, these differences cannot be

22   explained by subsequent memory performance because all three groups showed no differences

23   between the Scene single and Object single words (high imagery group: t_{13}=0.35, p=0.74; mid

24   imagery group: t_{13}=0.40, p=0.69; low imagery group: t_{13}=1.18, p=0.26).
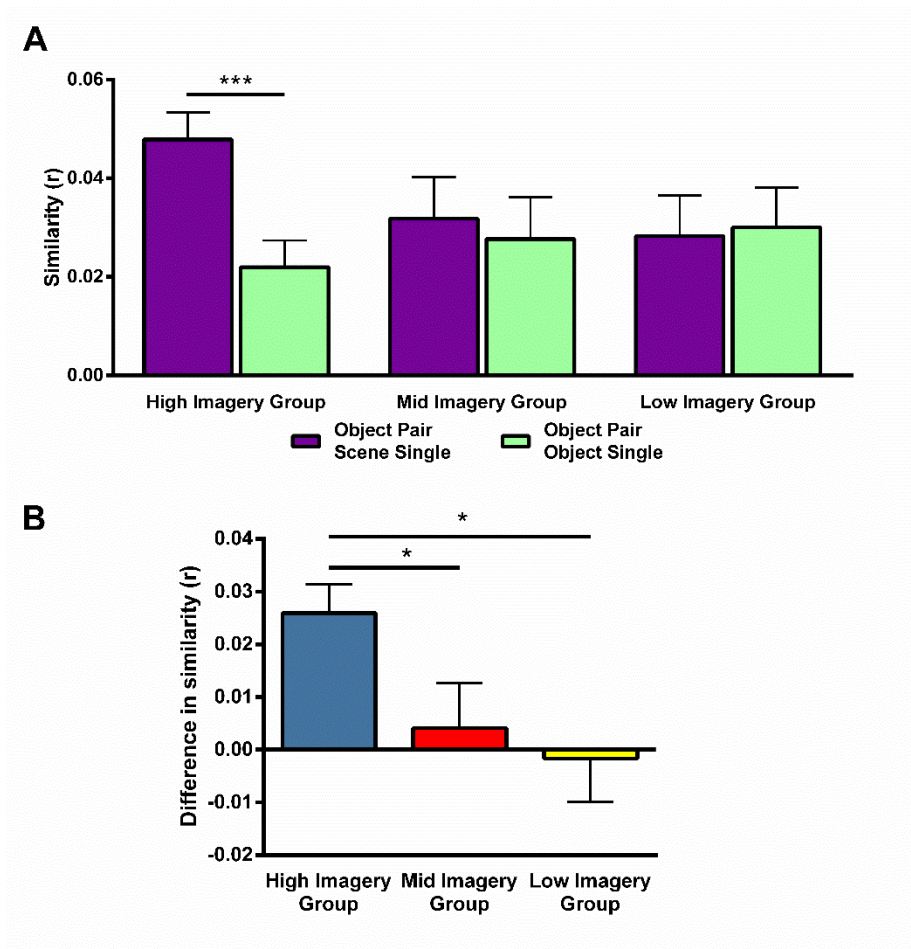
Figure 7. **RSA comparisons of the three imagery groups. (A)** The neural similarity of Object word pairs, Scene single words and Object single words when split by self-reported imagery use. Object Pair Scene Single – the similarity between Object word pairs and Scene single words. Object Pair Object Single – the similarity between Object word pairs and Object single words. **(B)** The difference in similarity between Object word pairs and Scene single words compared to Object words pairs and Object single words in the imagery groups. Error bars represent 1 standard error of the mean. **p < 0.01, ***p<0.001.

In summary, the neural patterns in anterior hippocampus for Object word pairs showed greater similarity with the Scene single words in the high imagery group, whereas for the mid and low imagery groups this was not the case. This provides further evidence linking the anterior hippocampus with the processing of Object word pairs through scene imagery.

1 **Discussion**

2 The aim of this study was to understand the role of the hippocampus in encoding verbal paired

3 associates (VPA). There were five findings. First, we observed greater anterior hippocampal

4 activity for imageable word pairs than non-imageable word pairs, highlighting the influence of

5 visual imagery. Second, non-imageable word pairs compared to non-imageable single words

6 revealed no differences in hippocampal activity, adding further support for the significance of

7 visual imagery. Third, increased anterior hippocampal engagement was apparent for Scene

8 word pairs more than Object word pairs, implicating specifically scene imagery. Fourth, for

9 Object word pairs, fMRI activity patterns in the anterior hippocampus were more similar to

10 those for scene imagery than object imagery, further underlining the propensity of the

11 hippocampus to respond to scene imagery. Finally, our examination of high, mid and low

12 imagery users found that the only difference between them was the use of scene imagery for

13 encoding Object word pairs by high imagers, which in turn was linked to scene-related activity

14 patterns in the hippocampus. Overall, our results provide evidence that anterior hippocampal

15 engagement during VPA seems to be closely related to the use of scene imagery, even for

16 Object word pairs.

17

18 Previous findings have hinted that visual imagery might be relevant in the hippocampal

19 processing of verbal material such as VPA. Work in patients with right temporal lobectomies,

20 which included removal of some hippocampal tissue, suggested that while memory for

21 imageable word pairs was impaired, memory for non-imageable word pairs was preserved

22 (Jones-Gotman and Milner, 1978). Furthermore, instructing these patients to use visual

23 imagery strategies impaired both imageable and non-imageable word pair performance (Jones-

24 Gotman, 1979). We are unaware of any study examining VPA in patients with selective

25 bilateral hippocampal damage that has directly compared imageable and non-imageable word

1   pairs (Clark and Maguire, 2016). More recent fMRI findings also support a possible distinction

2   in hippocampal engagement between imageable and non-imageable word pairs. Caplan and

3   Madan (2016) investigated the role of the hippocampus in boosting memory performance for

4   imageable word pairs, concluding that imageability increased hippocampal activity. However,

5   greater hippocampal activity for high over low imagery word pairs was only observed at a

6   lenient whole brain threshold (p<0.01 uncorrected, cluster size ≥ 5), possibly because their low

7   imagery words (e.g., muck, fright) retained quite a degree of imageability. Furthermore, they

8   did not examine the influence of different types of visual imagery on hippocampal engagement.

9

10   Our different word types were extremely well matched across a wide range of features, with

11   the abstract words being verified as non-imageable, and the scene and object words as reliably

12   eliciting the relevant type of imagery. Using these stimuli we showed that hippocampal

13   involvement in VPA is not linked to visual imagery in general but seems to be specifically

14   related to scene imagery, even when each word in a pair denoted an object. This supports a

15   prediction made by Maguire and Mullally (2013; see also Clark and Maguire, 2016), who noted

16   that a scene allows us to collate a lot of information in a quick, coherent and efficient manner.

17   Consequently, they proposed that people may automatically use scene imagery during

18   encoding and retrieval of imageable verbal material. For instance, we might visualise the scene

19   within which a story is unfolding, or place the objects described in word pairs in a simple scene

20   together.

21

22   If verbal tasks can provoke the use of imagery-based strategies, and if these strategies involve

23   scenes, then patients with hippocampal amnesia would be expected to perform poorly on VPA

24   tasks involving concrete imageable words because they are known to have difficulty with

25   constructing scenes in their imagination (e.g. Hassabis et al., 2007a; Andelman et al., 2010;

1   Race et al., 2011; Mullally et al., 2012; Kurczek et al., 2015). This impairment, which was not

2   apparent for single objects, prompted the proposal of the scene construction theory which holds

3   that scene imagery constructed by the hippocampus is a vital component of memory and other

4   functions (Hassabis and Maguire, 2007; Maguire and Mullally, 2013). Findings over the last

5   decade have since linked scenes to the hippocampus in relation to autobiographical memory

6   (Hassabis et al., 2007b; Hassabis and Maguire, 2007) but also widely across cognition,

7   including perception (Graham et al., 2010; Mullally et al., 2012), future-thinking (Hassabis et

8   al., 2007a; Schacter et al., 2012), spatial navigation (Maguire et al., 2006; Clark and Maguire,

9   2016) and decision-making (Mullally and Maguire, 2014; McCormick et al., 2016).

10

11 Our hippocampal findings were located in the anterior portion of the hippocampus. Anterior

12 and posterior functional differentiation is acknowledged as a feature of the hippocampus,

13 although the exact roles played by each portion are not widely agreed (Moser and Moser, 1998;

14 Fanselow and Dong, 2010; Poppenk et al., 2013; Strange et al., 2014; Ritchey et al., 2015). In

15 particular, our anterior results seem to align with anterior medial hippocampus in the region of

16 the presubiculum and parasubiculum. These areas have been highlighted as being consistently

17 implicated in scene processing (reviewed in Zeidman and Maguire, 2016) and were recently

18 proposed to be neuroanatomically determined to process scenes (Dalton and Maguire, 2017).

19

20 An important point to consider is whether our results can be explained by the effectiveness of

21 encoding, as measured in a subsequent memory test. It is certainly true that people tend to

22 recall fewer abstract than concrete words in behavioural studies of memory (Paivio, 1969;

23 Jones, 1974; Paivio et al., 1994). We tested memory for both single words and paired words.

24 Memory performance for scene, object and abstract words was comparable when tested singly.

25 Memory for the word pairs was significantly lower for the non-imageable Abstract word pairs

compared to the Scene word pairs and Object word pairs. Nevertheless, performance for all conditions was above chance, which was impressive given the large number of stimuli to be encoded with only one exposure. Increased hippocampal activity was apparent for both Scene word pairs and Object words pairs compared to the Abstract word pairs when all stimuli or only the subsequently remembered stimuli were analyzed. This shows that our results cannot be explained by encoding success.

There is a wealth of research linking the hippocampus with associative binding (e.g. Davachi, 2006; Konkel and Cohen, 2009; Eichenbaum and Cohen, 2014; Schwarb et al., 2015; Rangel et al., 2016). We do not deny this is the case, but suggest that our results provoke a reconsideration of the underlying reason for apparent associative effects. We found that the encoding of non-imageble Abstract word pairs did not elicit an increase in hippocampal activity compared to single abstract words, even when only subsequently remembered stimuli were considered. If binding per se was the reason for hippocampal involvement in our study, then this contrast should have revealed it. We suggest that the role of imageability, particularly involving scenes, has been underestimated or ignored in VPA and other associative tasks despite potentially having a significant influence on hippocampal engagement.

Our participants were self-declared low, mid or high imagery users as measured by the VVIQ. They differed in the degree of scene imagery usage in particular during the encoding of Object word pairs, with high imagers showing the greatest amount, and concomitant hippocampal engagement. Given that scene imagery has been implicated in functions across cognition, it might be predicted that those who are able to use scene imagery well might have more successful recall of autobiographical memories and better spatial navigation. Individual differences studies are clearly required to investigate this important issue in depth, as currently

there is a dearth of such work. In the present study, increased use of scene imagery by the high imagery group did not convey a memory advantage for the Object word pairs. However, in the real world, with more complex memoranda like autobiographical memories, we predict that scene imagery would promote better memory.

In conclusion, we showed a strong link between the anterior hippocampus and performance on a VPA task mediated through scene imagery. This offers a way to reconcile hippocampal theories with a visuospatial bias and the memory for verbal material. Moreover, we speculate that this could hint at a verbal memory system in humans piggy-backing on top of an evolutionarily older visual (scene) mechanism. We believe it is likely that other common verbal tests, such as story recall and list learning, which are highly imageble, may similarly engage scene imagery and the anterior hippocampus. Greater use of non-imageble abstract verbal material would seem to be prudent in future verbal memory studies. Indeed, an obvious prediction arising from our results is that patients with selective bilateral hippocampal damage would be better at recalling abstract compared to imageable word pairs, provided care is taken to match the stimuli precisely. Our data do not speak to the issue of whether or not scene construction is the primary mechanism at play within the hippocampus, as our main interest was in examining VPA, a task closely aligned with the hippocampus. What our results show, and we believe in a compelling fashion, is that anterior hippocampal engagement during VPA seems to be best explained by the use of scene imagery.

References

Andelman F, Hoofien D, Goldberg I, Aizenstein O, Neufeld MY (2010) Bilateral hippocampal lesion and a selective impairment of the ability for mental time travel. Neurocase 16: 426-435.

Ashburner J, Friston KJ (2000) Voxel based morphometry - The methods. Neuroimage 11: 805-821.

Ashburner J (2009) Computational anatomy with the SPM software. Magn Reson Imaging 27: 1163-1174.

Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, Neely JH, Nelson DL, Simpson GB, Treiman R (2007) The English Lexicon Project. Behav Res Methods 39: 445-459.

Bellmund JLS, Deuker L, Navarro Schröder T, Doeller CF (2016) Grid-cell representations in mental simulation. eLife 5: e17089.

Brysbaert M, Warriner A, Kuperman V (2014) Concreteness ratings for 40 thousand generally known English word lemmas. Behav Res Methods 46: 904-911.

Caplan JB, Madan CR (2016) Word imageability enhances association-memory by increasing hippocampal engagement. J Cogn Neurosci 28: 1522-1538.

Chadwick MJ, Jolly AEJ, Amos DP, Hassabis D, Spiers HJ (2015) A goal direction signal in the human entorhinal/subicular region. Curr Biol 25: 87-92.

Clark IA, Maguire EA (2016) Remembering preservation in hippocampal amnesia. Annu Rev Psychol 67: 51-82.

Cohen NJ, Eichenbaum H (1993) Memory, amnesia and the hippocampal system. Cambridge: MIT Press.

Dalton MA, Maguire EA (2017) The pre/parasubiculum: a hippocampal hub for scene-based cognition? Curr Opin Behv Sci 17: 34-40.

Dalton MA, Zeidman P, Barry DN, Williams E, Maguire EA (2017) Segmenting subregions of the human hippocampus on structural magnetic resonance image scans: An illustrated tutorial. Brain and Neuroscience Advances 1: 1-36.

Davachi L (2006) Item, context and relational episodic encoding in humans. Curr Opin Neurobiol 16: 693-700.

Deichmann R, Schwarzbauer C, Turner R (2004) Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T. Neuroimage 21: 757-767.

Deuker L, Bellmund JLS, Navarro Schröder T, Doeller CF (2016) An event map of memory space in the hippocampus. eLife 5: e16534.

Eichenbaum H, Cohen NJ (2014) Can we reconcile the declarative memory and spatial navigation views on hippocampal function? Neuron 83: 764-770.

Fanselow MS, Dong H-W (2010) Are the dorsal and ventral hippocampus functionally distinct structures? Neuron 65: 7-19.

Graham KS, Barense MD, Lee ACH (2010) Going beyond LTM in the MTL: A synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. Neuropsychologia 48: 831-853.

Hassabis D, Maguire EA (2007) Deconstructing episodic memory with construction. Trends Cogn Sci 11: 299-306.

Hassabis D, Kumaran D, Vann SD, Maguire EA (2007a) Patients with hippocampal amnesia cannot imagine new experiences. Proc Natl Acad Sci 104: 1726-1731.

Hassabis D, Kumaran D, Maguire EA (2007b) Using imagination to understand the neural basis of episodic memory. J Neurosci 27: 14365-14374.

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC (1998) Enhancement of MR images using registration for signal averaging. J Comput Assist Tomogr 22: 324-333.

Hsieh L-T, Gruber MJ, Jenkins LJ, Ranganath C (2014) Hippocampal activity patterns carry information about objects in temporal context. Neuron 81: 1165-1178.

Hsieh L-T, Ranganath C (2015) Cortical and subcortical contributions to sequence retrieval: Schematic coding of temporal context in the neocortical recollection network. Neuroimage 121: 78-90.

Jones-Gotman M, Milner B (1978) Right temporal-lobe contribution to image-mediated verbal learning. Neuropsychologia 16: 61-71.

Jones-Gotman M (1979) Incidental learning of image-mediated or pronounced words after right temporal lobectomy. Cortex 15: 187-197.

Jones MK (1974) Imagery as a mnemonic aid after left temporal lobectomy: Contrast between material-specific and generalized memory disorders. Neuropsychologia 12: 21-30.

Kim M, Jeffery KJ, Maguire EA (2017) Multivoxel pattern analysis reveals 3D place information in the human hippocampus. J Neurosci 37: 4270-4279.

Konkel A, Cohen NJ (2009) Relational memory and the hippocampus: Representations and methods. Front Neurosci 3: 166-174.

Kosslyn SM, Brunn J, Cave KR, Wallach RW (1984) Individual differences in mental-imagery ability: A computational analysis. Cognition 18: 195-243.

Kousta S-T, Vigliocco G, Vinson DP, Andrews M, Del Campo E (2011) The representation of abstract words: Why emotion matters. J Exp Psychol Gen 140: 14-34.

Kriegeskorte N, Mur M, Bandettini PA (2008) Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2

Kuperman V, Stadthagen-Gonzalez H, Brysbaert M (2012) Age-of-acquisition ratings for 30,000 English words. Behav Res Methods 44: 978-990.

Kurczek J, Wechsler E, Ahuja S, Jensen U, Cohen NJ, Tranel D, Duff M (2015) Differential contributions of hippocampus and medial prefrontal cortex to self-projection and self-referential processing. Neuropsychologia 73: 116-126.

Lakens D (2013) Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 4

Maguire EA, Nannery R, Spiers HJ (2006) Navigation around London by a taxi driver with bilateral hippocampal lesions. Brain 129: 2894-2907.

Maguire EA, Mullally SL (2013) The hippocampus: A manifesto for change. J Exp Psychol Gen 142: 1180-1189.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci 92: 8135-8139.

Marks DF (1973) Visual imagery differences in the recall of pictures. Br J Psychol 64: 17-24.

McAvinue LP, Robertson IH (2007) Measuring visual imagery ability: A review. Imagination, Cognition and Personality 26: 191-211.

McCormick C, Rosenthal CR, Miller TD, Maguire EA (2016) Hippocampal damage increases deontological responses during moral decision making. J Neurosci 36: 12157-12167.

Mechelli A, Price CJ, Friston KJ, Ashburner J (2005) Voxel based morphometry of the human brain: Methods and applications. Curr Med Imaging Rev 1: 105-113.

Milivojevic B, Vicente-Grabovetsky A, Doeller CF (2015) Insight Reconfigures Hippocampal-Prefrontal Memories. Curr Biol 25: 821-830.

Morey RD (2008) Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorial in Quantitative Methods for Psychology 4: 61-64.

Moser M-B, Moser EI (1998) Functional differentiation in the hippocampus. Hippocampus 8: 608-619.

Mullally SL, Intraub H, Maguire EA (2012) Attenuated boundary extension produces a paradoxical memory advantage in amnesic patients. Curr Biol 22: 261-268.

Mullally SL, Maguire EA (2014) Counterfactual thinking in patients with amnesia. Hippocampus 24: 1261-1266.

Murray EA, Bussey TJ, Saksida LM (2007) Visual perception and memory: A new view of medial temporal lobe function in primates and rodents. Annu Rev Neurosci 30: 99-122.

O'Keefe J, Nadel L (1978) The hippocampus as a cognitive map. Oxford: Clarendon Press.

Paivio A, Yuille JC, Madigan SA (1968) Concreteness, imagery and meaningfulness values for 925 nouns. J Exp Psychol 76

Paivio A (1969) Mental imagery in associative learning and memory. Psychol Rev 76: 241-263.

Paivio A, Walsh M, Bons T (1994) Concreteness effects on memory - When and why. J Exp Psychol Learn 20: 1196-1204.

Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013) Long-axis specialization of the human hippocampus. Trends Cogn Sci 17: 230-240.

Race E, Keane MM, Verfaellie M (2011) Medial temporal lobe damage causes deficits in episodic memory and episodic future thinking not attributable to deficits in narrative construction. J Neurosci 31: 10262-10269.

Rangel LM, Rueckemann JW, Riviere PD, Keefe KR, Porter BS, Heimbuch IS, Budlong CH, Eichenbaum H (2016) Rhythmic coordination of hippocampal neurons during associative memory processing. eLife 5: e09849.

Ritchey M, Montchal ME, Yonelinas AP, Ranganath C (2015) Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. eLife 4: e05025.

Rorden C, Brett M (2000) Stereotaxic display of brain lesions. Behav Neurol 12: 191-200.

Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK (2012) The future of memory: Remembering, imagining, and the brain. Neuron 76: 677-694.

Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM (2016) Statistical learning of temporal community structure in the hippocampus. Hippocampus 26: 3-8.

Schuck Nicolas W, Cai Ming B, Wilson Robert C, Niv Y (2016) Human orbitofrontal cortex represents a cognitive map of state space. Neuron 91: 1402-1412.

Schwarb H, Watson PD, Campbell K, Shander CL, Monti JM, Cooke GE, Wang JX, Kramer AF, Cohen NJ (2015) Competition and cooperation among relational memory representations. PLoS One 10: e0143832.

Stadthagen-Gonzalez H, Davis C (2006) The Bristol norms for age of acquisition, imageability, and familiarity. Behav Res Methods 38: 598-605.

Staresina BP, Henson RNA, Kriegeskorte N, Alink A (2012) Episodic reinstatement in the medial temporal lobe. J Neurosci. 32: 18150-18156.

Strange BA, Witter MP, Lein ES, Moser EI (2014) Functional organization of the hippocampal longitudinal axis. Nat Rev Neurosci 15: 655-669.

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

Vigliocco G, Kousta S-T, Della Rosa PA, Vinson DP, Tettamanti M, Devlin JT, Cappa SF (2014) The neural representation of abstract words: The role of emotion. Cereb Cortex 24: 1767-1777.

Warriner AB, Kuperman V, Brysbaert M (2013) Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav Res Methods 45: 1191-1207.

Wechsler D (2008) Wechsler adult intelligence scale–Fourth edition (WAIS–IV). San Antonio, TX: NCS Pearson.

Wechsler D (2009) WMS-IV.: Wechsler memory scale-Administration and scoring manual. Psychological Corporation.

Wechsler D (2011) Test of premorbid functioning. UK version (TOPF UK). London: Pearson Assessment.

Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. Neuroimage 33: 493-504.

Zeidman P, Maguire EA (2016) Anterior hippocampus: the anatomy of perception, imagination and episodic memory. Nat Rev Neurosci 17: 173-182.