

Memory Encoding and Retrieval: The Role of Attention, Representations and Networks

by

Benjamin Ryan Geib

Department of Psychology and Neuroscience
Duke University

Date: _____

Approved:

Roberto Cabeza, Supervisor

Marty Woldorff, Supervisor

John Pearson

Tobias Egner

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Psychology and Neuroscience in the Graduate School
of Duke University

2020

ABSTRACT

Memory encoding and retrieval: The role of attention, representations and networks

by

Benjamin Ryan Geib

Department of Psychology and Neuroscience
Duke University

Date: _____

Approved:

Roberto Cabeza, Supervisor

Marty Woldorff, Supervisor

John Pearson

Tobias Egner

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Psychology and Neuroscience in the Graduate School of
Duke University

2020

Copyright by
Benjamin R. Geib
2020

Abstract

Episodic memory, as a cognitive construct, exists only in relation to those other cognitive constructs that reference it. It is, as Ribot suggests: the tactile, the muscular, the auditory and so forth. And it is even more than this, extending to a breadth of cognitive operations, including, for example, attention and cognitive control, both of which are generally believed to facilitate episodic memory encoding and episodic memory retrieval. Without these types of sensory and cognitive referents, episodic memory does not exist. Accordingly, these types of referents are critical to an understanding of episodic memory. Therefore, in this dissertation I examine how different cognitive constructs serve to facilitate episodic memory.

Chapter 2 examines attention-related subsequent memory effects. Many studies of subsequent memory rely upon a reverse inference, i.e. increased activity in attention-related networks during memory encoding is related to better subsequent memory, ergo increased attention predicts better memory. However, it is only through direct manipulation of attentional states and the examination of specific neural markers that this claim can be strongly established. Additionally, attention is a multifaceted process, and claims that attention in general facilitates memory ignore the fact that attention consists of a set of rapidly unfolding processes. To address these issues, I designed a modified visual-search EEG experiment with a subsequent long-term memory test. The utilization of a visual-search paradigm has advantages, as the search process evokes a

series of independent and well-established attention-related EEG markers which can be linked to subsequent memory. All of the attentional effects examined were found to also predict subsequent memory, suggesting that these attentional processes associated with visual search, aid long-term memory formation as well.

Chapter 3 examines how large-scale network dynamics affect long-term memory retrieval. Until now, *all* studies of long-term memory have focused on individual regions, pair-wise connections between regions, or, very rarely, complex interactions between a small subset of regions. In a pair of fMRI studies, I use mathematical concepts from network science to examine the large-scale brain networks associated with successful remembering and forgetting. In doing so, I demonstrate that the hippocampus increases its integration with the rest brain when individuals successfully remember an item as compared to when they do not.

Chapter 4 examines how individual items are represented in the brain with machine-learning techniques and fMRI data. Studies of episodic memory often focus on things that are common across a set of items, while ignoring the uniqueness of individual events. However, an event's uniqueness is what defines it as being episodic with respect to memory. A primary reason unique events are not often studied is the difficulty of decoding brain states associated with individual events. In Chapter 4, I develop a machine-learning framework, utilizing cross-subject single-item decoding, to predict what image or word a left-out subject is viewing. This establishes a robust way

to detect individual events which could be used in service of better understanding episodic memory.

By examining long-term memory from these perspectives, I provide evidence of how different cognitive constructs facilitate episodic memory. In Chapter 2, I focus on the role of attentional processes with respect to episodic memory encoding, in Chapter 3, I focus on how large-scale network interactions facilitate episodic memory retrieval, and in Chapter 4 I focus on the representational nature of unique events. In all cases, the examination centers on how diverse processes coordinate in order to facilitate episodic memory.

Contents

Abstract.....	iv
List of Tables.....	xiii
List of Figures.....	xiv
Acknowledgements	xvi
1. General Introduction.....	1
1.1 Linking Episodic Memory to Other Cognitive Functions.....	3
1.2 Linking Episodic Memory to Large-Scale Network Interactions	6
1.3 Linking Episodic Memory to Individual Memory Traces	10
2. Attentional contributions to encoding.....	15
2.1 Linking the rapid cascade of visuo-attentional processes to successful memory encoding	15
2.1.1 Introduction.....	16
2.1.2 Materials and methods	18
2.1.2.1 Participants.....	18
2.1.2.2 Paradigm.....	19
2.1.2.3 EEG Processing	21
2.1.2.4 EEG Statistical Analyses.....	24
2.1.3 Results.....	26
2.1.3.1 Behavioral.....	26
2.1.3.2 ERP DM	27
2.1.3.3 Lateralized ERP attention effects at encoding: N2pc & SPCN	28

2.1.3.4 Oscillatory Contralateral Reduction in Alpha (CRA) [8-12 Hz]	30
2.1.3.5 Oscillatory Theta [4-7 Hz]	31
2.1.3.6 Independent and dependent contributions to memory	32
2.1.4 Discussion.....	35
2.1.4.1 Rapid-onset attentional processes as indexed by the N2pc, SPCN, and CRA predict subsequent-memory.....	37
2.1.4.2 The early attentional processes indexed by N2pc, SPCN, and CRA do not contribute to the DM	39
2.1.4.3 The magnitude of the mid-frontal theta was associated with the DM effect	40
2.1.5 Conclusion.....	41
3. Hippocampal Networks	43
3.1 Hippocampal contributions to the large-scale episodic memory network predict vivid visual memories	43
3.1.1 Introduction.....	44
3.1.2 Methods	51
3.1.2.1 Participants.....	51
3.1.2.2 Behavioral Methods.....	52
3.1.2.3 MRI scanning and image processing.....	53
3.1.3.4 Functional brain network construction	54
3.1.3.5 Graph Theory Measures	56
3.1.3.6 Statistical Testing	63
3.1.3 Results.....	64
3.1.3.1 Behavioral Results.....	64

3.1.3.2 Effects of hippocampal network properties on memory vividness	65
3.1.3.3 Effects of hippocampal network properties on memory vividness compared to other network nodes	69
3.2.3.4 Effects of hippocampal first-step connection reorganization on memory vividness	72
3.2.3.5 Extra-hippocampal global network effects on memory vividness.....	75
3.1.4 General Discussion.....	76
3.1.4.1 Right hippocampal network interactions significantly impact memory vividness	77
3.1.4.2 Network measures for the right hippocampus tracked memory vividness more than any other brain region	80
3.1.4.3 Hippocampal first-step connections substantially reorganize to support vivid memory retrieval	83
3.1.4.4 Extra-hippocampal global network changes support vivid memory retrieval	84
3.1.4.5 Conclusions	85
3.2 From hippocampus to whole-brain: The role of integrative processing in episodic memory retrieval	86
3.2.1 Introduction	87
3.2.2 Methods	92
3.2.2.1 Participants.....	92
3.2.2.2 Experimental Design.....	93
3.2.2.3 Data Acquisition and Pre-Processing	94
3.2.2.4 Generating Whole Brain Networks	96
3.2.2.5 Graph Theory Measures	98
3.2.2.6 Retrieval Assembly Construction	103

3.2.2.7 Statistical Testing of Graph Theory Measures	104
3.2.3 Results.....	105
3.2.3.1 Behavioral Results.....	105
3.2.3.2 Network Results.....	106
3.2.4 Discussion.....	120
3.2.4.1 Hippocampal integration: increased global efficiency and degree centrality for remembered compared to forgotten items	121
3.2.4.2 Assembly organization: Successful memory is supported by increased hippocampal connectivity with a <i>retrieval assembly</i>	122
3.2.4.3 Assembly integration: Global efficiency increases and participation coefficient increases for remembered compared to forgotten items.....	125
3.2.4.4 Whole brain integration: Decreased modularity for remembered compared to forgotten items.....	126
3.2.4.5 Conclusions	128
4. Single-item representations.....	131
4.1 The science of the singular: single-item decoding with multivariate pattern analysis	131
4.1.1 Introduction.....	132
4.1.2 Methods	137
4.1.2.1 Paradigms.....	137
4.1.2.2 Stimulus-Similarity Matrices.....	139
4.1.2.3 fMRI Acquisition and Processing	140
4.1.2.4 Logistic Regression (L-2 norm).....	142
4.1.2.5 Confusion Matrix	144
4.1.2.6 <i>Analyses for Decoding Individual Items</i>	145

4.1.2.7 Analyses for Measuring Similarity Effects	147
4.1.2.8 Single Item Decoding Code Repository	150
4.1.3 Results	150
4.1.3.1 Decoding Individual Items (picture-to-picture)	151
4.1.3.2 Effects of Similarity (picture-to-picture).....	153
4.1.3.3 Word-to-word, picture-to-word, word-to-picture results	156
4.1.4 Discussion.....	160
4.1.4.1 Decoding Individual Items.....	161
4.1.4.2 Effects of Similarity	164
4.1.4.3 Limitations and Caveats	166
4.1.4.4 General Conclusion.....	167
4.1.3 Supplementary Materials	167
4.1.3.1 Determining whether single-items can be decoded	167
4.1.3.2 Determining optimal LR-LOPO decoding parameters	171
4.1.3.3 Effects of Similarity	178
5. Conclusions and Future Directions	179
5.1 Linking Episodic Memory to Other Cognitive Functions.....	179
5.1.1 Summary.....	180
5.1.2 Future Directions.....	182
5.2 Linking Episodic Memory to Large-Scale Network Interactions	184
5.2.1 Summary.....	184
5.2.2 Future Directions.....	186

5.3 Linking Episodic Memory to Individual Memory Traces	187
5.3.1 Summary.....	188
5.3.2 Future Directions.....	190
5.4 Overarching Conclusions.....	191
Appendix A.....	192
Materials & Methods	192
Participants.....	192
Stimuli.....	193
Semantic feature norms.....	193
Procedure	195
Encoding	195
Retrieval	196
MRI Acquisition.....	196
References.....	199
Biography	217

List of Tables

Table 1. Behavioral Results. Summary of behavioral results. Note that Remembered and Forgotten items encompassed all responses to known objects only. Reported values.....	27
Table 2. Summary of medians and p-values obtained using the permutation framework for each network measure at the right hippocampus and for the entire brain network. .	67
Table 3. Provides a summary of the behavioral results. Only high-confidence hits were included in the remembered network, whereas both high- and low-confidence	105
Table 4. Changes in hippocampal connectivity with nodes comprising the retrieval assembly. Nodes included in the assembly exhibited a significant increase in	110
Table 5. Global efficiency for all nodes included in the retrieval assembly. Statistical significance for changes in global efficiency between remembered and forgotten	113
Table 6. Participation coefficients for all nodes included in the retrieval assembly. Statistical significance for changes in participation coefficients between	118
Table 7. Summarized decoding results: rows in blue are from the scene study and rows in green are from the object study. All measures are derived from t-tests within	159
Table 8. Results from equation 1. A.Left. Result from combined analysis, across the scene and object study {df=1280=32 models*40 participants} B.Right(Blue) Results from	170
Table 9. Results from equation 2 {df=320=8 models*40 participants for Object & Scene; 8*21=168 for Scene Study and 8*19=152 for Object Study}. Each column of the.....	173
Table 10. Results from all scene-paradigm models. The format is identical to table S1 with the inclusion of additional columns (left) pertaining to model	176
Table 11. Results from all object-paradigm models. The format is identical to Table 10 with the inclusion of additional columns (left) pertaining to model	177
Table 12. Results for items {df=2016=96*21 for the scene paradigm (blue) and df=5700=300*19 for the object paradigm (green)} ***p<0.001, **p<0.01, *p<0.05,	178

List of Figures

- Figure 1. Visual depiction of nodes with high and low integration in small toy networks. The node on the left (low) has low integration whereas the node on the right (high). 9
- Figure 2. Provides a brief graphical overview of our analysis problem. Typical univariate analyses essentially average *all* stimuli across subjects within some condition prior to .. 12
- Figure 3. Examples of encoding and retrieval trials. (A) For the encoding phase, subjects were instructed to covertly shift their attention to the lateralized object circumscribed . 21
- Figure 4. DM effect in central electrodes during. Running t-tests identified an epoch from 300 to 850ms (shaded gray) where subsequently remembered items had a..... 28
- Figure 5. Both the N2pc and the SPCN were larger for subsequently remembered than forgotten items. Running t-tests identified an epoch from 200 to 800ms. 30
- Figure 6. Mid-frontal theta was greater for subsequently remembered than forgotten items (all $p < 0.05$). 32
- Figure 7. A multi-level mediation model showed that superior-central positivity of the DM effect partially mediated the relationship between pre-DM theta power..... 35
- Figure 8. EEG effects enhancing subsequent-memory. The schematic figure displays a set of parallel processes that all serve to independently contribute to successful memory... 36
- Figure 9. Provides an overview of the experimental design. The same design was presented in Wing et al. (2015). (A) During encoding, pictures of scenes were..... 47
- Figure 10. Average (across subjects) adjacency matrices derived from beta series correlations are presented for dim and vivid retrieval conditions and split by..... 56
- Figure 11. The change in (z-scored) degree centrality values are presented for all nodes that shift in a positive direction (i.e., nodes that exhibit higher degree centrality for 68
- Figure 12. Provides a summary of how potential memory-related network nodes change between dim and vivid retrieval with respect to the four nodal measures of interest..... 70
- Figure 13. Provides a summary of the t-values associated with the average change in connectivity strength between the right hippocampus and the set of potential 74
- Figure 14. Average (across participants) adjacency matrices derived from beta series correlations are represented for remembered and forgotten networks and split 98

Figure 15. Global efficiency of the left and right hippocampus for remembered and forgotten trials, respectively. The permutation framework developed by.....	107
Figure 16. Provides an overview of the retrieval assembly and modular nodal assignments. All nodes in the assembly (depicted as the large nodes in the figure).....	109
Figure 17. (A) Average global efficiency of all nodes in the retrieval assembly for both remembered and forgotten networks. The permutation framework developed by.....	112
Figure 18. Provides a summary of the modularity findings. A: The permutation procedure developed by Simpson et al. [2013] in conjunction with the Jaccardized.....	115
Figure 19. Format of the two studies included in the analyses	139
Figure 20. Similarity structure of the two studies. (A: Left) Feature similarity for the scene paradigm and (B: Right) Semantic similarity for the object paradigm. Note.....	140
Figure 21. displays above-chance classification accuracy for the scene-paradigm (A: Blue) and the object-paradigm (B: Green) with within category accuracy being	152
Figure 22. displays above-chance classification accuracy for the scene paradigm (A: Blue) and the object paradigm (B: Green), with the within category evidence being shown..	153
Figure 23. displays the average classification evidence within (on diagonal) and between (off diagonal) categories for the scene paradigm (A) and object paradigm (B).	154
Figure 24. Regression results from equation 1 (see Table S3 for exact beta values) for picture-to-picture decoding. For visualization purposes, we set a baseline to be high..	156
Figure 25. Results from equation 1 plotted as a function of beta estimates from the associated linear model. The results indicate that the LR-LOPO models, in general.....	171
Figure 26. Effect of debiasing the data. (Left) displays raw data from a subset of trials and voxels from an example subject in the scene-study. The histograms on the.	174
Figure 27. Effect of normalization on the confusion matrix (example subject from the scene study). As can be seen, the normalization procedure removes streaks in the	175
Figure 28. Task Paradigm.....	196

Acknowledgements

First, I would like to thank my parents, Monica and Doug, for all their love and support. They held me up, when I no longer could, and for that I am grateful.

I would like to thank my advisors, Dr. Marty Woldorff and Dr. Roberto Cabeza. Coming into a psychology program with an engineering background was certainly a shift. But through their guidance, support, and mentorship I found my footing and began to thrive. They propelled me through graduate school, and I feel very well prepared as I continue my career as a post-doctoral researcher.

I would like to thank my informal advisors and mentors who helped guide me through graduate school: Dr. Erik Wing, Dr. Simon Davis, and Dr. Felipe deBrigard. Their insights and support have been invaluable, and I thoroughly enjoyed the all projects we collaboratively engaged in.

I would like to thank those fellow graduate students I've had the privilege of co-authoring with: Matt Stanley, Zach Monge, Berry van den Berg, Lifu Deng, Bryce Gessell and Peter Whitehead. These projects were all

I would like to thank my committee, Dr. John Pearson and Dr. Tobias Egner for providing valuable feedback and comments pertaining to my research.

I would like to thank both the Woldorff Lab and Cabeza lab for their invaluable support throughout graduate school as well. Further thanks to my undergraduate

mentees: Zain Azeem, Kaya Schemem, and especially Patrick Cardell and Grace Carlson.

Thanks to the best cohort, the class of 2013, Dianna Amasino, Charlie Giattino, Ellie Hanna, and Taylor Jackson.

A special thanks to all the great friends I've made at Duke: Berry van den Berg, Charlie Giattino, Johnathon Winkle, Matt Stanley, Natasha Parikh, Sam Dore, Shawn Willet and apologies to those I forgot to add.

Finally, thanks to my beautiful girlfriend Tabitha Steendam for all her love and support over this past year.

1. General Introduction

We have thus, as M. Ribot says, not memory so much as memoires.

The visual, the tactile, the muscular, the auditory ...

Principles of Psychology, William James, 1890

Episodic memory, as a cognitive construct, exists only in relation to those other cognitive constructs that reference it. It is, as Ribot suggests: the tactile, the muscular, the auditory and so forth. And it is even more than this, extending to a breadth of cognitive operations, including, for example, attention and cognitive control, both of which are generally believed to facilitate episodic memory encoding and episodic memory retrieval. Without these types of sensory and cognitive referents, episodic memory does not exist. Accordingly, these types of referents are critical to an understanding of episodic memory.

The role of additional cognitive operations, such as attention and cognitive control, are critical to memory encoding, yet exactly how these types of specific processes aid in the successful formation of memory is often unclear. In Chapter 2, I present a novel electroencephalogram (EEG) study that directly manipulates rapid attentional demands during episodic memory encoding in order to draw clear links between specific attentional processes and successful encoding. Links between cognitive control processes and successful encoding are also surmised from the results from this study according to existing literature.

Having thus demonstrated how these cognitive operations aid encoding, I next examine how large-scale network dynamics aid episodic memory. Here the focus is *not* so much on how specific cognitive functions aid episodic memory, but on how the whole brain works in coordination to facilitate episodic memory retrieval process.

While the hippocampus has been the focus of memory retrieval processes, it is well known that the episodic memory retrieval requires more than the hippocampus alone. In Chapter 3, in two functional magnetic resonance imaging (fMRI) studies, I examine how large-scale network dynamics function in the service of episodic memory retrieval; with a particular focus on how the hippocampus interacts with the rest of the *entire* brain to facilitate episodic memory retrieval. Again, the focus is on how episodic memory requires coordination across multiple cognitive subsystems in order to facilitate general retrieval processes.

However, due to methodological and technical limitations, an examination of network dynamics with fMRI requires the concatenation of multiple trials, thus limiting conclusions to more general memory processes common across large sets of trials. Accordingly, this approach fails to address the critical role of individual sensory representations (associated with individual trials) with respect to episodic memory – which is addressed next.

Chapter 3 and Chapter 4 examine how general cognitive processes facilitate episodic memory. However, these chapters do not address the nature of individual

(trial-wise) memory traces, which are the hallmark of episodic memory. Indeed, much of the literature doesn't address this key issue either, as the examination of individual trials is quite difficult given that the measured responses of single trials are very noisy. In Chapter 4, using two fMRI memory reactivation datasets, I introduce a novel application of machine-learning techniques to demonstrate the ability to decode individual trials by using cross-subject information. The ability to detect individual trial-level representations is of substantial value as it theoretically allows one to examine how sensory specific representations interact with more general cognitive processes to facilitate episodic memory encoding and retrieval.

1.1 Linking Episodic Memory to Other Cognitive Functions

The role of additional cognitive operations, such as attention, are critical to memory encoding, and the primary focus of Chapter 2 is on examining the role of attention during episodic memory encoding. Although *links between attention and successful memory encoding* are obvious, a surprisingly small amount of evidence supports this assertion in a specific functional way. Moreover, many of these *links are dependent upon a strong reverse inference* and/or treat attention as a uniform construct. Accordingly, in Chapter 2, I operationalize a paradigm that aims to minimize reverse inference while also focusing specifically on the role of *early attentional processes*, as, like memory, attention is multifaceted and decomposable into important subcomponents.

While a few cognitive neuroscience studies have directly manipulated early attentional processes during encoding e.g. (Chastelaine & Rugg, 2014; Turk-Browne, Golomb, & Chun, 2013; Uncapher, Hutchinson, & Wagner, 2011), all of these studies have utilized functional magnetic resonance imaging (fMRI), which lacks the temporal resolution to be able to delineate the cascade of visuo-attentional processes, such as distinguishing early versus late attentional modulations, given that this progression proceeds in less than second. In general, this lack of temporal specificity has led to many conclusions pertaining to attention and memory interactions to be driven by *reverse inference* (Poldrack, 2011).

In general, reverse inference is driven by a lack of specificity and/or a lack of experimental manipulation. Accordingly, to minimize the degree of reverse inference when linking attentional and mnemonic functions, I designed a visual-search EEG experiment that specifically manipulates visual spatial attention, followed by a subsequent memory test, to examine how temporally specific facets of the visuo-attentional cascade contribute to encoding success. This approach has some key advantages. With regards to specificity, visual-search tasks evoke a rapid sequence of well-characterized spatio-temporally specific EEG components that are very specific to the focusing of visual spatial attention. It is the specificity of these components that allows one to draw tight links between the associated attentional processes and their function role in memory encoding. Additionally, by engaging participants in a visual

search task, I ensure that this well-defined sequence of attentional operations is specifically and clearly elicited. I then look for modulations in these evoked attentional processes, as reflected by the modulation of their specific EEG neural markers that predict later memory performance. As mentioned above, visual search tasks evoke a cascade of EEG components that are modulated by attentional focus and demands, including the N2pc, the SPCN, and the contralateral reduction in alpha power (8-12 Hz).

(1) The N2pc (negative-polarity posterior-contralateral peaking at about 225 ms)(Luck & Hillyard, 1994) has been related to stimulus-specific attentional allocation toward a particular stimulus location (reviewed in: (Kiss, Van Velzen, & Eimer, 2008)), as well as by capture by specific stimulus items of particular salience or relevance (Qi, Zeng, Ding, & Li, 2013). (2) Following the N2pc, the SPCN (sustained posterior-contralateral negativity; >300ms) has been linked to sustained attentional processing for item-feature discrimination (Jolicoeur, Brisson, & Robitaille, 2008), and is generally larger when the target discrimination task is more difficult. (3) Finally, temporally co-occurring with the SPCN (>300ms), there is a post-stimulus reduction in oscillatory alpha power contralateral to the target, which we call here contralateral reduction in alpha (CRA). As with most other decreases of alpha, this effect is indicative of increased cortical activity e.g. (Laufs et al., 2003). The CRA has also been linked to longer-latency, attentionally induced, item-analysis processes (van Diepen, Miller, Mazaheri, & Geng, 2016), which

are cognitively distinct from the SPCN (Bae & Luck, 2017; De Vries, van Driel, & Olivers, 2017; Fukuda, Mance, & Vogel, 2015).

In summary, my first goal (Chapter 2: Study 1) was to establish more clear links between specific attentional processes and encoding success.

1.2 Linking Episodic Memory to Large-Scale Network Interactions

Memory retrieval most often focuses on individual regions, in particular the hippocampus, or small-scale interactions between the hippocampus and other regions (e.g., prefrontal cortex) – all the while ignoring the more complex interconnectedness across the entire cortex. I suggest that these more complex interactions are also foundational to long-term episodic memory retrieval. In particular, in chapter 3 I present two studies that focus on the role of the hippocampus with regards to the larger cortical network.

The contributions of the hippocampus to episodic memory have been the focus of thousands of lesion, electrophysiology, drug, and neuroimaging studies with both animals and humans. Lesion studies have provided clear evidence for the devastating effects of hippocampal damage on episodic memory, and functional neuroimaging studies with healthy adults have reliably shown greater hippocampal activity for successful than unsuccessful episodic retrieval and for rich than impoverished episodic memories (Eichenbaum, Yonelinas, & Ranganath, 2007; Kim, 2015; Skinner & Fernandes, 2007; Spaniol et al., 2009). The hippocampus is hypothesized to bind incoming

information from different neocortical regions, to store integrated event representations, and to allow access to cortical memory traces during retrieval (Alvarez & Squire, 1994; Danker & Anderson, 2010; McClelland, McNaughton, & O'Reilly, 1995; Ritchey, Wing, Labar, & Cabeza, 2013; Teyler & Rudy, 2007). During these processes, the hippocampus is assumed to interact very closely not only with regions storing representations but also with frontal, parietal, cingulate, and basal ganglia regions involved in a variety of attention, control, working memory, and decision making processes. Consistent with these assumptions, functional neuroimaging studies have shown that all these regions are co-activated with the hippocampus during episodic encoding and retrieval tasks (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Cabeza & Nyberg, 2000; Kim, 2013; Rugg & Vilberg, 2013; Spaniol et al., 2009). Even though all these regions are assumed to operate together as a complex large-scale network, existing functional connectivity studies of episodic memory have typically focused only on the relationship between a particular pair of regions, such as the interaction between the hippocampus and the prefrontal cortex e.g. (Schott et al., 2013; Wing, Marsh, & Cabeza, 2013). Only recently have researchers begun to use task-related functional connectivity analyses to investigate changes across larger sets of memory related brain regions (King, de Chastelaine, Elward, Wang, & Rugg, 2015; Schedlbauer, Copara, Watrous, & Ekstrom, 2014).

Clearly the hippocampus plays an important role in memory retrieval, as do many other cortical structures. The goal of this second investigation was not to contest this view, but to also suggest that the nature of the *interactions* between the hippocampus and specific cortical regions are critical. Therefore, the assertion that **memory is not based solely on a particular region, or even particular set of regions**, is grounded in hypothesis that it is the network -- i.e. the pattern of connectivity across the brain - that is critical. The core hypothesis of Chapter 3 (Study 2 and Study 3) is that the hippocampus becomes more **integrated** with the rest of the brain during successful long-term episodic memory retrieval.

Networks consist of two fundamental elements: **nodes**, which are information processing centers, and **edges**, which are the connections between nodes. A common example, with regards to social networks is that individuals are nodes and the connections between them (say friendship) are edges. In brains, cortical regions are nodes and the connections between them are edges. Nodes, regardless of the network, that have high integration can easily communicate with any other node in the network (Figure 1). This is not based just upon the number of connections a node has, but also on the nature of those connections. For example, imagine a large corporation as a network. The CEO and a lower-level employee can have an equal number of connections (colleagues) within the corporation, but the CEO, based upon the nature of his or her connections, can more easily communicate with any other individual in the network.

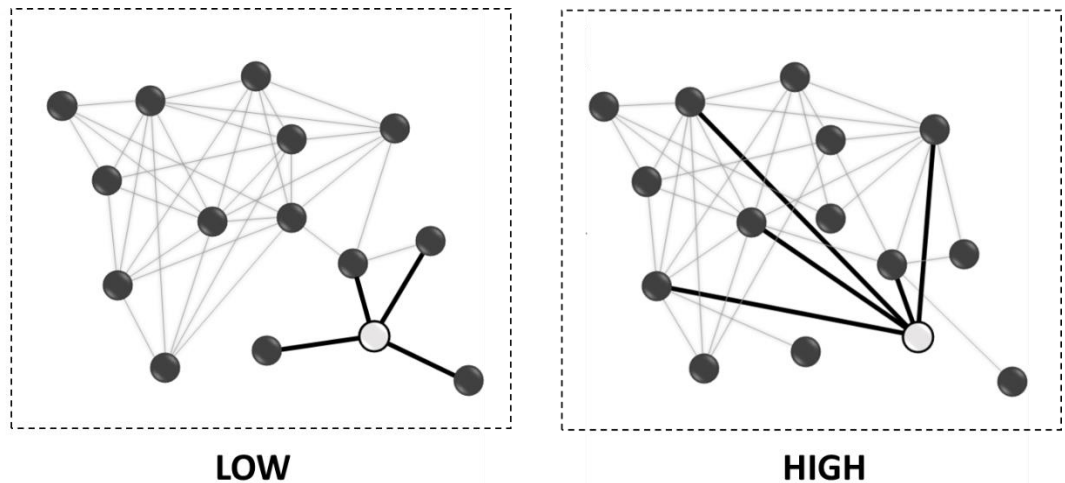


Figure 1. Visual depiction of nodes with high and low integration in small toy networks. The node on the left (low) has low integration whereas the node on the right (high) has high integration. Whereas both nodes are depicted as having four connections, it is easier for the node on the right to communicate with the rest of the network.

Through a similar thought process, I surmised the hypothesis that the hippocampus becomes more integrated with the rest of the brain during successful memory retrieval. That is, the hippocampus doesn't simply massively increase its number of connections during successful memory retrieval, it fundamentally changes its connections in a fashion that facilitates information transfer across the cortex. In Chapter 3, Study 2, I focus primarily on the role of the hippocampus, whereas in Chapter 3, Study 3, I expand this focus to those regions showing the strongest bivariate interactions with hippocampus.

1.3 Linking Episodic Memory to Individual Memory Traces

Most episodic memory studies focus on the commonality across a *set* of memories while ignoring the *individual* memory. The primary reason for this is methodological, as it is exceedingly difficult to decode responses to individual events in the brain. As such, in my final investigation (Study 4), I developed a machine-learning framework to decode responses to individual events, with the aim that this could be applied to investigating the nature of individual memories.

Investigations of episodic memory are severely hindered by the inability to examine individual events, given that the individuality of the stimulus event is the hallmark of the recollection process. When recalling a coastal sunset with friends and family, you recollect *your* family, *your* friends, that *specific* coast, all in relation to that *individual* event, and this collective recall of that individual event is the fundamental definition of recollection. The study of this uniqueness, however, presents episodic memory researchers with severe signal-to-noise limitations, as our stimuli are defined by their individuality, which, like the memory of that *specific* sunset, is unique.

The study of episodic long-term memory can therefore be divided into **general** (i.e. event non-specific) and **stimulus-specific** (i.e. event-specific) processes (Figure 2). With regards to episodic memory, more general processes are largely invariant to the particularities associated with individual stimuli; however, these more general processes constitute a core dogma of episodic memory theory. For example, the hippocampus is

critical to binding contextual and item-specific information into an integrated memory representation, regardless of the identity of the individual stimulus , and during retrieval, access to the hippocampal representation allows the recovery of these cortical memory traces, wherever they might be stored (Ranganath, 2010). Additionally, encoding and retrieval process are supported by control/attentional processes in prefrontal cortex (PFC) and parietal regions, again regardless of what specific memory is being retrieved. These general, or “bottleneck regions”, are typically emphasized in functional neuroimaging analyses that average across stimuli. For example, recollection has been strongly associated with a “core recollection network” consisting of the hippocampus, ventral parietal cortex, and posterior midline regions (King et al., 2015). However, missing from these general regions are regions associated with the particularities of individual stimuli, which are averaged out. Yet, these particularities are in some fundamental way what defines episodic memory.

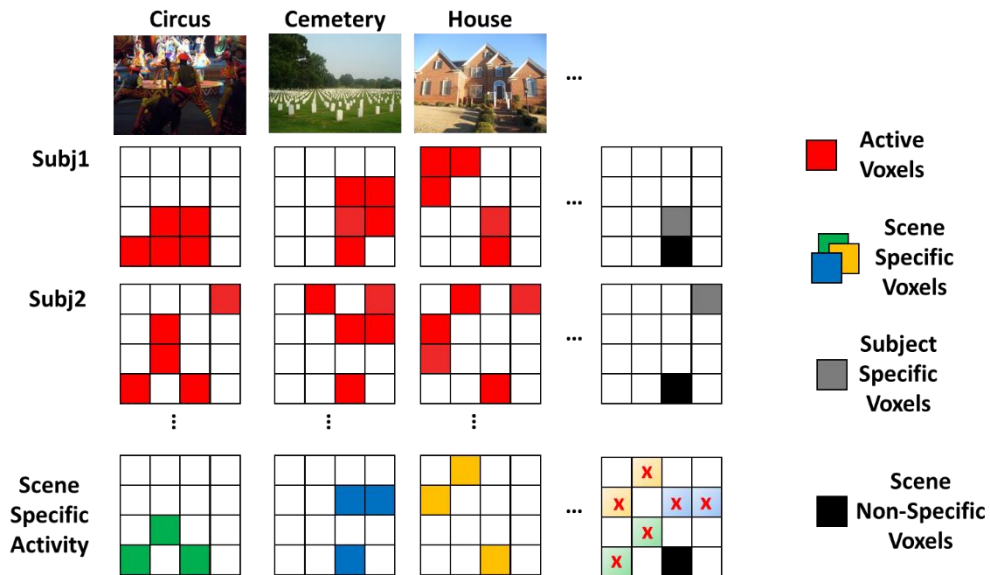


Figure 2. Provides a brief graphical overview of our analysis problem. Typical univariate analyses essentially average *all* stimuli across subjects within some condition prior to statistical tests. As consequence a single contrast (black) picks up on general stimulus related activity. However, this technique ignores the fact that there are likely consistent activations for complex individual stimuli (green, blue and gold). Thus, as shown in the bottom right a standard contrast misses a whole swath of regions which show consistent effects across subjects.

The study of stimulus-specific processing, in long term memory, is dominated by the theory of retrieval reactivation (reviewed in: (Danker & Anderson, 2010)), with only a few studies examining encoding strength e.g. (Gordon, Rissman, Kiani, & Wagner, 2014). According to reactivation theory, episodic memory traces are stored within the same brain regions that process each kind of information during encoding; visual memories are stored in visual cortex, auditory memories in auditory cortex, face information in face-sensitive cortex, and so forth. While many methods have been used to study reactivation, the most common are multivoxel pattern analyses (MVPA) e.g.

(Lee, Chun, & Kuhl, 2016; Polyn, Natu, Cohen, & Norman, 2005) and representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008); e.g. (Kuhl & Chun, 2014; Wing, Ritchey, & Cabeza, 2015).

MVPA consists of a wide assortment of machine learning techniques that are most often utilized to discriminate between different categories, and, as such, it is most advantageous when applied to associative encoding paradigms. As an example, during encoding, subjects have been presented with random words paired to either a face or place, and then during retrieval, the word alone is presented and subjects are asked what category of image it was associated with at encoding; MVPA-derived classification evidence for the paired associate is higher when subjects accurately remember the stimulus category – face vs. place (Lee et al., 2016). As such, in the context of MVPA, reactivation has always been measured at the categorical-level as opposed to stimulus-specific level.

Representational similarity analysis, unlike MVPA, treats all stimuli independently. Within the domain of memory, it has been used to assess encoding-retrieval similarity (ERS) (Kuhl & Chun, 2014; Wing et al., 2015) and as shown that for remembered items (as compared to forgotten ones) there is a greater pattern-similarity between matched encoding and retrieval trials at the stimulus-specific level. The major limitation of this technique is region-localization restrictions, as the similarity structure is only accessed within a region of interest (ROI). This demands that the ROI code for *all*

stimuli in the stimulus set. However, this is rarely the case, as different cortical regions appear to be specialized for certain types of stimuli (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Huth, Nishimoto, Vu, & Gallant, 2012). Additionally, ERS conflates the encoding and retrieval processes by uniting them into a single metric, and thus with this approach it is unclear if high ERS is driven by processes during encoding, retrieval, or a mix of both. To summarize, MVPA techniques often struggle when examining the representation of single stimuli, and while RSA can capitalize on subtle differences between the responses to single stimuli, it's limited to localized regions and mixes encoding and retrieval operations. In Chapter 4 (Study 4), I introduce a novel algorithm that offers a way to overcome these limitations to more accurately study encoding strength and retrieval reactivation by demonstrating the ability to classify individual items that have only been presented to once to subject¹.

¹ Note that I do not investigate memory-related effects in this chapter, but rather I only provide a framework for single-item decoding that could, in the future, be applied to study these concepts.

2. Attentional contributions to encoding

2.1 Linking the rapid cascade of visuo-attentional processes to successful memory encoding

While it is broadly accepted that attention modulates memory, the contribution of specific rapid attentional processes to successful encoding is largely unknown. To investigate this issue, we leveraged the high-temporal-resolution of electroencephalographic (EEG) recordings to directly link a cascade of visuo-attentional neural processes to successful encoding: namely (1) the N2pc (peaking ~200ms), which reflects stimulus-specific attentional orienting and allocation, (2) the SPCN (post-N2pc), which has been associated with sustained visual processing, (3) the contralateral reduction in oscillatory alpha power (CRA >200 ms), which has also been independently related to attentionally sustained visual processing. Each of these visuo-attentional processes were robustly predictive of successful encoding, and, moreover, each enhanced memory independently of the classic, longer-latency, conceptually related, difference-due-to memory (DM) effect. Early latency mid-frontal theta power also promoted successful encoding while additionally mediating the later latency DM effect and its relationship to successful encoding. These findings markedly expand current knowledge by helping to elucidate the intimate relationship between attentional modulations of perceptual processing and effective encoding for later memory retrieval.

2.1.1 Introduction

Every day we are exposed to a wide array of stimuli in our sensory environment, but we successfully encode into memory relatively few of them. The stimuli we later remember are usually among the ones we attended to during encoding, but attention is not a single process. Rather, it is a complex set of operations, and we know very little about the neural mechanisms whereby each of these attentional operations influences successful encoding. The most powerful method to examine the neural mechanisms of successful long-term memory encoding is the *subsequent-memory paradigm* (Paller, Kutas, & Mayes, 1987) (reviewed in: (Paller & Wagner, 2002)), in which stimulus responses at the study phase are classified as being later remembered or forgotten based on a subsequent-memory test. In subsequent-memory EEG studies, the most investigated activity marker is a broadly distributed, long-latency (400-800 ms) positive-polarity ERP wave known as the *difference-due-to-memory* or *DM* (Paller et al., 1987). Although the DM can be modulated by some types of attention (e.g., attention to meaning), other attentional processes onset well before 400 ms, and we know little about how they contribute to later memory. The current study takes advantage of the high temporal resolution of EEG in a modified visual search task to examine the impact these rapid visuo-attentional effects have on encoding processes and later memory retrieval.

Visual search tasks require identifying a target item amongst distractors. These types of tasks evoke a cascade of EEG components that are modulated by attentional

focus and demands, including the N2pc and the contralateral reduction in alpha power (8-12 Hz). (1) The *N2pc* (negative 200ms posterior-contralateral)(Luck & Hillyard, 1994) has been related to stimulus-specific attentional allocation (reviewed in: (Kiss et al., 2008)) and capture (Qi et al., 2013). (2) Following the N2pc, the *SPCN* (sustained posterior-contralateral negativity; >300ms) has been linked to sustained attentional processing for item-feature discrimination (Jolicoeur et al., 2008), and is generally larger when the target discrimination task is more difficult. (3) Finally, temporally co-occurring with the SPCN (>300ms), there is a post-stimulus reduction in oscillatory alpha power contralateral to the target, which we call here *contralateral reduction in alpha* (CRA). As with most other decreases of alpha, this effect is indicative of increased cortical activity e.g. (Laufs et al., 2003). The CRA has also been linked to longer-latency item-analysis processes (van Diepen et al., 2016), which are cognitively distinct from the SPCN (Bae & Luck, 2017; De Vries et al., 2017; Fukuda et al., 2015). The *onsets* of these three EEG components, the N2pc, the SPCN, and the CRA, occur well before the DM effect; how they modulate subsequent-memory, however, is unknown.

Given that these three components have been associated with enhanced perceptual processes, we hypothesized that *the early attentional processes these components index would predict subsequent-memory (Hypothesis 1)*. The relation between these components and the DM, however, is less clear. The DM is typically found during deep encoding tasks and an abundance of evidence suggests that this component is related to

conceptual processing of the stimuli (reviewed in: (Paller & Wagner, 2002)). On the one hand, enhanced early perceptual processes may lead to enhanced subsequent-memory by boosting later conceptual-level processing. If so, *the early attentional processes indexed by the N2pc, SPCN, and CRA should contribute to the DM (Hypothesis 2a)*. On the other hand, perceptual processes could enhance later memory success in parallel, relatively independently of conceptual processing. Accordingly, *the early attentional processes indexed by N2pc, SPCN, and CRA may not contribute to the DM (Hypothesis 2b)*. Finally, recent evidence has linked successful encoding to an increase in mid-frontal theta power, both pre- and post-stimulus (reviewed in: (Hsieh & Ranganath, 2014)), potentially reflecting interactions between prefrontal cortex and the hippocampus that promote later memory (Backus, Schoffelen, Szabenyi, Hanslmayr, & Doeller, 2016). We hypothesized that these interactions might enhance later conceptual processing such that *mid-frontal theta would be associated with the longer-latency DM effect, and thereby to memory encoding (Hypothesis 3)*.

2.1.2 Materials and methods

2.1.2.1 Participants

Forty-two subjects participated in the study. Six of these were excluded due to low memory performance ($d' < 0.35$ at memory retrieval for both all and high-confidence responses), five for excessive lateral eye-movements (see preprocessing), three for poor data quality, and one for post-experimental disclosure of Tourette's syndrome. Thus,

there were 27 subjects included in the final analysis (age = 22.1 (\pm 2.7) years old [15 males, 12 females]). All subjects were right handed and had normal or corrected-to-normal vision, and all gave written informed consent. The study was conducted in accordance with a protocol approved by the Duke Institutional Review Board. Participants were paid \$15/hour or received class credit from the Psychology Department for their participation in the study.

2.1.2.2 Paradigm

Participants engaged in a modified visual-search style paradigm in a memory encoding phase, followed by a retrieval test phase (Figure 1). During the encoding and retrieval phases, participants were seated approximately 60cm from a 24-inch LCD monitor and the stimuli were presented using the Presentation software package (Neurobehavioral Systems, Albany, CA).

During the encoding phase (Figure 1A), participants were briefly presented with two circumscribed objects (each object's size approximately 3.2 degrees; each ring size approximately 4.5 degrees), one to the left and one to the right of fixation (object center 2.6 degrees to each side). One of the objects was circumscribed in gold (218, 165, 32) and the other in pink (255, 51, 255), with these colors being chosen due to their minimal overlap with the object colors. Participants were instructed to covertly shift their attention (i.e., while maintaining fixation) to the object (*target object*) that was indicated by the target-designating color of the surrounding ring (e.g., circumscribed in pink), and

to indicate—as quickly as possible—whether the object was known or unknown. The target color was fixed for the entire encoding phase, counterbalanced across subjects. *Known objects* (80% of items) were everyday objects participants could identify and name (e.g, chair), whereas *unknown objects* (20% of items) were rare objects participants could not generally identify or name (e.g., a glass insulator). Before encoding, participants practiced the encoding task to ensure they understood the instructions and were able to minimize eye movements, which were tracked with bipolar HEOG. The objects were presented for 400ms and followed an intertrial interval (ITI) of 1600-2000ms (uniform distribution), during which the fixation cross remained on the screen. In total, 640 image-pair trials were presented across 16 blocks, consisting of 512 (80%) familiar object targets and 128 (20%) unknown object targets.

The retrieval phase (Figure 3A) immediately followed the encoding phase. In each retrieval trial, a single object was presented in the midline just beneath fixation for 1000 ms, and participants indicated whether the object was definitely old, probably old, probably new, or definitely new. The ITI during retrieval was 2100-2500 ms (uniform distribution) and 768 trials were presented across 16 blocks, consisting of 512 (67%) old and 256 (33%) new items.

This visual-search task was designed to be a deep-encoding task that forced subjects to conceptually identify the target stimulus (e.g. dog), with unknown items being included to serve as *catch* items to ensure that the subjects were engaged in the

task. As such, during the following retrieval phase, when memory was tested (Figure 3B), we only accessed memory for *known* items that were targets and were correctly identified as being known during the encoding phase, to be sure that these items were attentionally attended to, as well as accurately identified conceptually. All electrophysiological data reported here are from the encoding phase.

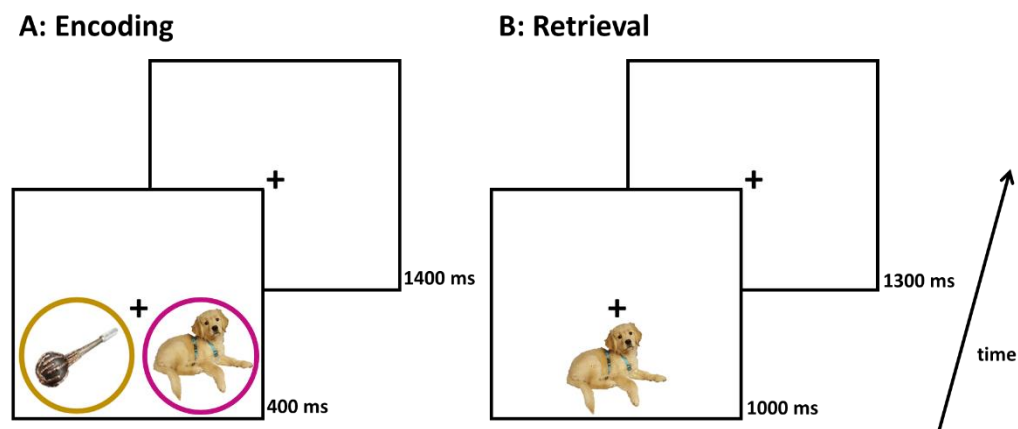


Figure 3. Examples of encoding and retrieval trials. (A) For the encoding phase, subjects were instructed to covertly shift their attention to the lateralized object circumscribed by a ring of a designated color (e.g., pink, counterbalanced across subjects) and to determine if this target object was known or unknown. Known objects could be identified and named, whereas unknown objects could not be. (B) During retrieval, subjects were presented with an old or new object in the midline (just below fixation) and asked to determine if the object was definitely old, probably old, probably new, or definitely new. All the electrophysiological data reported here are from the encoding phase (A).

2.1.2.3 EEG Processing

EEG was DC-recorded (500 Hz sampling rate – three-stage cascaded integrator low-pass comb filter [CIC-filter] with a corner frequency of 130 Hz), using a 64-channel, custom-layout, equidistant, extended-coverage electrode cap (Woldorff et al., 2002) and a Brainproducts Actichamp amplifier with active electrodes (Acticap). The data were

recorded referenced to the right mastoid. Data were also filtered offline using a 0.05 Hz highpass causal FIR filter and a 70 Hz low-pass filter. Data were then downsampled to 250 Hz in conjunction with application of a corresponding anti-aliasing filter.

Independent component analysis (ICA) was used to correct for eyeblinks, which were extracted using the extended infomax algorithm as implemented in EEGLab (using `pop_runica`, EEGLab13.4.4b; (Delorme & Makeig, 2004)). Independent components (ICs) that reflected eyeblinks (1 or 2 ICs per participant) were removed from the data. Finally, trials were epoched [-1000, 1800ms], re-referenced to the algebraically averaged mastoids, and baselined from -200 to 0 ms. Residual eye blinks were detected [-500, 1000ms] using a step response test [-100, 100uV], with trials that surpassed this threshold being rejected.

Lateral eye movements were detected with a unit step function convolutional filter (full length = 200ms) applied to the lateral eye electrodes left-minus-right difference wave, from 0 to 500ms. Trials with eye movements greater than 1.5 degrees (corresponding to ~24uV)(Luck, 2005) during object presentation were flagged for rejection due to fears that they might contaminate the N2pc analysis. Subjects with fewer than 200 remaining trials (total across all conditions) after artifact rejection (N=5) were removed from the analysis.

To optimally examine the posteriorly distributed, attention-related ERPs, a current-source-density (CSD) analysis was applied, which entails calculating the second-

spatial derivative of the EEG for each trial at each time point (ft_scalpcurrentdensity (default usage -> spherical spline)(Oostenveld, Fries, Maris, & Schoffelen, 2011). This processing, which derives the current sources and sinks at the scalp/skull, emphasizes the nearer cortical sources, which has an advantage when examining components such as the posteriorly distributed N2pc when there could be some residual eye-movement from frontal channels, as the CSD derivation makes it much less likely for activity from such frontal sources to propagate to the posterior channels measuring the N2pc.

Frequency information was computed from the CSD-transformed data, after the average ERP for each condition was subtracted from each trial. Frequency decomposition was performed on the EEG data using multitaper methods as implemented in the analysis software package Fieldtrip (Oostenveld et al., 2011), in which discrete prolate slepian sequences were used to estimate the power in logarithmically spaced frequencies from 2 to 30Hz. The window widths for the tapers were 2 cycles for 2-4Hz, 3 cycles for 4-7Hz, 5 cycles for 8-14Hz, 7 cycles for 15-20Hz, 10 cycles for 21-30Hz. Smoothing by means of multitapers was specified as $5 \times \log_{10}$ of each frequency. \log_{10} transformed, time-locked power spectra for each subject were subsequently binned and averaged according to the various conditions. No baseline normalization was performed.

The EEG cap utilized a slightly modified 10-10, equidistant-electrode montage with extended inferior occipital coverage (Woldorff et al., 2002). When reporting

electrode sites, we report standard 10-10 electrode names *if* the electrode in our montage is within 5mm of the corresponding 10-10 site. For those electrodes further than this (5-10mm), the name had a prime appended to it (e.g. O1'). No electrodes were further than about 1 cm from their closest standard 10-10 site.

2.1.2.4 EEG Statistical Analyses

Most of the ERP and oscillatory components under investigation in this study have been well characterized in the attention and memory literature. This prior knowledge was used to guide our statistical analysis. For all statistical tests of functional contrasts, except for the DM, electrode clusters were chosen via visual inspection of waveforms generated from the trials collapsed across all conditions, within the context of previous literature. For the DM, we selected a cluster centered on electrode Cz (Paller & Wagner, 2002). When determining temporal windows to test for significance, we generally relied upon the waveforms generated from *all* trials. For the N2pc, we identified the peak deflection of the grand average across all conditions and centered a 100ms window to test for modulation based upon subsequent-memory. For the CRA, we did the same, but utilized a window of 500ms to account for its longer duration and latency. For the SPCN, we tested a window following the N2pc and lasting until the SPCN returned to baseline (i.e., until there was no difference between activity over the contralateral and ipsilateral hemispheres, again collapsed across all conditions). For the DM, we relied upon prior literature, selecting an epoch between 400 to 800ms.

Mnemonic theta-related effects have been reported both pre- and post- stimulus (reviewed in: (Hsieh & Ranganath, 2014)), and thus we examined three temporal epochs [pre-stimulus: -500 to 0ms; early: 0 to 500ms; and late 500 to 1000ms] while using a Bonferonni correction for multiple comparisons $p < 0.017$. When plotting ERP timecourses, a 20Hz, butterworth 4th order filter was applied to the grand average; no statistics, however, were run on this smoothed data. When reporting numerical results we provide the mean and standard deviation (in parentheses) in addition to t-values (from paired t-tests), p-values, and temporal epochs.

In addition to these *a priori* defined tests, we also performed running t-tests, for all memory contrasts, in order to identify significant time epochs. For this purpose, the ERP data were segmented into 50-ms bins, and t-tests were run on each bin with a critical threshold of $p < 0.05$. For both the ERP and oscillatory data, statistically significant sequences of at least three of such 50-ms time bins were considered meaningful and are marked in gray on all plots. As expected, these analyses corroborated our *a priori* planned statistical test.

A mediation model was run with the toolbox from Tor Wager and colleagues (Wager, Davidson, Hughes, Lindquist, & Ochsner, 2008; Wager et al., 2009). This trial-to-trial, subject-wise multi-level model included all remembered and forgotten trials. In general, mediation models rely upon a regression framework to determine how two correlated variables (e.g., A, B) contribute to a common outcome (C), when they both

predict that outcome. If there is no mediation, then A and B are said to independently contribute to C. If there is a mediation, it suggests that A predicts B (or B predicts A), which then predicts C. In our study, A occurred temporally prior to B and thus testing the alternative path (B predicts A) was nonsensical. This mediation link is known as an indirect path and suggests that part of the way in which A predicts C is determined by its impact on B.

2.1.3 Results

2.1.3.1 Behavioral

Subjects' behavioral performance is summarized in Table 1. Retrieval performance had a mean d' of 0.82 (+/-0.30) across all trials. Response times (RTs) during encoding did not differ for subsequently remembered and forgotten items ($p = 0.67$), but during retrieval, they were slower for forgotten than remembered items [$t(26) = -3.29$, $p = 0.0029$]. To maximize trial counts for the electrophysiological results at encoding, the data were collapsed across high and low confidence for the later memory test responses. This resulted in subjects having at encoding an average of 221 (+/-55) later remembered trials and 120 (+/-36) later forgotten trials. Subjects also had significantly slower RTs for *known* versus *unknown* items at encoding [$t(26) = -10.52$, $p = 7.3e-11$] and worse performance [$t(26) = -3.99$, $p=0.00047$] when responding to catch trials (*unknown* objects) during encoding (i.e., when they incorrectly responded that *unknown* items were *known*), as would be expected given the rarity of the catch items. These behavioral differences

between known and unknown objects did not affect the EEG results, which were analyzed only for known objects.

Table 1. Behavioral Results. Summary of behavioral results. Note that Remembered and Forgotten items encompassed all responses to known objects only. Reported values are means and standard deviations.

	Remembered	Forgotten
All trials (%)	65 (8)	35 (8)
Encoding RT	690 (106)	690 (110)
Retrieval RT	905 (123)	980 (164)

2.1.3.2 ERP DM

As noted in the Introduction, the DM effect is a long-latency (400 to 800ms) positivity over central scalp regions (Paller & Wagner, 2002). Consistent with this literature, during a 300 to 850ms window, we found a robust DM [$t(26) = 3.05$, $p = 0.0052$, {400 to 800ms}] at the vertex and surrounding electrodes (Cz, FCz, CPz, FC1, FC2, C1a, C2a) (Figure 4).

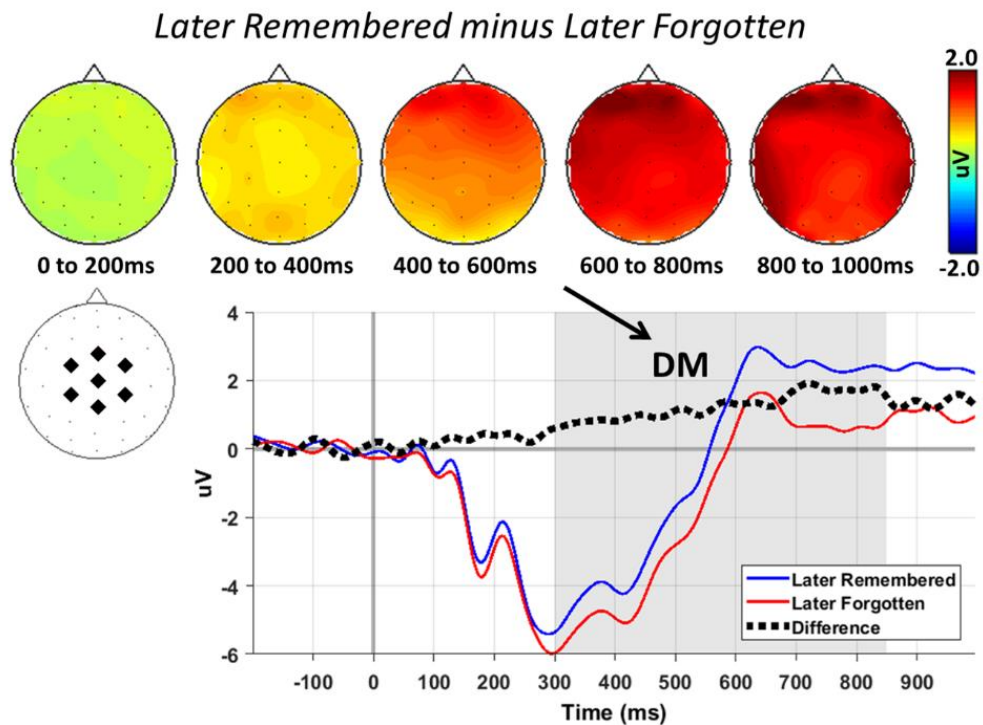


Figure 4. DM effect in central electrodes during. Running t-tests identified an epoch from 300 to 850ms (shaded gray) where subsequently remembered items had a greater positivity than subsequently forgotten items.

2.1.3.3 Lateralized ERP attention effects at encoding: N2pc & SPCN

Lateralized ERP attention effects at encoding were computed by directly subtracting the ERPs at the contralateral channels (relative to the relevant target object) minus the ERPs at the corresponding ipsilateral channels (contralateral *minus* ipsilateral i.e., c-i). This operation effectively provides twice the number of trial counts, as left target responses (right-hemisphere electrodes minus left-hemisphere responses) are combined with right target responses (left hemisphere electrodes minus right-hemisphere electrodes) and is very commonly utilized when computing the lateralized

attention-related N2pc and SPCN components. A visual analysis of *all* trials indicated that the N2pc peaked ~260 ms (Figure 5; black electrodes: O1', P3i). This activity peak at encoding, analyzed from 210-310ms (+/- 50 ms around the peak), showed a significant difference for later remembered *minus* later forgotten trials [$t(26) = -3.39$, $p = 0.0022$].

Following the N2pc, we identified a subsequent sustained negativity from 400 to 800ms [$t(26) = -3.25$, $p=0.0032$] reflecting an SPCN-like component. The end duration of this post- N2pc lateralized negativity was set to coincide with the time point at which *all* known-object trials no longer showed a contralateral vs. ipsilateral difference. It is also important to note that this negativity did not separate from the N2pc in time, with only a small dip in magnitude between the N2pc and this sustained negativity (i.e., the N2pc negativity here seemed to progress directly into the SPCN negativity). However, this sustained negativity extended far beyond what would be considered an N2pc, which is thought to reflect the initial orienting of attention toward a visual target in a visual search paradigm. Rather, its extended duration seems likely to reflect the continued processing of the lateralized target stimulus, similar to the classic SPCN. Accordingly, we will refer to this long-latency, sustained negativity, effect as the SPCN, while also providing alternative perspectives in the discussion.

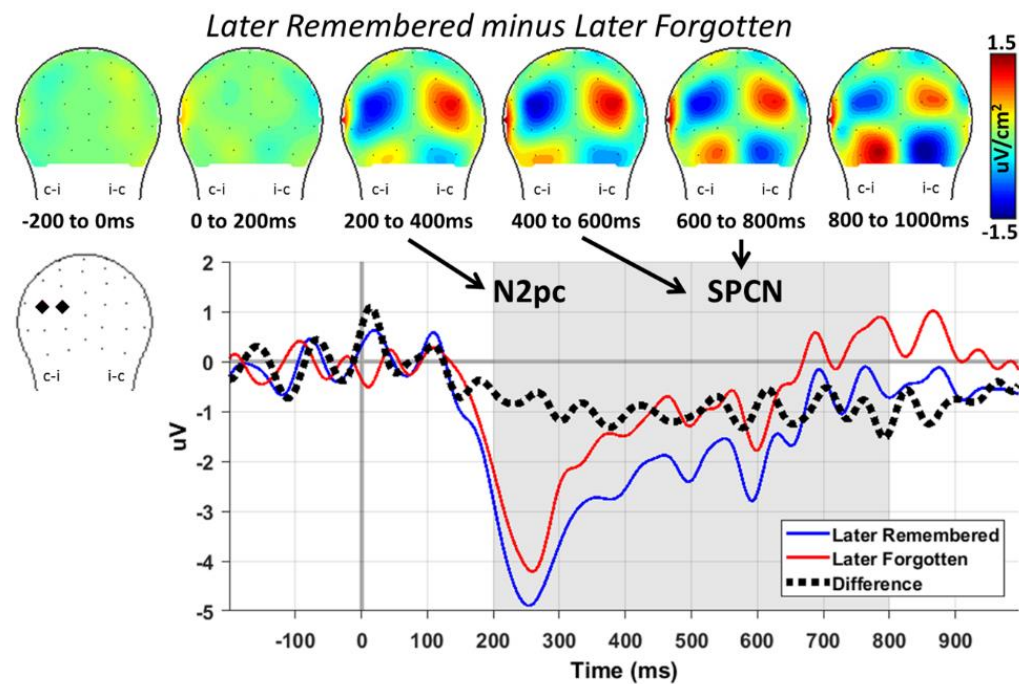


Figure 5. Both the N2pc and the SPCN were larger for subsequently remembered than forgotten items. Running *t*-tests identified an epoch from 200 to 800ms (shaded gray) where subsequently remembered items had a greater contralateral negativity than subsequently forgotten items.

2.1.3.4 Oscillatory Contralateral Reduction in Alpha (CRA) [8-12 Hz]

A robust CRA effect at encoding was identified for all known-object items (Figure ; black electrodes: O1', P3i, TO1), peaking at about 500ms. As predicted, the magnitude of this effect varied as a function of subsequent-memory [$t(26) = -2.76$, $p=0.011$, {250 to 750ms}]. Running *t*-tests identified an early epoch (200 to 500ms) and a later epoch (700 to 1000ms) where *decreased* contralateral alpha power was associated with better memory.

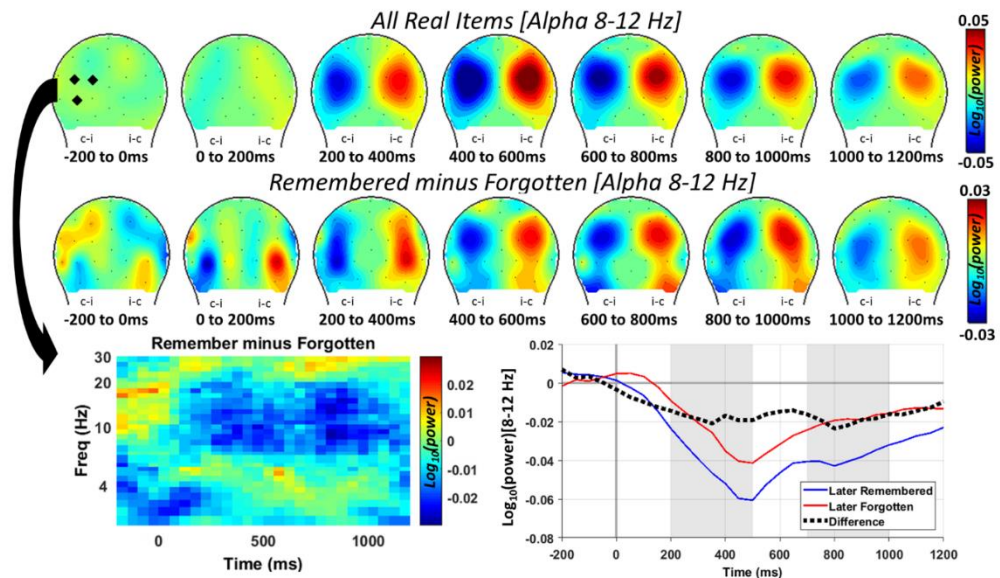


Figure 2. Differences in the contralateral reduction of alpha (CRA), reflecting increased cortical activity, were prominent both early in the trial (200 to 500ms) and later in the trial (700 to 1000ms).

2.1.3.5 Oscillatory Theta [4-7 Hz]

As illustrated by Figure 6, theta power over mid-frontal (FCz, CPz, FC1, FC2, C1a, C2a) electrodes (mid-frontal theta) was significantly higher for later remembered than for later forgotten items, pre-stimulus [$t(26)=3.47$, $p=0.0018$, $\{-500$ to $0\text{ms}\}$], early [$t(26)=3.45$, $p=0.0020$, $\{0$ to $500\text{ms}\}$], and late [$t(26)=3.35$, $p=0.0025$, $\{500$ to $1000\text{ms}\}$].

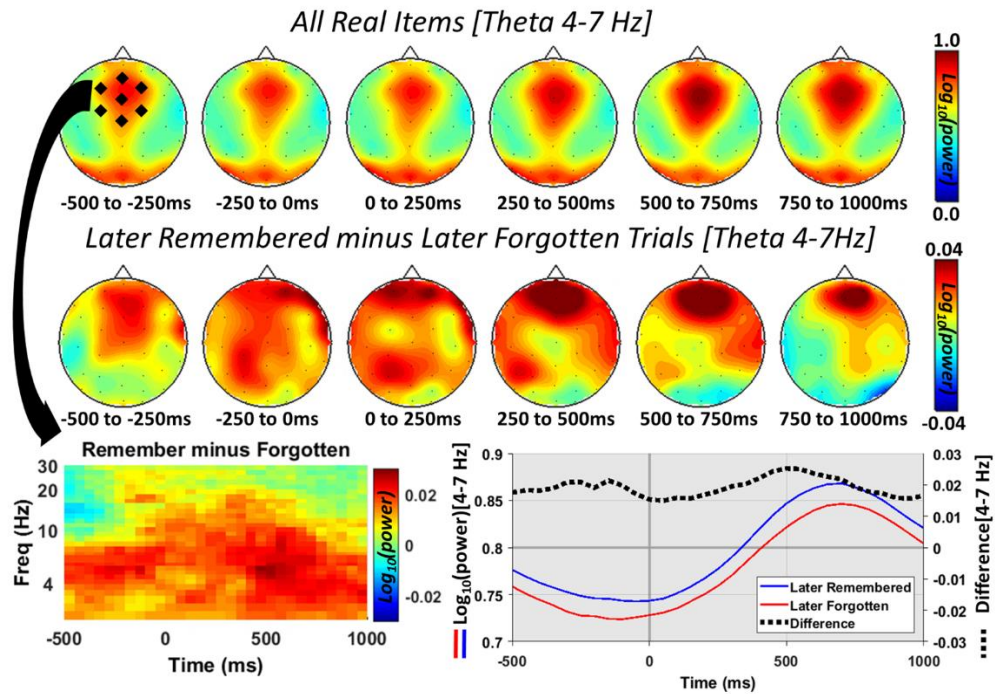


Figure 6. Mid-frontal theta was greater for subsequently remembered than forgotten items (all $p < 0.05$).

2.1.3.6 Independent and dependent contributions to memory

In sum, our results linked subsequent-memory not only to the traditional DM effect, but also to several rapid-onset, more spatially specific, attention-related EEG components including the N2pc and CRA. We also identified a sustained negativity following the N2pc that also predicts subsequent-memory and appears temporally and spatially similar to an SPCN. Finally, we also found that mid-frontal theta was greater for subsequently remembered than forgotten trials, even before the trial began. Having successfully identified this set of effects, we next examined interactions among them, based upon what we directly observed (i.e., not a priori hypothesized), to determine if

their contributions to memory were dependent or independent of each other. Dependent processes should be correlated with each other, whereas independent processes would be expected to be uncorrelated. We computed, for each subject, the Spearman correlation across trials, between the following eight EEG measures: N2pc (210 to 310ms), the SPCN (400 to 800ms), the early (200 to 500ms) and late (700 to 1000ms) CRA, the superior-central positivity underlying the DM (300 to 850ms), and the mid-frontal theta in the three analyzed epochs of pre-stimulus (-500 to 0ms), early (0 to 500ms), and late (500 to 1000ms). For each of these eight EEG measures, the trial-wise value included in the correlation analysis was based upon the average magnitude within the indicated temporal epoch. For these analyses we utilized a Bonferonni correction for multiple comparisons ($N = (8^2 - 8) / 2 = 28$, critical $p < 0.05/28 = 0.0018$).

The correlations yielded three key findings. First, neither the N2pc, the subsequent sustained SPCN-like negativity, nor the CRA were correlated with the superior positivity of the DM, suggesting they contribute independently to successful encoding. Second, in marked contrast to the N2pc, SPCN and CRA, early theta was strongly correlated with the DM [$\text{atanh}(r) = 0.074$ (0.056), $t(26) = 6.83$, $p = 2.97e-7$], suggesting the former contributes to the latter. Furthermore, the magnitude of this effect (DM to early theta) was far stronger than any possible relationships between other conceptually related processing (DM, theta) and attention-related processing (SPCN, N2pc, CRA; all $p < 0.0001$). Third, there was no significant correlation between the CRA

and SPCN, further suggesting that although these processes generically support sustained visual processing and overlap in time, they do so in different ways. Additionally, we found correlations between the similarly distributed N2pc and SPCN, as well as between the three theta epochs (all $p < 0.0018$); these correlations would seem unsurprising, however, given the spatial correspondence and temporal adjacency of these effects.

To clarify the significant correlation between the mid-frontal theta and the superior-central positivity of the DM, we ran an exploratory multi-level mediation model (Wager et al., 2008, 2009). In this analysis, we tested whether the magnitude of the superior-central positivity underlying the DM (600 to 800ms) mediated the relationship between early theta power (125 to 225ms) and subsequent-memory. These temporal epochs were chosen such that they were completely non-overlapping (given that the theta kernel is 750ms) and within the bounds in which a significant correlation was identified: theta (0 to 500ms) and superior-central positivity associated with the DM (300 to 850ms). The results of this model showed a partial mediation ($p = 0.0031$, proportion mediated about 7%; Figure 7), suggesting that part of influence of the theta-indexed process on subsequent-memory is driven in part by its relationship with the superior-central positivity underlying the DM.

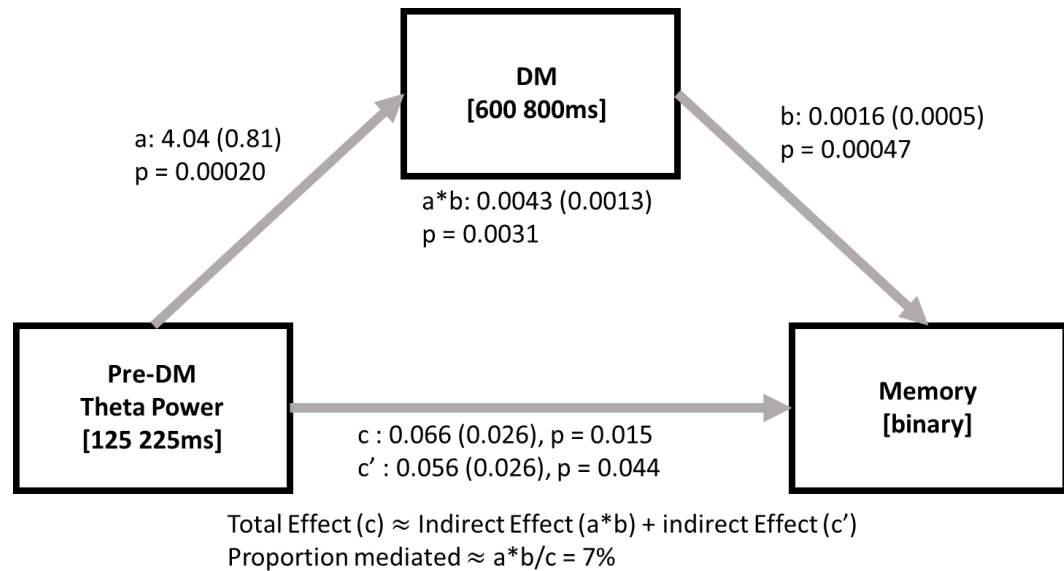


Figure 7. A multi-level mediation model showed that superior-central positivity of the DM effect partially mediated the relationship between pre-DM theta power and subsequent-memory.

2.1.4 Discussion

The contribution of rapid visuo-attentional processes to successful memory encoding is largely unknown as many paradigms and analyses focus solely on longer-latency mechanisms, such as the role of conceptual processing as indexed by the DM. By utilizing a novel spatial-focusing paradigm, we investigated the mnemonic contributions of a rapid cascade of visuo-attentional processes indexed by the N2pc (attentional orienting and allocation), SPCN (item-feature identification), and CRA (item-analysis processes) and their relationship to longer-latency conceptual processing as indexed by the DM. We additionally examined the role of the level of mid-frontal theta for subsequent-memory and, in an intermediating way, the DM effect. The study

yielded three main findings: (1) Each of the EEG markers of the rapid-onset attentional modulations of visual processing - the N2pc, SPCN and CRA -- predicted better subsequent-memory; (2) the N2pc, SPCN, and CRA visual-attention effects led to enhanced memory independently of the DM; (3) finally, mid-frontal theta power predicted both the DM effect and subsequent-memory. These findings are summarized in Figure 8.

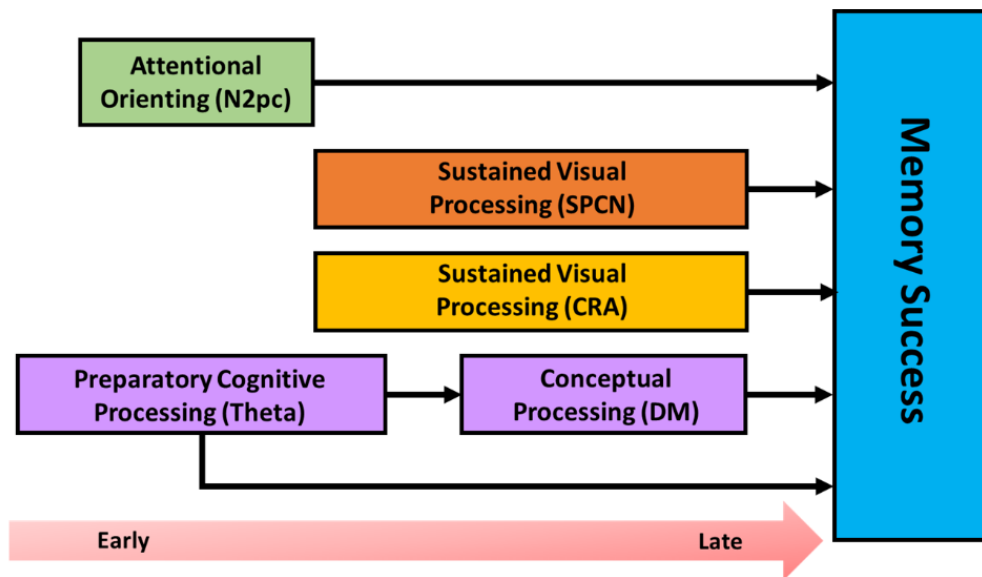


Figure 8. EEG effects enhancing subsequent-memory. The schematic figure displays a set of parallel processes that all serve to independently contribute to successful memory encoding including attentional orienting and allocation as indexed by the N2pc ERP (green), sustained item-specific visual processing as indexed by the SPCN ERP (orange), sustained visual processing as indexed by oscillatory alpha-band CRA (yellow; CRA), and preparatory cognitive-control processing followed by conceptual processing, indexed by theta-band EEG and the DM (purple).

2.1.4.1 Rapid-onset attentional processes as indexed by the N2pc, SPCN, and CRA predict subsequent-memory

Consistent with our first hypothesis, we found that markers of early attention processes, including the N2pc, SPCN and CRA, all predicted subsequent-memory, and all independently of the longer-latency DM. We discuss our interpretation of each of these markers below.

N2pc. The cascade of attentional processes that are being tracked here and facilitate memory encoding are initiated with attentional allocation towards the target item that is both spatially and stimulus specific. This shift is indexed by the N2pc (starting ~210ms), which was larger for subsequently remembered items. One potential explanation for why the magnitude of the N2pc predicted subsequent-memory is that the strength of the focusing of attention toward the relevant target stimulus will naturally fluctuate from trial to trial, due to factors such as randomly varying levels of task focus. That is, trials with strong attentional orienting and focusing, as indexed by the elicitation of larger N2pc's, lead to stronger perceptual processing and thus encoding of the stimulus, and thereby to better subsequent-memory. Another potential explanation relies upon attentional bias, as larger N2pcs have also been found for more rewarding stimuli (Donohue et al., 2013) and more fear-provoking stimuli (Eimer, & Kiss, 2007) as compared to neutral stimuli. In the current experiment, in which random objects were presented to the participants, it is possible that they more strongly oriented

to stimuli that were both targets and personally meaningful or salient – leading to an increased N2pc and enhanced subsequent-memory.

SPCN. The SPCN sustained negativity follows the N2pc, typically showing a similar scalp distribution, and tends to be larger for more difficult-to-discriminate stimuli (Jolicoeur et al., 2008). As hypothesized, we observed a sustained negativity attentional modulation following the N2pc, which also predicted subsequent-memory, suggesting that stimuli that were more difficult to discriminate, or might otherwise invoke more extended discrimination processes, might be more strongly encoded. Another potential explanation, however, might relate to a spatially similar slow-wave component, the *contra-lateral delay activity* (CDA). This component has been associated with increased working memory load (Luria, Balaban, Awh, & Vogel, 2016) in that it gradually increases in amplitude (with load) and is sustained until the information in working memory is discarded. As such, another interpretation is that the SPCN effect reflects extended manipulation of the item in working memory, which is also known to increase long-term memory retrieval (Khader, Ranganath, Seemüller, & Rösler, 2007).

CRA. Another marker of increased, spatially specific, visual processing that has been reported more recently is a *contralateral reduction in alpha* (CRA) power, which we also find to be linked to improved subsequent-memory. In general, reductions in alpha activity have been associated with increased cortical activity (Laufs et al., 2003). While proactive lateralized alpha in a spatial-attention cueing task has been well characterized

(Worden, Foxe, Wang, & Simpson, 2000), and even dissociated from lateralized slow-wave ERPs (Grent-'T-Jong, Boehler, Kenemans, & Woldorff, 2011), the function of lateralized *reactively* triggered alpha, and its relationship to ERPs is less clear. Analogous to proactive lateralized processes, preliminary evidence suggests that reactively triggered alpha (CRA) and lateralized slow-wave ERPs (SPCN), though both related to sustained visual processing, serve independent functions (Bae & Luck, 2017; De Vries et al., 2017; Fukuda et al., 2015), consistent with our finding that these measures were uncorrelated. With respect to visual search, one recent interpretation is that the CRA serves to “protect the sensory processing of stimuli” (Stankevich & Geng, 2014) from other distracting stimuli, and thus the early-latency CRA observed here might reflect this type of protection, which then contributes to better encoding. The CRA might also reflect working-memory-related processes, as the CRA magnitude has also been found to be modulated as a function of working memory load (Fukuda et al., 2015) and thus the enhanced CRA observed here might reflect working-memory-related processes that contribute to better encoding.

2.1.4.2 The early attentional processes indexed by N2pc, SPCN, and CRA do not contribute to the DM

Having found that the rapid cascade of early attentional effects indexed by the N2pc, SPCN, and CRA predicted subsequent-memory, we examined whether any of these early attention effects correlated with the hallmark long-latency DM effect.

Consistent with Hypothesis 2b, none of these early attention effects correlated with the

DM, suggesting that their contributions to subsequent-memory occur in parallel to or addition to the DM, and, moreover, such contributions are not mediated by the DM. Given that the DM has been linked to conceptual processing, whereas these early attention effects have been associated primarily with perceptual processes, the most parsimonious account is that perceptual and conceptual processes, including their modulation by attention, can make independent contributions to subsequent-memory performance in a recognition memory test.

2.1.4.3 The magnitude of the mid-frontal theta was associated with the DM effect

Finally, consistent with our third hypothesis, we found that the magnitude of early-latency mid-frontal theta (125-225ms) predicted not only subsequent-memory, but its relationship with subsequent-memory was partially mediated by the long-latency DM effect (600-800 ms). Post-stimulus frontal midline theta has previously been linked to memory encoding (reviewed in: (Hsieh & Ranganath, 2014)) and to cognitive control in tasks where there is need for top-down control (reviewed in: (Cavanagh & Frank, 2014)), both of these factors are likely relevant for our results. With regards to cognitive control, our task was rather difficult; objects were only presented for 300ms, and unknown items were, on average, successfully identified only 62% of the time, with known items being correctly identified 84% of the time. Thus, early-latency top-down cognitive control was likely essential to completing the object-discrimination task. The indirect effect of the early theta power (c), as reflected by a partial mediation of the DM,

suggests that this realized control may causally contribute to later-latency conceptual processing, which then in turn leads to stronger encoding. With regards to memory, theta also had direct (c') effects, suggesting a primary role as well. While the neuroanatomical geometry of the hippocampus renders it largely undetectable with scalp EEG (Nunez & Srinivasan, 2006), a few studies have reported that increased mid-frontal theta power is associated with prefrontal cortex and hippocampal connectivity, which is known to facilitate encoding (Backus et al., 2016; Garrido, Barnes, Kumaran, Maguire, & Dolan, 2015), and thus the direct effect might be reflective of these types of interactions.

2.1.5 Conclusion

In sum, our results indicate that, when processing a stimulus or event in our environment, a cascade of rapid-onset attentional processes operate in concert that can facilitate successful memory encoding of those stimuli, mostly independently of, and on-setting earlier than, conceptual-level processing that has been indexed by the longer-latency DM effect. This attentional cascade is initiated by increased stimulus-specific attentional orienting and allocation (N2pc) and is followed by sustained visual processing (SPCN and CRA) likely reflecting item-specific discrimination processes and/or working memory processes that also contribute to successful memory encoding. In parallel to these visuo-attentional processes, increased preparatory cognitive control processes, as indexed by early midline frontal theta, appears to lead to enhanced longer-

latency conceptual processing (DM). These parallel item-focused visuo-attentional processes and enhanced conceptual processes all appear to then serve to strengthen memory encoding and lead to better subsequent-memory retrieval.

3. Hippocampal Networks

3.1 Hippocampal contributions to the large-scale episodic memory network predict vivid visual memories

A common approach in memory research is to isolate the function(s) of individual brain regions, such as the hippocampus, without addressing how those regions interact with the larger network. To investigate the properties of the hippocampus embedded within large-scale networks, we used functional MRI and graph theory to characterize complex hippocampal interactions during the active retrieval of vivid vs. dim visual memories. The study yielded four main findings. First, the right hippocampus displayed greater communication efficiency with the network (shorter path length) and became a more convergent structure for information integration (higher centrality measures) for vivid than dim memories. Second, vivid minus dim differences in our graph theory measures of interest were greater in magnitude for the right hippocampus than for any other region in the ninety-region network. Moreover, the right hippocampus significantly reorganized its set of direct connections from dim to vivid memory retrieval. Finally, beyond the hippocampus, communication throughout the whole brain network was more efficient (shorter global path length) for vivid than dim memories. In sum, our findings illustrate how multivariate network analyses can be used to investigate the roles of specific regions within the large-scale network, while also accounting for global network changes.

3.1.1 Introduction

Although the vast majority of studies on the neural bases of episodic memory (declarative memory for personally-experienced, context-specific events) have focused on the contributions of specific brain regions, such as the hippocampus, it is generally accepted that no single region can support episodic memory unless it interacts with other components of the large-scale memory network. The properties of large-scale networks can be investigated by applying multivariate analytic methods, such as graph theory, to brain connectivity data collected from functional MRI (fMRI). Complex network analyses using graph theory provide the means to reliably quantify properties of brain networks with a small number of neurobiologically meaningful measures that capture the interactions between all brain regions simultaneously (Bullmore & Sporns, 2009; Rubinov & Sporns, 2010). Yet, most complex network analyses have focused on global measures of resting-state networks rather than on the contributions of specific regions in task-related networks. The current large-scale network study provides a unique contribution to this literature by focusing primarily on the role of a specific brain region, the hippocampus, within the context of the whole brain network supporting vivid episodic memory retrieval.

The contributions of the hippocampus to episodic memory have been the focus of thousands of lesion, electrophysiology, drug, and neuroimaging studies with both animals and humans. Lesion studies have provided clear evidence for the devastating

effects of hippocampal damage on episodic memory, and functional neuroimaging studies with healthy adults have reliably shown greater hippocampal activity for successful than unsuccessful episodic retrieval and for rich than impoverished episodic memories (Eichenbaum et al., 2007; Kim, 2015; Skinner & Fernandes, 2007; Spaniol et al., 2009). The hippocampus is hypothesized to bind incoming information from different neocortical regions, to store integrated event representations, and to allow access to cortical memory traces during retrieval (Alvarez & Squire, 1994; Danker & Anderson, 2010; McClelland et al., 1995; Ritchey et al., 2013; Teyler & Rudy, 2007). During these processes, the hippocampus is assumed to interact very closely not only with regions storing representations but also with frontal, parietal, cingulate, and basal ganglia regions involved in a variety of attention, control, working memory, and decision making processes. Consistent with these assumptions, functional neuroimaging studies have shown that all these regions are co-activated with the hippocampus during episodic encoding and retrieval tasks (Cabeza et al., 2008; Cabeza & Nyberg, 2000; Kim, 2013; Rugg & Vilberg, 2013; Spaniol et al., 2009). Even though all these regions are assumed to operate together as a complex large-scale network, existing functional connectivity studies of episodic memory have typically focused only on the relationship between a particular pair of regions, such as the interaction between the hippocampus and the prefrontal cortex e.g. (Schott et al., 2013; Wing et al., 2013). Only recently have researchers begun to use task-related functional connectivity analyses to investigate

changes across larger sets of memory related brain regions (King et al., 2015; Schedlbauer et al., 2014).

A powerful approach for investigating the operation of large-scale brain networks, such as the one supporting episodic memory, is to analyze functional connectivity data using multivariate analytic methods, such as graph theory. Most large-scale network analyses have provided valuable insights into the topology of resting-state networks e.g. (Dosenbach et al., 2007; Hayasaka & Laurienti, 2010; Power, Fair, Schlaggar, & Petersen, 2010; Vogel et al., 2013), but those resting-state networks do not necessarily match network architecture during specific cognitive tasks e.g. (Bassett et al., 2010; Braun et al., 2015; Cao et al., 2014; Meunier et al., 2014; Malaak Nasser Moussa et al., 2011, 2014; Rzucidlo, Roseman, Laurienti, & Dagenbach, 2013; Stanley, Dagenbach, Lyday, Burdette, & Laurienti, 2014). The topological properties of brain networks do not remain static and fixed. Depending upon the demands on the system, there are continuously changing patterns of functional interactions between regions, circuits, and systems in the brain (Sporns, 2013). Furthermore, most of these large-scale network analyses have focused on global network properties rather than on the contributions of specific brain regions to the network, and hence, their findings have been somewhat disconnected from the main body of research on the functions of these regions. To address these issues, we applied graph theory measures to functional connectivity data during an episodic memory retrieval task, focusing in particular on how the

contributions of the hippocampus to the large-scale network are related to the vividness of visual memories.

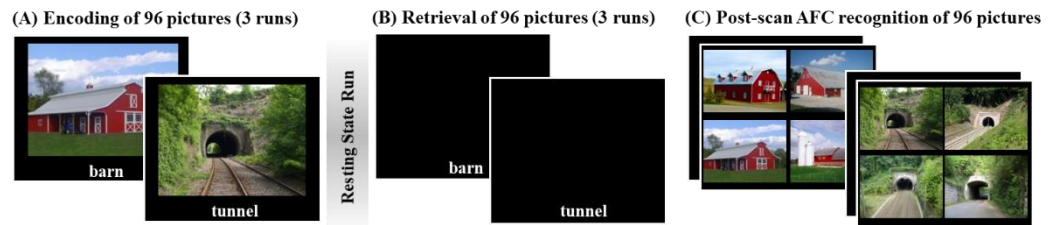


Figure 9. Provides an overview of the experimental design. The same design was presented in Wing et al. (2015). (A) During encoding, pictures of scenes were presented to participants with a descriptive label while participants judged image composition. (B) At retrieval, only the descriptive labels for previously encoded scenes were presented. Participants rated how detailed (vivid) their memory was for the corresponding picture on a 4-point scale. (C) After the scan, all scenes from encoding were presented in a forced-choice recognition task that included three similar scene exemplars. Participants chose the specific image they believed was presented at encoding and then rated their confidence on a 4-point scale.

The behavioral paradigm we investigated, which is depicted in Figure 9, had four phases. First, participants encoded a series of labeled scenes (e.g., barn, tunnel) by rating the representativeness of each photo. Second, participants underwent a resting-state scan. Third, participants recalled the previously-viewed scenes in response to their labels, rating the vividness of their memories (from 1 to 4). Finally, participants performed a forced-choice scene recognition test outside the scanner, in which they discriminated between each viewed scene and three similar distractors. Post-scan recognition accuracy and confidence increased with in-scan vividness ratings, indicating that these ratings provide a valid measure of memory quality. In the current study, we focused on functional connectivity using graph theory measures during scene recall in

order to identify network differences between vivid memories (ratings 3-4) and dim memories (ratings 1-2).

We investigated four main questions. First, how do complex patterns of hippocampal interactions with the rest of the network change for the retrieval of vivid vs. dim memories? To answer this question, we measured vivid minus dim differences in four nodal measures: path length, degree centrality, page rank centrality, and leverage centrality. Path length measures the efficiency with which information can flow between any two nodes in the network and is computed as the average of the shortest paths between a node and every other node in the network. Shorter path lengths promote functional integration by allowing communication between any two network nodes with few intermediate steps, thereby reducing the effects of noise and signal degradation (Rubinov & Sporns, 2010; Sporns, 2013). Centrality measures index the convergence and joint processing of distributed information at central, influential nodes, and can be defined by diverse criteria. Degree centrality identifies nodes with many connections to other nodes in the network. Page rank centrality identifies nodes that are connected to nodes that are themselves central within the network. Finally, leverage centrality identifies nodes that are connected to more nodes than their immediate neighbors. We predicted that the hippocampus would display a shorter path length and become a more central, influential node to support the convergence of information in facilitating vivid memory retrieval, thereby confirming the assumption that

hippocampal contributions to the episodic retrieval network have a significant impact on retrieval success.

Second, how do the changes in hippocampal network properties from dim to vivid memory retrieval compare to the changes in other network nodes? Even if there are stark shifts in hippocampal network properties from dim to vivid memory retrieval, it is possible that many other nodes in the network are changing in more dramatic ways. To address this issue, we computed the aforementioned four nodal measures for all nodes in the network. Memory theories generally assume that the hippocampus is a “bottleneck” or “convergence zone” in the episodic memory network because it integrates information from several different brain regions, each with particular functions related to memory retrieval (Damasio, 1989; Mišić, Goñi, Betzel, Sporns, & McIntosh, 2014; Moll & Miikkulainen, 1997). Thus, we predicted that when comparing dim to vivid memories, the hippocampus would show the most dramatic changes in path length and centrality measures compared to all other regions in the network. Several other graph theory analyses of resting-state connectivity data (Buckner et al., 2009; Tomasi & Volkow, 2010) did not find the hippocampus among the most central, influential network nodes, or among the nodes occupying critical positions along shortest paths in the network. It remains possible, however, that hippocampal function in the context of larger networks is most sensitive to memory differences that are observed within the same memory task, rather than in resting state networks. As such,

we were interested in investigating network properties capturing the *relative shift* --i.e. how the hippocampus relates to the rest of the network—between dim to vivid memory retrieval.

Third, to what extent does the hippocampus reorganize its set of direct (first step) connections from dim to vivid memory retrieval? Even though shifts in multivariate hippocampal network properties can only be explained in full by appealing to the entire network architecture, one can ask whether those shifts are better explained by (1) a substantial reorganization of connectivity strengths for first step (direct) connections or (2) a substantial reorganization of connectivity patterns beyond first step connections (indirect connections). To investigate this question, we calculated the extent to which first step connections reorganize themselves for each and every node between dim and vivid retrieval networks using the novel first step reorganization measure. If, in fact, a network node exhibits a dramatic change in first step reorganization between networks, we can reasonably conclude that stark changes in its set of direct connections are at least partly accounting for observed changes in path length, page rank centrality, and leverage centrality. In contrast, if a network node changes minimally in first step reorganization between networks, then any observed significant changes in path length, page rank centrality, and leverage centrality for that node are primarily due to changes in indirect connectivity taking place elsewhere in the network.

Finally, do properties of the entire episodic retrieval network change from dim to vivid memory retrieval, and are these changes driven by the hippocampus? As noted before, most large-scale network analyses of functional neuroimaging data have focused on global network measures rather than on the operation of individual brain regions. Although these analyses focus on the role of the hippocampus, we believe that investigating the network properties of this region alone is insufficient to fully explain the neural basis of episodic memory. Even if our predictions about the role of the hippocampus within the episodic retrieval networks are confirmed, we assume that many other regions contribute to memory vividness. Thus, to assess global network changes in the efficiency with which information can be integrated in the network, we calculated the normalized average path length for the entire brain network. We predicted that the network as a whole would facilitate more efficient global communication (shorter global path length) for vivid than dim memory retrieval. And even after removing the hippocampus from the network, we predicted that we would find significant global network differences between vivid and dim memories in the remaining nodes.

3.1.2 Methods

3.1.2.1 Participants

Twenty-two participants completed the experiment. One participant, who lacked functional data due to a technical error, was excluded from these analyses. All analyses

were performed with the remaining 21 participants (12 female, age range: 18-30, $M = 23.5$, $SD = 3.0$). Participants were healthy, right-handed, fluent English speakers with normal or corrected-to-normal vision. Written informed consent was obtained from each participant in accordance with a protocol approved by the Duke University Institutional Review Board.

3.1.2.2 Behavioral Methods

The behavioral paradigm is depicted in Figure 9. Before beginning the scan, participants completed a short practice session so that they were familiar with the instructions at each phase of the study. The scan session contained 3 encoding runs, a resting state run, and 3 retrieval runs (analyses comparing spatial patterns of activation across encoding and retrieval were reported in Wing et al. (2015)). Stimuli consisted of the pictures of 96 nameable outdoor and indoor scenes, and the corresponding verbal labels (e.g. “island”, “concert hall”). During three encoding runs, each scene and its label were presented for 4-sec, and participants rated how well the picture matched the label. During three retrieval runs, each label was presented alone. Participants tried to recall the corresponding picture in as much detail as possible and then rated the amount of detail (vividness) in the image they generated from 1 (little or no detail) to 4 (highly detailed). For the purpose of the present network analyses, trials with responses of 1-2 were classified as “dim memory” trials, and trials with responses of 3-4 as “vivid memory” trials. During both encoding and retrieval, the 8-sec inter-trial intervals were

filled by an active baseline task in which participants made even/odd judgments in response to randomly presented digits ranging from 1 to 9.

Immediately after exiting the scanner, participants completed a four-alternative forced-choice recognition test assessing memory for all 96 pictures. Each trial had two phases. First, the target picture and three distractor pictures for the same label were presented in different quadrants of the computer screen. Participants selected the picture they believed they saw in the scanner. During the second phase of each recognition trial, participants used a 4-point scale to rate how confident they were in the preceding recognition decision (1 = guess, 4 = very confident).

3.1.2.3 MRI scanning and image processing

Imaging data were collected using a 3T GE scanner. Following a localizer scan, functional images were acquired using a SENSE spiral-in sequence (TR = 2000ms, TE = 30ms, FOV = 24cm, 34 oblique slices with voxel dimensions of 3.75 x 3.75 x 3.8mm). Functional data were collected during six task runs of equal length during the performance of the memory task described above. A resting state scan lasting for five minutes was collected following the third run but was not used in the current analyses. Stimuli were projected onto a mirror at the back of the scanner bore, and responses were recorded using a 4-button fiber optic response box. Scanner noise was reduced with ear plugs, and head motion was minimized with foam pads. A high-resolution anatomical image (96 axial slices parallel to the AC-PC plane with voxel dimensions of

0.9x0.9x1.9mm) was collected following functional scanning. Our preprocessing procedure mirrored that of Fornito et al. (2011), who have published data on a beta series derived network analysis. Briefly, using SPM8 software (www.fil.ion.ucl.ac.uk/spm) and custom MATLAB scripts, all data were high-pass filtered, motion corrected, and each individual's brain map was registered to standard MNI space.

3.1.3.4 Functional brain network construction

To create episodic retrieval networks, we used a beta time series analysis (Rissman, Gazzaley, & D'Esposito, 2004), which assumes that two regions are functionally coupled during a task if the activity of both regions is significantly correlated across trials (Fornito et al., 2011; Schedlbauer et al., 2014). Each beta value reflected the fit shape of the hemodynamic response evoked by a given trial during the retrieval phase of the procedure. These observed beta values were then sorted in accordance with level of detail scores reported by participants during each trial of the retrieval condition. Beta values obtained during trials for which participants reported high levels of detail with which they could remember the specific picture (ratings of 3 and 4) were concatenated to generate a beta series for networks of vivid memory retrieval; beta values obtained during trials for which participants reported low levels of detail with which they could remember the specific picture (ratings of 1 and 2) were concatenated to generate a beta series for dim memory retrieval.

Networks can be represented as graphs, which consist of a set of nodes together with pairwise relationships between those nodes (edges). In these functional brain networks, each node represented a discrete brain region, and edges represented measured correlations between pairs of nodes. To measure functional connectivity among all regions simultaneously, the brain was first parcellated into 90 discrete anatomical regions of interest (45 ROIs in each hemisphere) defined in accordance with the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). A commonly used nodal parcellation scheme in functional brain network analyses is this AAL template (Stanley et al., 2013), which parcellates the cortex and subcortical structures by identifying gyral and sulcal boundaries. Each anatomical ROI from the atlas served as a network node. Pairwise Pearson correlations between regional mean beta series were computed to generate $\{90 \times 90\}$ functional connectivity matrices, or adjacency matrices, with the correlation coefficients representing functional connectivity strength between nodes. These correlations between regional beta series reflected correlated variations in evoked hemodynamic responses within vivid and dim retrieval networks, respectively. The matrices were not thresholded, and each complete matrix served as an undirected, weighted graph (Rubinov & Sporns, 2010). Adjacency matrices (averaged across subjects for display purposes) for dim and vivid memory retrieval conditions are presented in Figure 10.

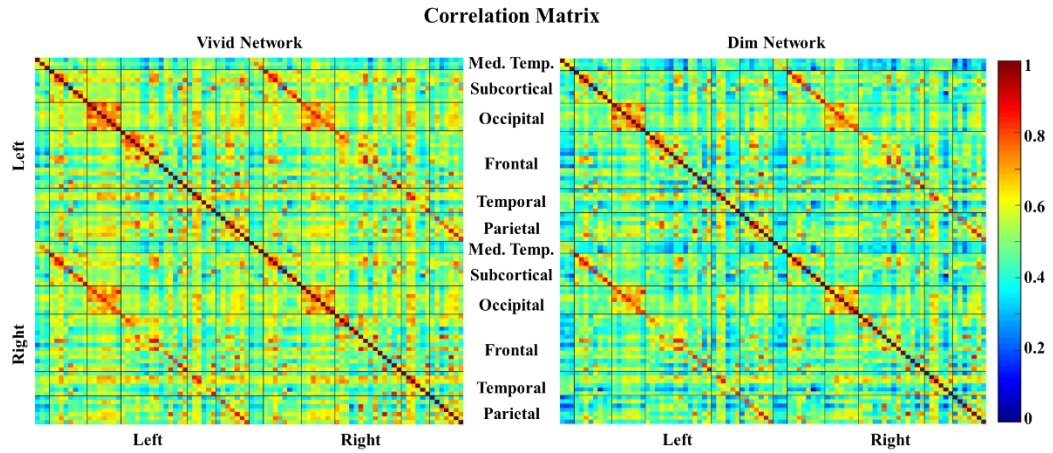


Figure 10. Average (across subjects) adjacency matrices derived from beta series correlations are presented for dim and vivid retrieval conditions and split by hemisphere (right and left). For ease of visualization, regions of interest are ordered in accordance with the procedure implemented by Salvador et al. (2005).

3.1.3.5 Graph Theory Measures

3.1.3.5.1 Path Length

Path length measures the overall capacity for efficient information transfer across a network. From each weighted graph, two global network measures were computed: weighted characteristic path length (L_{net}^w) and the normalized weighted characteristic path length (λ). Both L_{net}^w and λ were calculated individually for each node and subsequently averaged over the entirety of the graph. The path length values computed for each individual node (L_i^w) were also used in our analyses. The path length between nodes v_i and v_j is defined as the sum of the edge lengths along the path where each edge's length is obtained by computing the reciprocal of the edge weight $1/W_{ij}$, such that the weighted shortest path L_{ij}^w between nodes v_i and v_j is the length of the shortest

path between the nodes. L_{net}^w is then computed by measuring the shortest path lengths between all nodes in the network:

$$L_{net}^w = \frac{1}{\frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N \frac{1}{L_{ij}^w}}$$

where N is the number of nodes in the network. The average L_{net}^w reflects the global integration of a network. Short path lengths ensure that information quickly and easily spreads throughout the network, making efficient distributed and parallel information processing possible. Link lengths are inversely related to link weights, as large weights typically represent strong associations between nodes (Rubinov & Sporns, 2010).

3.1.3.5.2 Normalized Path Length

In order to normalize the computed values of L_{net}^w for differences in overall connection strength between vivid and dim memories, we randomly rewired each observed network. Network randomization was performed by randomly rewiring edges an average of 10 times for each full network. This randomization procedure was accomplished a total of 100 times for each corresponding real network, and the path lengths were computed for each random network. Then, the mean path length was computed for all 100 networks, and the mean of the mean path length for the 100 networks was computed to serve as L_{rand}^w in the calculation of λ :

$$\lambda = \frac{L_{net}^w}{L_{rand}^w}$$

L_{net}^w and λ provide summary statistics estimating global integration properties of an entire network. As such, these measures should be used in conjunction with other measures that provide further information regarding localized shifts in network topology.

3.1.3.5.3 Degree Centrality

Degree centrality in weighted networks measures the overall strength of a node's connections in terms of the total summed weights of their connections.

$$DC_i^w = \sum_j^N w_{ij}$$

where w_{ij} represents the weighted adjacency matrix in which $w_{ij} > 0$ if v_i is connected to v_j , and the value given to the link is the weight of the connection. Nodes with a high degree centrality directly interact with many other nodes in the network, are likely to be highly influential over the behavior of the network, and play a critical role in the flow of information or resources throughout the network. While degree centrality often identifies critical network elements, a highly essential node in the brain network may not necessarily have ubiquitous connections to other nodes in the network as assumed by degree centrality; thus, degree centrality should be used in conjunction with other centrality metrics to fully grasp the influence of a given node in a network.

3.1.3.5.4 Page Rank Centrality

Page rank centrality, a variant of eigenvector centrality ideal for small networks, is unique in that it considers the centrality of the immediate neighbors of node v_i in computing the centrality of v_i itself (van den Heuvel & Sporns, 2013). Although eigenvector centrality has been used more widely in complex brain network analyses, the distributions of centrality scores for page rank centrality tend to follow a normal distribution unlike eigenvector centrality scores (Zuo et al., 2011), making page rank centrality more amenable to standard statistical analyses. Much like degree centrality, page rank centrality favors nodes strongly connected with many other network nodes. In contrast to degree centrality it favors nodes that are connected to nodes that are themselves central within the network. Because of this recursive property, page rank centrality captures information regarding global features of the graph. Recent work has shown that page rank centrality is able to identify certain brain regions that are not as widely connected throughout the brain but are connected with key hubs (Zuo et al., 2011). This allows those nodes with high page rank centrality to integrate information throughout the entire network efficiently and effectively in relatively few steps. Mathematically, page rank centrality is defined as the stationary distribution achieved by instantiating a Markov chain on a graph, meaning that the page rank centrality of v_i is proportional to the number of steps spent at v_i as a result of that process (Ding, Yan, Frazho, & Caverlee, 2009). For ease of interpretation, page rank centrality in binary, unweighted networks is standardly defined as

$$PRC_i = (1 - d) \frac{1}{N} + d \sum_{i=1}^k \frac{PR(p_i)}{L(p_i)}$$

where p_1, p_2, \dots, p_N are the nodes under consideration, $PR(p_i)$ is the set of nodes that link to p_i , $L(p_i)$ is the number of connections for node p_i , and d is the damping factor. This equation is easily generalizable to weighted graphs (Ding et al., 2009; Rubinov & Sporns, 2010). The page rank index is modified by the addition of a damping factor, d , which specifies the fraction of time that a random walker will transition to one of its neighboring nodes in order to handle walking traps on graphs (Boldi, Santini, & Vigna, 2009). It is standard for the damping value to be set at $d = 0.85$, the value used in the present study. One limitation of the page rank centrality measure is that it fails to account for the disparity in the degree of a node with respect to its neighbors, which has important implications depending upon the network's assortativity, or the tendency for nodes to be connected to other nodes of a similar degree more often than would be expected by chance alone (Joyce, Laurienti, Burdette, & Hayasaka, 2010). We included in our analyses the recently developed measure of leverage centrality, which is able to overcome this limitation inherent in the calculation of page rank centrality.

3.1.3.5.5 Leverage Centrality

Leverage centrality considers the degree of a node, v_i , relative to its neighbors by identifying nodes in the network that are connected to more (or fewer) nodes than their immediate neighbors (Joyce et al., 2010).

$$LC_i^w = \frac{1}{k_i} \sum_{N_i} \frac{k_i - k_j}{k_i + k_j}$$

where k_i is the degree (summed strength) of node v_i . Nodes with high leverage centrality tend to be more connected to nodes of lower degree, whereas nodes with low leverage centrality tend to be more connected to nodes of higher degree. Leverage centrality is unique from other centrality measures in that it does not assume serial transportation of information, but rather allows for parallel information processing, which is fundamentally characteristic of certain systems such as the brain.

<h3>3.1.3.5.6 Nodal Reorganization of First Step Connections</h3>

The above nodal measures are dependent upon the entire set of connections between all network nodes (with the exception of degree centrality, which relies only on first step connections). That is, these functional brain networks are interdependent, nonlinear systems. Even though stark shifts in nodal network properties can only be explained in full by appealing to the entire network architecture, these shifts can only occur because of (1) the reorganization of connectivity strengths for the set of first step (direct) connections and/or (2) the reorganization of connectivity patterns beyond first step connections (i.e., indirect connections). Therefore, we sought to quantify the manner in which a given node's first step connections reorganize via a novel metric, First Step Reorganization (FSR_i^w). FSR_i^w is computed by comparing the strength of each individual connection from node v_i in the dim retrieval networks to the strength of the same

corresponding connection in the vivid retrieval networks. As such, this measure is conceptually and mathematically distinct from computing a change in degree centrality, because changes in degree centrality are merely dependent upon the value of the summed strength of connections for node v_i for each network. Formally, for the entire distribution of scores, FSR^w is calculated as

$$FSR^w = zscore \left\{ -\operatorname{arctanh} \left(\frac{\operatorname{Cov}(A_i, B_i)}{\sigma_{A_i} \sigma_{B_i}} \right) \right\}$$

where A_i is the connectivity profile of node i in network A , B_i is the connectivity profile of node i in network B , and the connectivity profile of node i is defined to be its distinct set of weighted connections (i.e., the column in the adjacency matrix corresponding to that particular node). A Fisher transformation ($\operatorname{arctanh}$) was implemented to render the distribution Gaussian, followed by a z-scoring of that distribution. Due to the standardization of the measure by means of z-scoring, the FSR_i^w value of a given node is always relative to the rest of the nodes in the network. Once the FSR_i^w value is obtained for each node, paired-samples t -tests can be used to determine whether a node exhibits significant reorganization of connections at the group level. Higher values of FSR_i^w indicate that the set of direct connections from node i exhibit greater reorganization between networks; lower values of FSR_i^w indicate that the set of direct connections from node i exhibit less reorganization between networks. If, in fact, a network node exhibits a significant change in FSR_i^w between networks, we can reasonably conclude that changes in the set of direct connections are at least partly

accounting for observed changes in path length, page rank centrality and leverage centrality. In contrast, if a network node changes minimally in FSR_i^w between networks, then any observed significant changes in path length, page rank centrality, and leverage centrality for that node are primarily due to changes in indirect connectivity elsewhere in the network.

3.1.3.6 Statistical Testing

3.1.3.6.1 Behavioral Data Analyses

Paired-samples t -tests were used to confirm the validity of the subjective in-scan memory ratings for the level of detail recollected at retrieval by comparing those scores to the responses for post-scan recognition and confidence ratings. Behavioral data was analyzed using SPSS and R statistical software.

3.1.3.6.2 Network Data Analyses

We used the permutation framework developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski index (Schubert, 2013; Schubert & Teles, 2014) to assess significance for differences in each graph theory measure at the hippocampus between dim and vivid memory retrieval. This same permutation framework was also used to assess differences in whole-brain network properties between dim and vivid memory retrieval. This permutation procedure allows for comparing groups of networks while accounting for the diverse topological features

inherent in each individual network. All p -values presented are uncorrected for multiple comparisons across measures. In assessing nodal changes in graph theory measures, it is important to know not only which nodes showed significant changes in network properties (L_i^w , DC_i^w , PRC_i^w , and LC_i^w) from dim to vivid memory retrieval, identified with the permutation test, but also which nodes showed the largest changes in relation to all other nodes in the network. To compare changes across nodes, we z-scored each network metric within subjects, and subsequently calculated the normalized difference between vivid and dim retrieval for each metric separately. All group-level statistics were formulated based upon this *nodal change score*. All network data analyses were implemented using standard and custom scripts in Matlab.

3.1.3 Results

Below we report behavioral results and the results of network analyses addressing our four questions concerning: (1) nodal changes in the hippocampus, (2) nodal changes in the hippocampus compared to other network nodes, (3) the extent to which the hippocampus reorganizes its set of first step connections, and (4) global network changes.

3.1.3.1 Behavioral Results

During the image recall task in the scanner, participants distributed their responses across the four vividness ratings (mean proportion of responses for ratings 1 to 4 were 19.9%, 22.6%, 28.8%, and 28.7%, respectively). Confirming the validity of these

ratings, the results of the post-scan forced-choice scan scene recognition test showed that accuracy was significantly greater ($t_{20} = 4.20, p < .001$) for vivid ($M = 82.3\%$; $SD = 1.2\%$) vs. dim ($M = 72.0\%$; $SD = 1.4\%$) memory retrieval. Furthermore, when considering only successfully recognized old scenes (hits), recognition confidence was reliably higher ($t_{20} = 7.85, p < .001$) for vivid memories (ratings 3-4, $M = 3.64, SD = 0.12$) than for dim memories (ratings 1-2; $M = 3.14, SD = 0.34$).

3.1.3.2 Effects of hippocampal network properties on memory vividness

The purpose of our first question was to investigate how complex patterns of hippocampal interactions with the rest of the network change from dim to vivid memory retrieval. To answer this question, we directly compared vivid vs. dim memories using the permutation procedure developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index (Schubert, 2013; Schubert & Telecs, 2014) for the four nodal measures of interest at the hippocampus: path length, degree centrality, page rank centrality, and leverage centrality. Consistent with our prediction, all four measures showed significant differences between vivid and dim memories in the hippocampus. Possibly due to the visual nature of the stimuli, all of these significant effects were significant in the right but not in the left hippocampus (see Table 2 for a summary of results). (1) Path length was significantly shorter for vivid than dim memories ($p = .0002$), which we interpret as a capacity for more efficient communication between the right hippocampus and the rest of the network supporting vivid

remembering. (2) Degree centrality was greater for vivid than dim memories ($p = .00009$) suggesting that stronger right hippocampal interactions with the rest of the network promote vivid remembering. (3) *Page rank centrality* was also higher for vivid than dim memories ($p = .013$) implying that the right hippocampus is connected to more central, influential nodes for enhancing memory vividness. (4) Finally, *leverage centrality* was also lower for vivid than dim memories ($p = .0029$) suggesting that the right hippocampus is connected with nodes of higher degree than itself promoting vividness in episodic memories. Given that the degree centrality of the right hippocampus significantly increased from dim to vivid memory retrieval *and* the right hippocampus still became more connected with nodes of a higher degree than itself for vivid memory retrieval, this indicates that the right hippocampus dramatically strengthens its connections with the more central, influential nodes in the network to support memory vividness. In sum, all nodal measures investigated support the idea that nodal changes in the right, but not necessarily the left, hippocampus are especially important in supporting vivid remembering. However, there was no significant dissociation between the right and left hippocampus for any graph theory measure. Because degree centrality is the most fundamental graph theory measure, Figure 11 displays the change in the degree centrality for all nodes exhibiting higher degree centrality during vivid than dim memory retrieval in standard brain space.

Table 2. Summary of medians and p-values obtained using the permutation framework for each network measure at the right hippocampus and for the entire brain network.

Network Measures		Dim Memory	Vivid Memory	<i>p-value</i>
<i>Right hippocampus</i>				
Path Length	L_i^w	2.517	1.959	.0002***
Degree Centrality	DC_i^w	15.216	22.477	.00009***
Page Rank Centrality	PRC_i^w	0.0094	0.0108	.013*
Leverage Centrality	LC_i^w	1.119	0.712	.0029**
<i>Global Measures</i>				
Path Length	L_{net}^w	2.062	1.887	.015*
Normalized Path Length	λ	1.045	1.022	.0001***

p < .05, **p < .01, *p < .001*

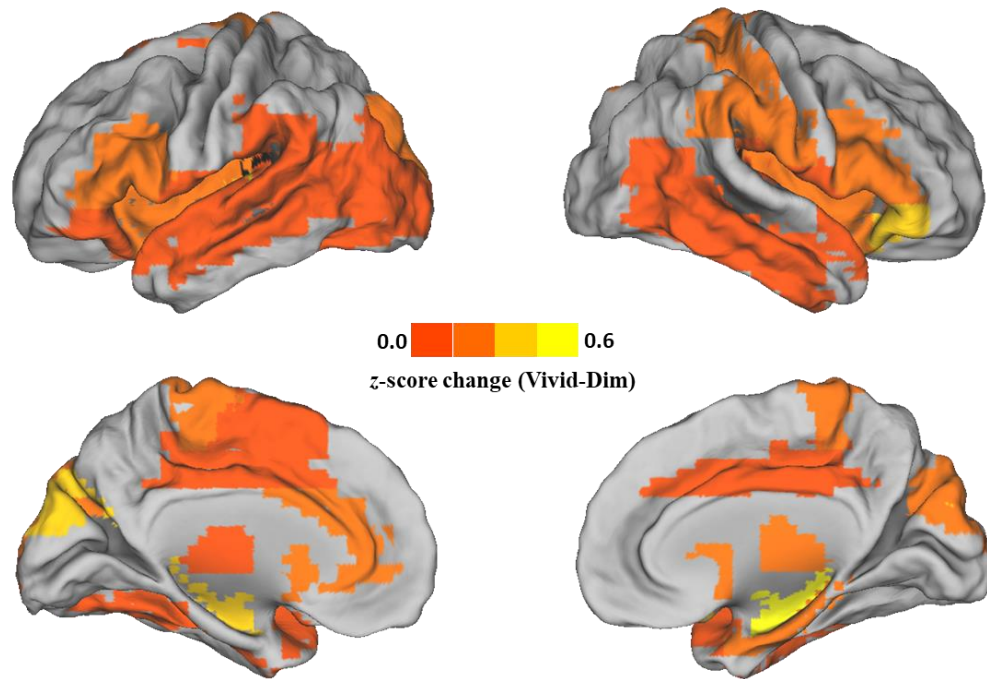


Figure 11. The change in (z-scored) degree centrality values are presented for all nodes that shift in a positive direction (i.e., nodes that exhibit higher degree centrality for vivid compared to dim retrieval) in standard brain space.

We then sought to determine whether univariate measures of retrieval dimness/vividness in the right or left hippocampus correlated (across subjects) with our four graph theory measures of interest (path length, degree centrality, page rank centrality, and leverage centrality). Initial group-level effect paired-samples t -tests on parameter estimates of vivid vs. dim memory retrieval yielded significant retrieval vividness effects (vivid > dim) for both the left and right hippocampal nodes (Left: $t(20) = 4.82, p < .0001$; Right: $t(20) = 5.04, p < .0001$). However, across subjects, univariate measures of activity within the vivid and dim conditions were *not* significantly correlated with graph theory metrics computed on the corresponding vivid and dim

networks, nor was the vivid > dim univariate difference correlated with the corresponding difference of any graph theory measures (all p 's > .25).

3.1.3.3 Effects of hippocampal network properties on memory vividness compared to other network nodes

Our second question addressed how the magnitude and direction of hippocampal nodal changes supporting episodic memory vividness is related to observed changes in all other network nodes. Despite significant shifts in hippocampal network properties from dim to vivid memory retrieval, it is possible that properties of other network nodes are more dramatically tracking differences in memory vividness. To address this question, we (1) calculated the same four graph theory measures of interest (path length, degree centrality, page rank centrality, and leverage centrality) for each of the 90 nodes in the network and computed the nodal change score; (2) averaged these nodal change scores across participants and z-scored the result to determine which nodes exhibited the greatest changes in our four measures of interest in support of memory vividness.

Consistent with our predictions, the hippocampus (right) exhibited the greatest magnitude change of any network node for supporting vivid episodic remembering. From dim to vivid memory retrieval, the right hippocampus exhibited the largest decrease in path length and leverage centrality as well as the largest increase in degree and page rank centrality compared to all other network nodes (see Figure 12). These results demonstrate for the first time that changes in the topological properties of the

right hippocampus more closely track visual episodic retrieval performance than any other regions in the brain.

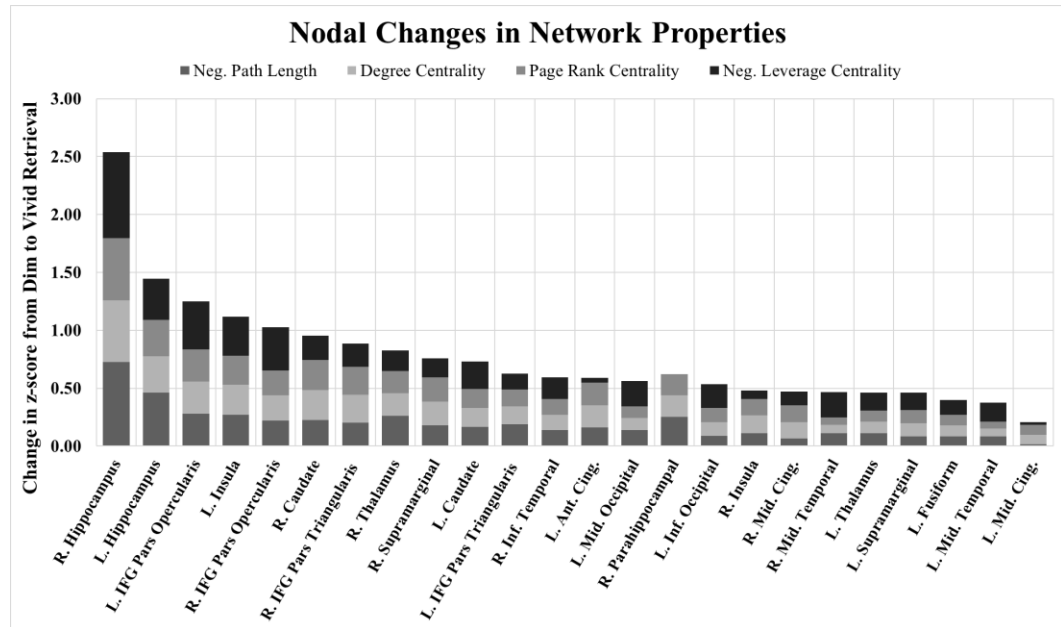


Figure 12. Provides a summary of how potential memory-related network nodes change between dim and vivid retrieval with respect to the four nodal measures of interest: path length, degree centrality, page rank centrality, and leverage centrality. Only nodes for which each measure changes in the same direction as the hippocampus are included, because only changes in those directions result in a node more efficiently communicating with the rest of the network and becoming more central or influential in the network. The signs of the z-scored values for path length and leverage centrality were flipped on the figure, such that higher z-scores for path length and leverage centrality represent shorter path lengths and lower leverage centrality, respectively. The largest shifts in the entire network for each of these four nodal measures occur for the right hippocampus. Furthermore, while changes in page rank centrality and degree centrality appear to be inversely related to changes in path length and leverage centrality, respectively, this is not a mathematical certainty, as each measure characterizes unique features of the network in accordance with different criteria.

Although we predicted that the hippocampus would display stark shifts in network properties as a function of memory vividness, we did not expect this region to

necessarily be among the most central, influential nodes in the network independent from the shift in memory performance. In keeping with this idea, if instead of ranking the magnitude of the changes between vivid and dim conditions, we rank the absolute value of nodal measures within one of these conditions, such as the vivid condition, then the hippocampus is not among the subset of nodes with shortest path length, highest degree centrality, highest page rank centrality, or lowest leverage centrality. In the vivid condition, for example, the right hippocampus ranked 68th for path length, 67th for degree centrality, 67th for page rank centrality, and 61st for leverage centrality. That is, for vivid memories, the right hippocampus actually displayed a higher path length, lower degree centrality, lower page rank centrality, and higher leverage centrality than the average node in the network for each of these respective measures. Perhaps more surprisingly, for the dim memory condition, the right hippocampus had some of the lowest rankings in path length (85th), degree centrality (86th), page rank centrality (89th), and leverage centrality (88th). That is, the right hippocampus was among the select subset of nodes with the highest path length, lowest degree centrality, lowest page rank centrality, and highest leverage centrality for dim retrieval.

These results suggest that when participants were unable to vividly recollect the previously encoded scene, the right hippocampus did not participate in efficient information transfer with other nodes across the network, interacted with relatively few other nodes in the network, and did not connect with other central, influential nodes in

the network. Collectively, these results suggest that the right hippocampus was not involved in integrative processes in the case of dim memories but became sufficiently involved in integrative processes to support the recollection of vivid memories. The fact that the hippocampus displayed the greatest magnitude *shift* for each graph theory measure from dim to vivid retrieval but was not among the subset of nodes with lowest path length, highest degree centrality, highest page rank centrality or lowest leverage centrality at either dim or vivid retrieval, separately, emphasizes the importance of examining shifts in nodal properties between dim and vivid memory retrieval as opposed to the absolute ranking of a node within each network separately.

3.2.3.4 Effects of hippocampal first-step connection reorganization on memory vividness

The purpose of our third question was to determine the extent to which the hippocampus reorganizes its set of first step (direct) connections from dim to vivid memory retrieval. To answer this question, we created a novel measure, called *First Step Reorganization* (FSR_i^w). FSR_i^w is computed by comparing the strength of each individual connection from node v_i in the dim retrieval condition to the strength of the same *corresponding connection* in the vivid retrieval condition. In addition to significant changes in each graph theory measure, the right hippocampus also exhibited a significant reorganization of first step connections between dim and vivid memory retrieval, $t(20) = 3.33, p = .003$. Because the right hippocampus exhibited a dramatic change in FSR_i^w between dim and vivid retrieval networks, we can reasonably conclude

that this substantial reorganization of direct connections is at least partly responsible for observed changes in path length, page rank centrality and leverage centrality. Thus, the observed changes in path length, page rank centrality, and leverage centrality are not merely due to changes in connectivity elsewhere in the network that do not directly involve interactions with the right hippocampus. The extent to which a node reorganizes its set of connections is not trivially related to any of the four nodal measures of interest displayed for each node in Figure 12. Quantitatively, none of the four nodal measures of interest were significantly correlated with FSR_i^w (all p 's > .15, uncorrected).

To further investigate how the right hippocampus reorganizes its set of first step connections, we computed the average change in connectivity strength between the right hippocampus and every other network node from dim to vivid memory retrieval. As illustrated by Figure 13, the right hippocampus exhibited substantial increases in average connectivity strength with nodes commonly assumed to be functionally connected with the hippocampus during successful and/or vivid episodic retrieval, such as occipital regions supposed to store visual memory traces and frontal regions supposed to mediate retrieval control operations (King et al., 2015; Schedlbauer et al., 2014; Spaniol et al., 2009).

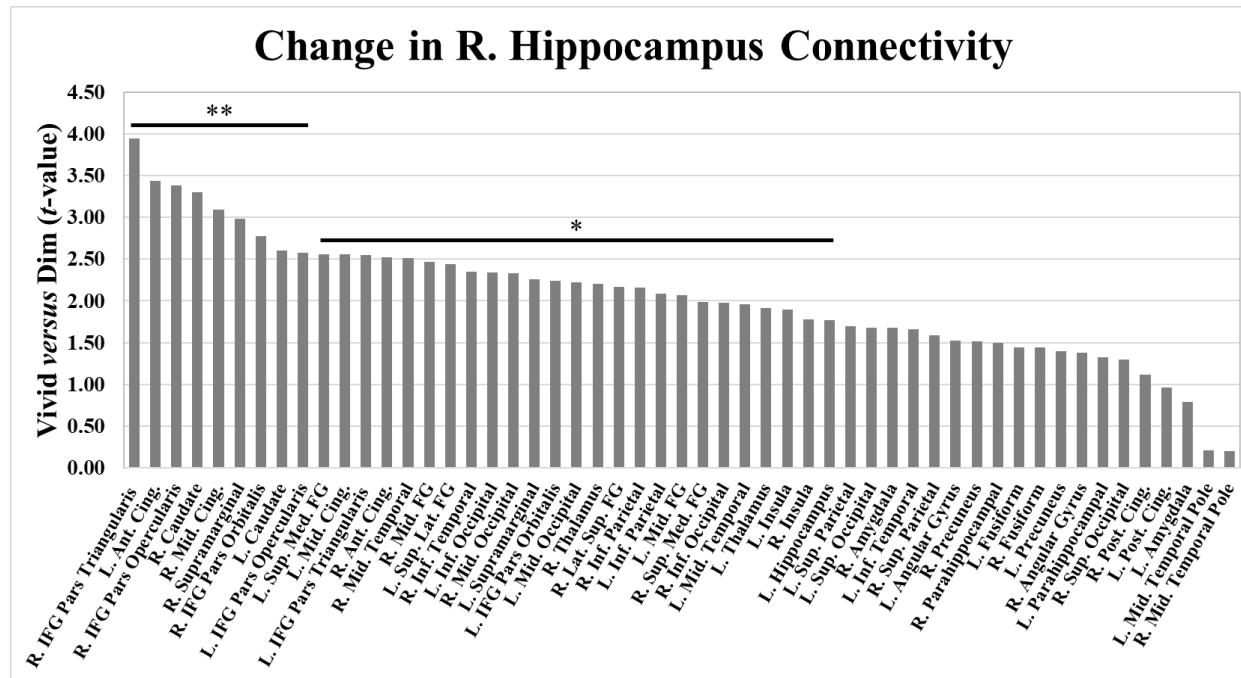


Figure 13. Provides a summary of the t-values associated with the average change in connectivity strength between the right hippocampus and the set of potential memory-related nodes from dim to vivid memory retrieval (* $p < .05$, ** $p < .01$). Only nodes for which the right hippocampus exhibited increased connectivity from dim to vivid retrieval are presented in the figure.

The right hippocampus exhibited the largest magnitude increases in its connectivity strength with inferior frontal regions thought to mediate retrieval control, caudate, inferior lateral temporal regions thought to be critical for visual memory, and occipital regions thought to store visual memory information. This demonstrates that the right hippocampus was, in fact, communicating more strongly with the select subset of regions thought to be highly involved in episodic memory for visual stimuli. It is important to note that this figure is meant only to show which nodes the right hippocampus tends to become more connected with for vivid memory retrieval (i.e., which nodes contributed more than others to the increase in hippocampal degree centrality); as such, we did not correct for multiple comparisons.

3.2.3.5 Extra-hippocampal global network effects on memory vividness

Whereas our first three questions were about the topological properties of the hippocampus, our last question concerned the capacity for effective information integration in the entire brain network. More specifically, are there *global* changes in path length from dim to vivid memory retrieval, and do those global changes persist even after removing the right hippocampus from the networks? To answer these questions, we computed the vivid-dim difference in path length across the entire brain network (L_{net}^w). Using the permutation framework developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski index, we found a significant difference in L_{net}^w between vivid and dim memory retrieval conditions, such that the path length in the entire brain network was shorter for vivid than dim memory retrieval (see Table 2). We then normalized L_{net}^w by dividing by L_{rand}^w to obtain values of lambda (λ) for vivid and dim retrieval conditions, respectively. Even after normalizing the path length across networks, the permutation revealed that the vivid retrieval networks demonstrated a greater capacity for efficient communication across the entirety of the network (see Table 2). Even after removing both hippocampi from the network and re-calculating the global path length, the permutation revealed that the global path length was shorter for vivid than dim retrieval ($p < .05$). Furthermore, after removing both hippocampi from the network, re-calculating the global path length, and normalizing the global path length, the permutation revealed that the normalized global path length was still shorter for the

vivid retrieval condition than dim retrieval condition ($p < .001$). This demonstrates that dramatic changes in path length at the hippocampi are not exclusively responsible for driving the average global change in path length from dim to vivid retrieval.

While the hippocampus (right) maintained the largest magnitude change in path length of any network node, the fact that the network as a whole exhibited a shorter average path length during vivid retrieval (even after removing the hippocampi from the network) suggests that many other nodes exhibited a lower path length at vivid retrieval as opposed to dim retrieval. In fact, Figure 12 shows which other nodes, identified as involved in memory-related processes in activation (Kim, 2015; Spaniol et al., 2009) and connectivity analyses (King et al., 2015; Schedlbauer et al., 2014), tended to exhibit shorter path lengths at vivid retrieval as opposed to dim retrieval relative to each other. Interestingly, these tend to be the same nodes identified in Figure 13 that significantly increase in connectivity strength with the hippocampus to support vivid memory retrieval.

3.1.4 General Discussion

The purpose of this study was to investigate properties of the large-scale network underlying episodic memory by using graph theory measures, focusing in particular on the contribution of the hippocampus in supporting patterns of connectivity underlying vivid memories for visual scenes. The study yielded four main findings. First, when recollecting vivid compared to dim memories, the hippocampus displayed

greater efficiency in communication with the rest of the network (shorter path length), connected more strongly with the rest of the network (greater degree centrality), preferentially connected with the most central nodes in the network (higher page rank centrality), and still connected with nodes of a higher degree than itself (lower leverage centrality). Second, among all ninety nodes in the network, nodal changes in the right hippocampus made the largest contributions to visual memory vividness. Third, the stark shifts in hippocampal network properties were at least partly due to the fact that the hippocampus massively reorganizes its set of direct connections to support vivid memory retrieval. Finally, beyond the hippocampus, the brain network as a whole displayed a greater capacity for efficient communication throughout the network (shorter global path length) to facilitate vivid memory retrieval. This shift in the functional profile of the hippocampus to support vivid memory retrieval provides novel insight into why the hippocampus is a critical brain area for episodic memory processes. These four main findings are discussed in separate sections below.

3.1.4.1 Right hippocampal network interactions significantly impact memory vividness

While episodic memory is subserved by complex neural interactions and the continuous exchange of information between circuits distributed across several brain regions, the hippocampus in particular stands out as the critical structure for the encoding, storage, and retrieval of such memories (Battaglia, Benchenane, Sirota, Pennartz, & Wiener, 2011; Watrous & Ekstrom, 2014). Determining how the

hippocampus supports the unique demands of retrieving experienced events from memory is fundamental to understanding the biological basis of episodic memory. Although lesion studies have provided clear evidence for the destructive effects of hippocampal damage on episodic memory and prior functional neuroimaging studies of healthy adults have demonstrated greater hippocampal activations for successful than unsuccessful episodic memories (Eichenbaum et al., 2007; Rugg et al., 2012; Skinner & Fernandes, 2007; Spaniol et al., 2009), these methods have not characterized how the hippocampus dynamically interacts (directly or indirectly) with other relevant brain regions during intact episodic memory retrieval. Given that the human hippocampus is thought to provide the vital integrative link that receives relevant sensory information during encoding and ultimately facilitates memory retrieval via reactivation of memory traces with input from frontal regions mediating retrieval control operations (Alvarez & Squire, 1994; Danker & Anderson, 2010; McClelland et al., 1995; Ritchey et al., 2013; Teyler & Rudy, 2007), understanding the role of the hippocampus in the large-scale episodic memory network is critical for understanding the neural basis of memory. While this prior work has provided evidence for the importance of the hippocampus to episodic memory, our results provide a unifying framework showing that the direct and indirect connections between the hippocampus and all other network nodes contribute to stark changes in the complexity of the system to support vivid episodic memory retrieval.

The information processing properties of the hippocampus have traditionally been studied at the more microscopic, local level with a focus on information flow and plasticity exclusively within the hippocampal formation (Battaglia et al., 2011). Watrous and Ekstrom (2014) recently proposed that both cell assembly firing patterns and global patterns of brain oscillatory activity within hippocampal-neocortical networks form the basis of a memory. Our results also suggest the hippocampus is critical for information processing at the more macroscopic level of the entire functional brain network to facilitate vivid visual memory retrieval, building on and extending considerable work regarding the role of hippocampal-cortical projections related to memory processes (Bai et al., 2009; McCormick, Moscovitch, Protzner, Huber, & McAndrews, 2010; McIntosh, Nyberg, Bookstein, & Tulving, 1997; Ranganath, Heller, Cohen, Brozinsky, & Rissman, 2005; Ritchey et al., 2013; Robin et al., 2015; Rolls, 2000; Sadeh, Shohamy, Levy, Reggev, & Maril, 2010; Schott et al., 2013; Takahashi, Ohki, & Kim, 2008). The integration of information from diverse functional domains at a more macroscopic level may serve as a key feature enabling the hippocampus to support not only the vivid retrieval of episodic memories, but also their translation into more complex, adaptive behaviors.

Several recent studies have extended existing task-based functional connectivity analyses of memory retrieval by identifying changes in bivariate connections (King et al., 2015) or basic graph theory measures (Schedlbauer et al., 2014) between small subsets of a priori selected, memory-related brain regions. Of particular relevance to the

current study, Schedlbauer et al. (2014) demonstrated that several areas within the medial temporal (including the hippocampus), frontal and parietal lobes exhibited significantly greater connectivity with several other brain regions and were more often along shortest paths in the network during successful retrieval, underscoring the importance of integrative processes in supporting memory. However, there may be differences in multivariate graph theory measures between small networks comprised of a small proportion of brain space and larger-scale networks that include all cortical and subcortical regions (Stanley et al., 2013). To the best of our knowledge, no existing whole-brain network analyses during active memory retrieval have been reported in the literature. Our study also extends this prior work by using more data-driven analysis methods, utilizing multiple multivariate graph theory measures to better assess integrative properties in the networks, and creating an innovative new measure to aid in understanding why these shifts in multivariate graph theory measures were observed.

3.1.4.2 Network measures for the right hippocampus tracked memory vividness more than any other brain region

When separate brain networks were constructed in accordance with measured performance during the retrieval of visual scenes, the topological properties of the right hippocampus shifted more than any other node in the network between vivid and dim memory retrieval. The right hippocampus exhibited a greater capacity for efficient communication with the rest of the network (shorter path length), communicated more strongly with directly connected nodes (higher degree centrality), interacted with more

central nodes (higher page rank centrality), and preferentially strengthened connections with more connected nodes than itself (lower leverage centrality). In other words, compared to all other brain areas, the hippocampus was the brain region where improvements in the capacity for efficient communication and integration of disparate information most closely tracked reports of memory vividness.

These findings provide novel insight into existing memory theories that identify the hippocampus as a “bottleneck” or “convergence zone” at which information from distributed brain regions is processed and integrated to facilitate memory retrieval (Damasio, 1989; Mišić et al., 2014; Moll & Miikkulainen, 1997). Though it is widely assumed that the hippocampus operates as a convergence zone during the encoding and retrieval of episodic memories, the lack of an adequate, multivariate analytic methodology has stymied progress on properly investigating this idea in healthy, living humans. With the explosion of interest in large-scale complex network analyses coupled with recent advances in statistical physics that have produced relevant graph theory measures, we now have the requisite measures for identifying brain regions that facilitate the convergence and joint processing of specialized information from distributed brain regions while participants actively engage in memory-related tasks. The ability to identify the hippocampus as the network node that adjusts its integrative properties the most in accordance with the vividness of retrieved memories provides novel evidence for the hippocampus as a convergence zone in the brain. Future studies

using high-resolution fMRI and manual ROI tracing might be able to investigate whether episodic memory for object features involves other convergence zones besides the hippocampus, such as perirhinal cortex (Bussey, Saksida, & Elisabeth, 2005; McLelland, Chan, Ferber, & Barense, 2014; Ryan et al., 2012).

Despite the assumption that the hippocampus occupies a convergence zone in the brain, graph theory analyses of structural (Gong et al., 2008; Hagmann et al., 2008; van den Heuvel, Kahn, Goñi, & Sporns, 2012) and functional (Buckner et al., 2009; Tomasi & Volkow, 2010) whole-brain networks *at rest* in humans have consistently failed to demonstrate the topological centrality of the hippocampus. The hippocampus has not been identified as among the most highly connected network nodes, nor has it been shown to occupy a critical position along shortest paths in the network. Similarly, we found that the hippocampus was not among the most highly connected, influential network nodes in vivid or dim retrieval networks taken separately. This does not imply, however, that the hippocampus should not be considered a convergence zone in these episodic networks; it is the *relative change* in network properties at the hippocampus in accordance with memory performance that provide insight into the integrative functions of the hippocampus for memory retrieval. A more general implication of this finding is that to investigate the properties of brain regions within large-scale memory networks, it is necessary to focus on functional connectivity patterns associated with memory

performance. This typically requires the use of event-related fMRI and the construction of separate connectivity matrices for trials that differ in behavioral performance.

3.1.4.3 Hippocampal first-step connections substantially reorganize to support vivid memory retrieval

Even though observed shifts in network properties of the right hippocampus can only be explained in full by appealing to the entire network architecture, those shifts are at least partly due to a substantial reorganization of connectivity strengths for the set of first step (direct) connections. Because the right hippocampus exhibited a substantial reorganization of its direct connections between dim and vivid retrieval, we can reasonably conclude that this substantial reorganization is at least partly responsible for observed changes in path length, page rank centrality and leverage centrality. Thus, the changes in these measures are not merely due to changes in connectivity elsewhere in the network that do not directly involve interactions with the right hippocampus. The finding that the hippocampus drastically reorganized its set of direct connections from dim to vivid retrieval provides complementary support for the idea that this region can alter its connectivity profile to operate as a convergence zone during successful episodic memory.

To acquire more specific information about how the right hippocampus reorganizes its set of first step connections, we identified the magnitude and direction of changes in connectivity strength between the right hippocampus and every other network node from dim to vivid memory retrieval. The right hippocampus exhibited

significant increases in connectivity strength with nodes commonly assumed to directly interact with the hippocampus during vivid episodic retrieval, such as occipito-temporal regions supposed to store visual memory traces and frontal regions supposed to mediate retrieval control operations (see Figure 13). This finding is consistent with recent work from King et al. (2015), who have demonstrated that the hippocampus shows increased connectivity with many diffuse, functionally distinct brain regions for recollecting picture-word pairs. Relatedly, the stark reorganization in first step connections to the right hippocampus suggests that the significant overall increase in connectivity strength (degree centrality) of the right hippocampus from dim to vivid memory retrieval is *not* due to a relatively even strengthening of all hippocampal connections. Instead, there is tremendous variability in the magnitude of connectivity strength shifts between corresponding connections during vivid vs. dim memories.

3.1.4.4 Extra-hippocampal global network changes support vivid memory retrieval

Beyond the hippocampal structures, studies with brain-damaged patients and more traditional functional activation analyses in healthy adults have demonstrated that numerous other brain regions are critically important for episodic memory retrieval, including: posterior parietal regions, precuneus, prefrontal cortex, thalamus, retrosplenial and posterior cingulate regions, occipito-temporal regions, and other medial temporal lobe structures (Cabeza et al., 2008; Cabeza & Nyberg, 2000; Huijbers, Pennartz, Cabeza, & Daselaar, 2011; Rugg & Vilberg, 2013; Spaniol et al., 2009; Wagner,

Shannon, Kahn, & Buckner, 2005). Importantly, the prefrontal cortex, precuneus, visual cortex, thalamus, posterior parietal regions, and other medial temporal lobe structures (in addition to the hippocampus) have been identified as densely interconnected during successful episodic memory retrieval (Preston & Eichenbaum, 2013; Schedlbauer et al., 2014; Staresina, Cooper, & Henson, 2013). These results collectively emphasize the idea that a dense, interconnected network of disparate brain regions, each with a particular function, facilitates episodic memory retrieval. Although the hippocampus may be necessary for vivid, rich episodic memory retrieval, its proper functioning in the network is likely not sufficient. Thus, it is reasonable to hypothesize that more global integrative properties of the network are predictive of vivid memories of recently encoded events. Supporting this hypothesis, our results demonstrated that the network as a whole exhibited a greater capacity for effective information integration supporting vivid memory retrieval. Even after removing the right hippocampus from the networks, vivid retrieval networks still exhibited a greater capacity for efficient information integration.

3.1.4.5 Conclusions

In sum, the results obtained in this study demonstrate that the manner in which the network as a whole efficiently integrates information and the specific role of the right hippocampus in integrating information changes significantly between vivid and dim retrieval of recently encoded scenes. Our results, therefore, provide a new

perspective on the neural basis of episodic memory, capturing the importance of optimal integration in the large-scale network as a whole and among critically important network nodes embedded within the network in generating vivid, detailed episodic memories.

3.2 From hippocampus to whole-brain: The role of integrative processing in episodic memory retrieval

Multivariate functional connectivity analyses of neuroimaging data have revealed the importance of complex, distributed interactions between disparate yet interdependent brain regions. Recent work has shown that specific topological properties of functional brain networks are associated with individual and group differences in cognitive performance, including in episodic memory. After constructing functional whole-brain networks derived from an event-related fMRI study of memory retrieval, we examined differences in functional brain network architecture between forgotten and remembered words. This study yielded three main findings. First, multivariate graph theory analyses showed that successfully remembering compared to forgetting was associated with significant changes in the connectivity profile of the left hippocampus and a corresponding increase in efficient communication with the rest of the brain. Second, bivariate functional connectivity analyses indicated stronger interactions between the left hippocampus and a *retrieval assembly* for remembered versus forgotten items. This assembly included the left precuneus, left caudate, bilateral supramarginal gyrus, and the bilateral dorsolateral superior frontal gyrus. Integrative

properties of the *retrieval assembly* were also greater in the remembered condition than in the forgotten condition. Third, whole-brain modularity analyses revealed that successful memory retrieval was associated with increased integration, as indexed by a less segregated modular architecture in the network. Additionally, the magnitude of the decreases in modularity between remembered and forgotten retrieval networks was closely related to memory performance. These findings indicate that increases in integrative properties at the *nodal*, *retrieval assembly* and *whole-brain* topological levels facilitate memory retrieval, while also underscoring the potential of multivariate brain connectivity approaches for providing valuable new insights into the neural bases of memory processes.

3.2.1 Introduction

The human brain is a large-scale complex system comprised of diverse yet interconnected brain regions (Bullmore & Sporns, 2009). Most existing neuroimaging research investigating the neural bases of cognitive processes has focused on the localization of specific functions using univariate activation methodologies. For example, consistently observed increases in hippocampal activity during memory retrieval tasks for remembered versus forgotten items suggest that the hippocampus is critical for successful episodic memory retrieval (J. Kim, Kim, & Lee, 2013; Rugg & Vilberg, 2013; Spaniol et al., 2009). This corpus of work has focused on identifying encapsulated, circumscribed regions thought to be involved in certain cognitive

processes. However, individual regions, such as the hippocampus, can only support episodic memory by dynamically interacting with other diverse and spatially distributed brain regions (Geib, Stanley, Wing, Laurienti, & Cabeza, 2017; Mišić et al., 2014; Watrous & Ekstrom, 2014). In the current study, we utilized graph theory measures to investigate how complex patterns of functional interactions at different topological levels underlie episodic memory retrieval for words.

Graph theory provides a particularly powerful framework for characterizing brain networks derived from functional neuroimaging data (Bullmore & Sporns, 2009; Rubinov & Sporns, 2010). Using the mathematical formalisms of graph theory, networks of brain regions can be represented as graphs consisting of a set of nodes with the pairwise relationships between them, known as edges. In functional neuroimaging analyses, each node represents a discrete brain region, and the edges represent the measured functional connectivity between pairs of nodes.

There is general agreement that healthy and effective brain network architectures require both segregated and integrative processing (Bassett & Bullmore, 2006; Bressler & Menon, 2010; Medaglia, Lynall, & Bassett, 2015; Sporns, 2013; Tononi, Sporns, & Edelman, 1994). Segregation refers to specialized processing within an individual brain region (Tononi et al., 1994) or within a small subset of tightly interconnected regions (Bullmore & Sporns, 2009). In contrast, integration refers to the assimilation and transfer of information between specialized, spatially distributed brain regions (Rubinov &

Sporns, 2010). In the present paper, we focus on the importance of integrative properties in functional brain networks during episodic memory retrieval.

Most graph theory analyses of functional neuroimaging data have investigated network topology during resting state e.g. (Buckner et al., 2009; Hayasaka & Laurienti, 2010; He et al., 2009; Malaak N Moussa, Steen, Laurienti, & Hayasaka, 2012; Stevens, Tappan, Garg, & Fair, 2012; van den Heuvel, Stam, Kahn, & Pol, 2009), with relatively few studies having investigated alterations in network topology during cognitive tasks. Among the studies that have linked graph theory measures to cognitive performance on a task e.g. (Bassett et al., 2010; Braun et al., 2015; Cao et al., 2014; Meunier et al., 2014; Malaak Nasser Moussa et al., 2011, 2014; Stanley et al., 2014, 2015), only three have investigated patterns of functional network topology subserving episodic memory (Geib et al., 2017; King et al., 2015; Schedlbauer et al., 2014). In particular, Schedlbauer et al. (2014) showed that a substantial proportion of the shortest topological paths between disparate brain regions pass through the hippocampus to support successful memory retrieval. King et al. (2015) showed that a subset of brain regions—*a priori* identified from previous activation analyses as being involved memory retrieval—become more strongly connected with one another to support successful remembering and that the strength of these connections is related to recollection accuracy. Lastly, Geib et al. (2017) showed that the hippocampus, embedded within the whole-brain network, re-organized its set of functional connections, displayed greater communication efficiency with the

rest of the brain, and became a more convergent structure for information integration supporting the vivid compared to dim retrieval of visual scenes. All of these studies underscore the importance of integrative processing for successful memory retrieval, expanding beyond prior emphasis on segregated processing.

To support episodic memory retrieval, recent work has shown that the hippocampus participates in more efficient communication with many other brain regions and serves as the critical integrative structure for the convergence and joint processing of information (Battaglia et al., 2011; Geib et al., 2017; Mišić et al., 2014; Watrous & Ekstrom, 2014). However, this prior work has largely focused on the role of an individual node (i.e., the hippocampus) embedded within a larger network. The current study builds on this prior work by characterizing network architecture at three different organizational levels: a nodal level, a subnetwork or retrieval assembly level, and the whole-brain network level. We maintain that the hippocampus serves as a critical integrative structure for successful memory retrieval, but we also suggest that the functional interactions of the hippocampus with a retrieval assembly facilitates efficient integrative processing and successful memory retrieval. We define a retrieval assembly as a subset of network nodes that become more strongly connected with the hippocampus during remembered trials compared to forgotten trials. Furthermore, prior work has also suggested that the extent to which the functional network as a whole

exhibits certain integrative properties facilitates successful memory retrieval (Geib et al., 2017).

Specifically, we made three primary hypotheses in the current study. First, extending prior work investigating topological properties of the hippocampus embedded within the entire functional brain network (Geib et al., 2017; Schedlbauer et al., 2014), we hypothesized that the hippocampus would exhibit an increase in its role in integrative processing (i.e., more efficient communication with the rest of the brain) for remembered versus forgotten items.

Second, we hypothesized that successfully remembering compared to forgetting items would be associated with stronger connectivity between the hippocampus and a subset of regions involved in memory retrieval, subsequently identified as a “retrieval assembly.” More specifically, we hypothesized that certain subregions within the prefrontal cortex (PFC) and the ventral parietal cortex (VPC) would serve as critical nodes in this retrieval assembly. The reasoning for this is that the PFC is thought to support retrieval control processes, such as retrieval search and monitoring (Anderson, Bunce, & Barbas, 2016; Preston & Eichenbaum, 2013; Wagner et al., 2005), and the VPC is thought to mediate attentional processes involved in memory retrieval (Cabeza et al., 2008; Ciaramelli, Grady, & Moscovitch, 2008). There is also mounting evidence from bivariate functional connectivity analyses indicating that both PFC (Blumenfeld & Ranganath, 2007; Preston & Eichenbaum, 2013) and VPC are functionally connected to

the hippocampus during episodic retrieval (King et al., 2015; Robin et al., 2015).

Additionally, we hypothesized that this retrieval assembly would exhibit increased integration with the rest of the brain for remembered compared to forgotten items.

Third, extending prior work showing that increased overall whole-brain integrative processing is closely related to cognitive performance (Cao et al., 2014; Geib et al., 2017; Meunier et al., 2014; Stanley et al., 2014, 2015; Stevens et al., 2012; van den Heuvel et al., 2009), we hypothesized that successful memory retrieval would be associated with a more globally integrated (and less segregated) network architecture across the entire brain. To address this question, we implemented data-driven modularity algorithms to assess the relative balance between segregated and integrative network architectures in remembered and forgotten episodic memory retrieval networks.

3.2.2 Methods

3.2.2.1 Participants

Nineteen right-handed healthy young adult participants (mean age = 22.29 years, $SD = 3.20$, range [18, 30]; 7 female, 12 male) completed the experiment. Participants with a history of neurological difficulties or psychiatric illness, alcoholism, drug abuse, or learning disabilities were excluded from the study. Due to technical problems during data acquisition, two subjects were excluded from the analyses: one lacked functional data, and the other lacked behavioral data. All analyses were performed on the

remaining 17 participants. Written informed consent was obtained from each participant in accordance with a protocol approved by the Duke University Institutional Review Board. All participants were monetarily compensated for their time.

3.2.2.2 Experimental Design

Three-hundred sixty concrete nouns were chosen from the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm) and used as stimuli. During the encoding task, a total of 240 words was presented, equally divided across 4 functional runs. Each run included five on-blocks (task being performed, 60 sec each) and 6 off-blocks (rest, 15 sec each). During each on-block, participants were presented with 12 words serially, for a duration of 2 seconds each. During the first two encoding runs, the participants were asked to make an animacy decision for each word (i.e., living /non-living). Each word was presented in the center of the screen, and the living/non-living choice was symmetrically placed to the left and right below each word. During the last two encoding runs of the four, participants were asked to read and remember each word and press the button corresponding to the location of the word 'press' after they had completed reading the word. Again, the stimulus word was presented in the center of the screen and the two options for responding at the bottom of the screen ('press' and 'xxxx') were evenly spaced below each word. Button responses and response times (RTs) were recorded using a magnetically shielded 4-button box held in the participant's right hand. Stimuli were separated with a jittered inter-

stimulus-interval (ISI) which ranged from 1 to 4 seconds to facilitate deconvolution and extraction of the hemodynamic response. Off-block stimuli consisted of a continuously presented crosshair figure at fixation, which was also used for memory-task on-block ISI periods.

The encoding runs were followed by 20 minutes of anatomical and DTI scanning. Participants then completed a recognition test in the scanner. Across six retrieval runs, participants were presented with each of the 240 words that had been presented during encoding along with 120 new words. Word duration, ISI distribution, block length, and block order mirrored that of encoding. In response to each presented word, participants were asked to make an old/new judgment and indicate their confidence in that judgement (definitely old, probably old, probably new, definitely new). Again, words were presented one at a time in the center of a computer screen. The old/new confidence judgement was displayed below each word, and participants pressed a corresponding key to indicate whether the word was definitely old, probably old, probably new, or definitely new. The current study was based on retrieval trials for the encoding trials for which they were explicitly told to remember the word (i.e. runs 3 and 4).

3.2.2.3 Data Acquisition and Pre-Processing

Imaging data were collected on a 3T GE scanner. Following a localizer scan, functional images were acquired using a SENSE spiral-in sequence (TR = 2000ms, TE =

27ms, FOV = 24cm, 34 oblique slices with voxel dimensions of 3.75x3.75x3.8mm). Stimuli were projected onto a mirror at the back of the scanner bore, and responses were recorded using a 4-button fiber optic response box. Scanner noise was reduced with ear plugs, and head motion was minimized with foam pads. A high-resolution anatomical image (96 axial slices parallel to the AC-PC plane with voxel dimensions of 0.9x0.9x1.9mm) was collected following the functional scanning.

Following acquisition, each participant's volumes were slice-time corrected, realigned to the first volume, and normalized to a standard EPI template in SPM8. Single-trial beta estimates of the event-related hemodynamic responses were computed using a method recently developed by Mumford et al. (2012) that better represents true activation magnitudes in fast event-related designs. Single-trial estimates also included six motion regressors (x-dim, y-dim, z-dim, roll, pitch, and yaw) in addition to block-wise run regressors. Following single-trial estimation, each voxel's time series was winsorized at three standard deviations above and below the mean. Overall translational movement (Euclidian distance from the origin) was minimal $M = 0.31$ ($SD = 0.16$ cm). Additionally, only 5% of the volumes had motion exceeding 1 cm and <1% of the volumes had motion exceeding 2 cm. Therefore, the effects of motion were negligible.

3.2.2.4 Generating Whole Brain Networks

To create episodic retrieval networks, we used a beta time-series analysis, which assumes that two regions are functionally coupled during a task if their activity is significantly correlated across trials (Fornito et al., 2011; Rissman et al., 2004; Schedlbauer et al., 2014). Each beta value reflected the magnitude of the hemodynamic response evoked by a particular trial during the retrieval phase of the procedure. These observed beta values were then sorted for each individual participant separately in accordance with memory performance during the corresponding trial during the retrieval phase of the experiment. “Definitely old” responses to old items constituted the remembered-trial category. “Probably new” and “definitely new” responses to old items were used together to constitute the forgotten-trial category. “Probably old” responses were not included in order maximize the memory signal and to ensure that the number of remembered-trials was not significantly different from the number of forgotten-trials ($t(16) = -1.56, p > 0.10$) Networks constructed using beta-series connectivity have been shown to have the advantage of being more sensitive to variability in the shape of hemodynamic response compared with psychophysiological-interaction (PPI) analyses (Cisler, Bush, & Steele, 2014); this advantage that is particularly important for whole-brain connectivity analyses (Handwerker, Ollinger, & D’Esposito, 2004).

To measure functional connectivity between all regions simultaneously, the brain was first parcellated into 90 discrete anatomical regions of interest (45 ROIs in each

hemisphere and excluding all cerebellar nodes) defined in accordance with the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). This AAL template has been the most widely used nodal partitioning scheme in functional brain network analyses (Stanley et al., 2013), and it parcellates the cortex and subcortical structures by identifying gyral and sulcal boundaries. However, it is important to note that many other nodal partitioning schemes have been used in the literature, and the best possible method for defining nodes remains an open question (Stanley et al., 2013). Each anatomical ROI from the atlas served as a network node. Pairwise Pearson correlations between regional mean beta series were computed to generate $\{90 \times 90\}$ functional connectivity matrices, with the correlation coefficients representing the strength in connectivity between any two nodes. These $\{90 \times 90\}$ cross-correlation matrices (i.e., adjacency matrices) serve as the fundamental starting point for any graph theoretic analysis of neuroimaging data. In the current study, the adjacency matrices were not thresholded (i.e. weak connections were not removed), and each matrix constituted an undirected, weighted graph. Two separate whole-brain functional networks were created for each individual participant: a *forgotten network* constructed using the concatenated beta values from forgotten-trial category and a *remembered network* constructed using the concatenated beta values from the remembered-trial category. For display purposes, Figure 14 depicts remembered and forgotten adjacency matrices extracted from this approach, averaged across all participants.

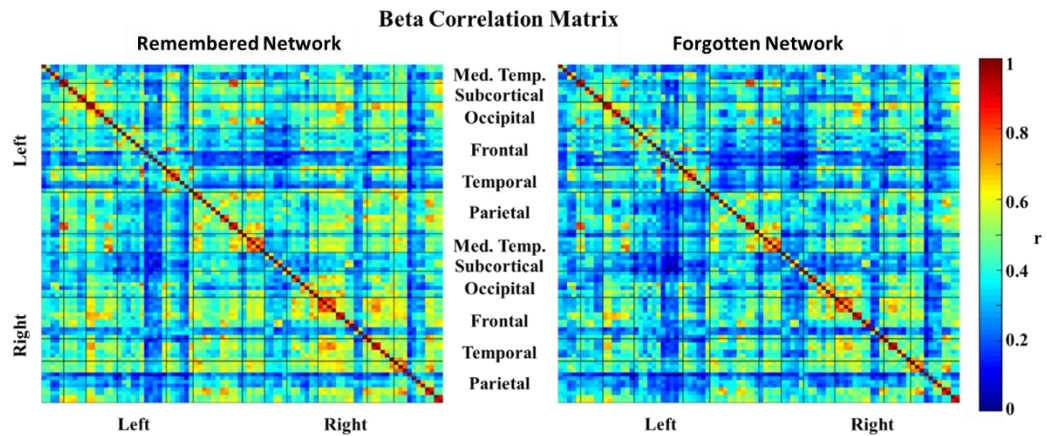


Figure 14. Average (across participants) adjacency matrices derived from beta series correlations are represented for remembered and forgotten networks and split by hemisphere (right and left). For ease of visualization, regions of interest are sorted in accordance with the procedure implemented by (Salvador et al., 2005).

3.2.2.5 Graph Theory Measures

3.2.2.5.1 Global Efficiency

Global efficiency is a measure of the capacity for efficient information transfer throughout the entirety of a network. It is closely related to another widely used graph metric—path length. When computed for an individual node in the network, path length is defined as the average shortest *distance* between that node i and all other nodes in the network. In weighted functional brain networks, connections with higher correlation strengths are considered to be *closer* together, whereas nodes with lower correlation strengths are considered to be *further* apart. A node's global efficiency is defined as the average of the inverse characteristic path length between that node and all other nodes in the network (Rubinov & Sporns, 2010), such that higher global

efficiency corresponds to shorter path lengths. In the formulation given below, d_{ij}^w represents the distance between two nodes in the network after taking into account the strength of each connection. The network-wide global efficiency is computed as an average of all nodal-wise global efficiency values.

$$E^w = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} (d_{ij}^w)^{-1}}{n - 1}$$

Negative connections were not included in the global efficiency analyses (for reasons noted in: (Cao et al., 2014; Telesford, Simpson, Burdette, Hayasaka, & Laurienti, 2011)). It is important to note that we are not using the term “distance” in a Euclidian sense. Distance here is a measure of how easily information may get from one node to another. Given this particular interpretation, using correlation strength as a measure of distance is reasonable, because information is more likely to travel (with less added noise) between two nodes that are strongly correlated.

<i>3.2.2.5.2 Degree Centrality</i>

Degree centrality indexes the overall strength of a given node’s connections by summing the weights of all their connections. Nodes with high degree centrality directly interact with many other nodes in the network and are likely to play an important role in the flow of information throughout the network. In order to better compare results between global efficiency and degree centrality, only positive connections were included

in the calculation of degree centrality for each network node. However, for all other computations, negative connections were retained in the functional brain networks.

3.2.2.5.3 First Step Reorganization (FSR)

Metrics such as global efficiency and degree centrality assess a node's capacity for information transfer without regard to how a node's connectivity profile changes between networks. As such, these measures are uninformative with respect to whether a node is connected to the same or different regions in different networks. However, *FSR* (Geib et al., 2017) can be used to partially disambiguate this by determining whether direct (first step) functional connections significantly change between conditions (e.g., remembered and forgotten retrieval networks). Mathematically, *FSR* is computed as follows:

$$FSR^w = zscore \left\{ -\operatorname{arctanh} \left(\frac{\operatorname{Cov}(A_i, B_i)}{\sigma_{A_i} \sigma_{B_i}} \right) \right\}$$

where A_i is the connectivity profile of node i in network A , B_i is the connectivity profile of node i in network B , and the connectivity profile of node i is defined as its distinct set of weighted connections (i.e., the column in the adjacency matrix representing that particular node's connections with all other nodes in the network). Due to the normalization procedure (z-score), the FSR_i^w value obtained for a given node is *always* relative to those *FSR* values obtained for the other nodes in the network. Traditional parametric statistics can then be applied at the group level in order to

identify nodes that consistently exhibit a reorganization of direct connections across participants.

3.2.2.5.4 Modularity (Q)

Modularity is computed using an optimization algorithm that identifies subsets of nodes that are more densely interconnected among themselves than with other nodes in the network (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Newman, 2006; Newman & Girvan, 2004). The extent to which the network can be subdivided into non-overlapping modules is quantified by the modularity Q statistic. Networks that can be clearly divided into non-overlapping modules, or functional communities, have larger Q values and are considered less integrated (i.e., more segregated) with the rest of the functional brain network. In contrast, networks that are not so clearly divided into distinctive functional communities have smaller Q values and are considered more integrated with the rest of the functional brain network. The modularity value assigned to a given partition of the entire functional brain network is:

$$Q = \sum_{i=1}^k \left[\frac{e_{ij}}{M} - \left(\frac{a_i}{M} \right)^2 \right]$$

where e_{ij} is a measure of within module connections in module i , a_i is the total degree (summed strength of all connections that a node has) of module i , and M is equal to the degree of the entire network. Modularity algorithms are designed to maximize the value of Q , by separating the network into non-overlapping subsets of nodes that

maximize within-module connectivity and minimize between-module connectivity. In the current study, in order to find the maximal value of Q , the weighted undirected Louvain algorithm was run 10 times on each individual subject's remembered and forgotten network (Blondel et al., 2008).

In contrast, a node's modular assignment was determined by running the Louvain algorithm 1,000 times on the averaged (across participants) remembered network. The averaged remembered network was used, as opposed to a general retrieval network that includes both remembered and forgotten trials, because the condition of interest is successful memory retrieval as opposed to a more general retrieval processing mode. The resulting 1,000 partitions were then divided into two groups based upon the number of modules discovered (i.e., networks with 5 modules and networks with 4 modules). These groups were then separately analyzed using the Jaccard Index to determine to the consistency of the module assignments across different iterations of the algorithm. The most consistently identified modular partition and the Q value itself were both considered in order to identify the best partition for further analyses.

3.2.2.5.5 Participation Coefficient

The positive participation coefficient measures the proportion of connections a node has within its own module versus other modules in the network (Guimerà & Nunes Amaral, 2005). Nodes with higher participation coefficients are more strongly

connected to nodes in other modules in the network, thereby facilitating communication between functional communities; in contrast, nodes with lower participation coefficients are predominantly connected to nodes within the module to which they were assigned.

Formally, a node's participation coefficient is defined as:

$$1 - \sum_{m \in M} \left(\frac{k_i^w(m)}{k_i^w} \right)^2$$

where M is the set of modules, and k_i^w is the weighted number of links between node i and all the other nodes in module m . The participation coefficient is computed independently for positive and negative connections, and we only report results from the positive participation coefficient. After we first identified the best modular partition using the *averaged* (across participants) *remembered network*, we then used that modular partition as a fixed template when computing participation coefficients within each individual participant's functional brain networks.

3.2.2.6 Retrieval Assembly Construction

In order to further investigate changes in graph theory metrics computed for the hippocampus, we identified a *retrieval assembly* defined as a set of regions (i.e., nodes) whose bivariate connectivity with the hippocampus was stronger during remembered than forgotten trials. These retrieval assembly nodes were identified by finding regions with a significant within-subject change in connectivity strength (alpha = 0.01, one-tailed, unc.) with the hippocampus. We intentionally used a relatively liberal, fixed alpha level, because these analyses were exploratory in nature. Additionally, compared

to the *remembered network* and the *forgotten network*, which were constructed for each individual participant and comprised all nodes in the 90-node AAL atlas in addition to all of their pairwise connections, the *retrieval assembly* is only comprised of the subset of those 90 nodes that alter their connectivity with the hippocampus between remembered and forgotten conditions. Consequently, there is only one *retrieval assembly*, because the assembly represents the subset of node-to-node connections that *change* between remembered and forgotten states.

3.2.2.7 Statistical Testing of Graph Theory Measures

We used the permutation framework developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index (Schubert, 2013; Schubert & Telecs, 2014) to assess significance for potential differences in each graph theory measure between remembered and forgotten memory retrieval conditions. This framework is based on the Jaccard Index (JI), which computes a ratio between two discrete data sets based upon the size of the intersection and union between those two sets. The permutation framework developed by Simpson et al. (2013) can accommodate continuous values (numeric values) and utilizes a permutation procedure to assess significance. This permutation framework is non-parametric, and we report median values and interquartile ranges (*IQR*) for global efficiency, degree centrality, modularity (*Q*), and participation coefficients (*PC*). Violin plots were used to visualize these distributions mapped on top of box plots. For all box plots the middle band represents

the median, the top of the box the 75th percentile, the bottom of the box the 25th percentile, and whiskers extend an additional 1.5*IQR. Outliers are depicted as black dots.

3.2.3 Results

3.2.3.1 Behavioral Results

Indicative of accurate memory performance, there was a significant difference between the accuracy of high-confidence correctly recognized items and false alarms ($t(32) = 10.7, p < 0.0001$), as well as between all correctly recalled items and false alarms ($t(32) = 9.9, p < 0.0001$). Average discrimination (d') across all correctly recalled items and false alarms was $M = 1.0$ ($SD = 0.5$). A summary of behavioral results is provided in Table 3.

Table 3. Provides a summary of the behavioral results. Only high-confidence hits were included in the remembered network, whereas both high- and low-confidence misses were included in the forgotten network. The forgotten and remembered networks were comprised of a similar number of total trials, on average (mean remembered = 54 (SD = 21; range = [25, 94]) trials, mean forgotten = 41 (SD = 15; range = [16, 68]), and not a significantly different number of total trials ($t(16) = 1.56, p > 0.10$).

	<u>New Items [Mean (SD)]</u>		<u>Old Items [Mean (SD)]</u>	
	<u>False Alarm</u>	<u>Correct Rejection</u>	<u>Miss</u>	<u>Hit</u>
High Confidence	27% (15)	73% (15)	19% (14)	81% (14)
Low Confidence	30% (10)	70% (10)	57% (13)	43% (13)
All Trials	28% (12)	72% (10)	35% (12)	65% (12)

3.2.3.2 Network Results

3.2.3.2.1 *Hippocampal integration is greater for remembered than forgotten items*

Our first hypothesis was that the hippocampus would show an increase in its capacity for efficient integration with the rest of the network for remembered compared to forgotten items. Consistent with our first hypothesis, there was a significant difference in global efficiency ($p = .02$) and degree centrality ($p = 0.04$) for the left hippocampus between remembered and forgotten retrieval networks, as revealed by the permutation procedure developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index. Only the global efficiency result remained significant after Bonferroni correction. However, in general, these results show that the left hippocampus exhibited a greater capacity for efficient communication with the rest of the network for remembered items compared to forgotten items. In contrast, no significant differences in global efficiency or degree centrality were identified for the right hippocampus (both p 's $> .10$) (Figure 15).

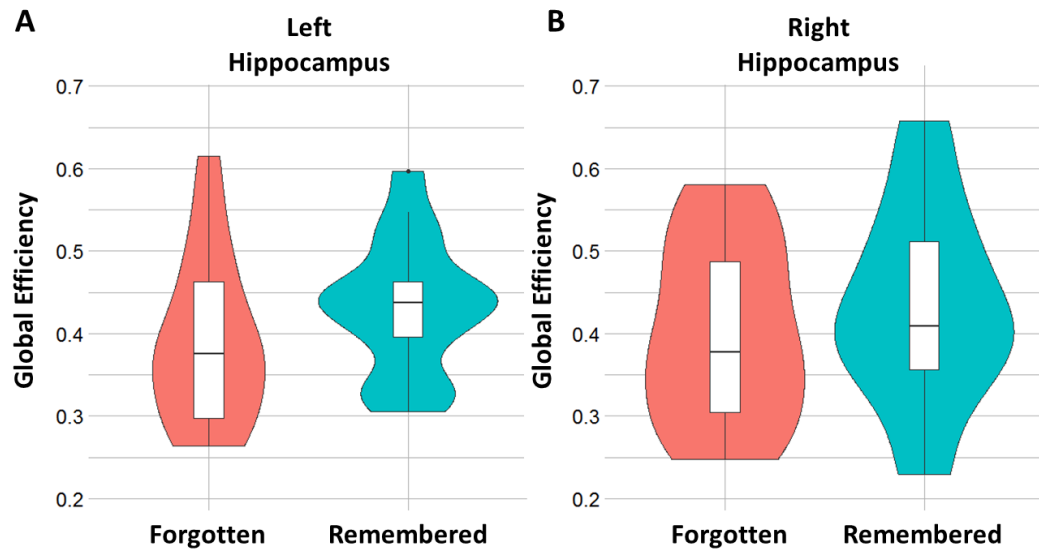


Figure 15. Global efficiency of the left and right hippocampus for remembered and forgotten trials, respectively. The permutation framework developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index revealed that left hippocampal global efficiency was greater in the remembered than the forgotten condition ($p < 0.05$), while the right hippocampus did not reach significance ($p = 0.17$).

The first step reorganization (*FSR*) analysis was used to clarify the extent to which direct changes in functional connectivity contribute to the observed changes in global efficiency and degree centrality for the left hippocampus. *FSR* was significant for the left hippocampus ($FSR = 0.63$, $t(16) = 2.37$, $p = 0.03$), indicating that the left hippocampus significantly reorganized its set of direct connections between remembered and forgotten conditions.

3.2.3.2.2 Successful memory is supported by a hippocampal retrieval assembly

Our second hypothesis was that successful memory retrieval would be associated with stronger connectivity between the hippocampus and a subset of nodes

comprising a “retrieval assembly.” To investigate this hypothesis, we identified a group of regions whose bivariate (i.e., direct) functional connectivity with the left hippocampus increased from forgotten to remembered items (Figure 16). A total of 12 brain regions showed an increase (within participants) in bivariate connectivity with the left hippocampus ($p < 0.01$, one-sided, unc.) from forgotten to remembered retrieval networks (Table 4). No node showed a decrease in connectivity with the left hippocampus from forgotten to remembered retrieval networks in accordance with our set alpha level ($p < 0.01$, one-sided, unc.). The 12 regions identified as members of the assembly included the bilateral dorsal superior frontal gyrus, left middle frontal gyrus, left caudate, bilateral supramarginal gyrus, left precuneus, bilateral precentral gyrus, left postcentral gyrus, right inferior temporal gyrus, and right middle occipital gyrus.

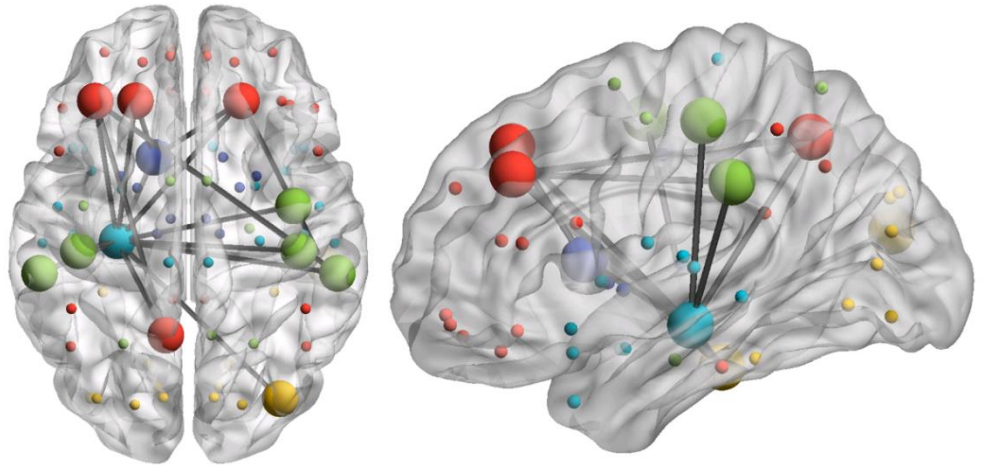


Figure 16. Provides an overview of the retrieval assembly and modular nodal assignments. All nodes in the assembly (depicted as the large nodes in the figure) showed significantly greater connectivity strength with the left hippocampus for remembered as compared to forgotten items (significant connections are depicted as dark lines; see Table 4 for further details). Significant connections within the assembly were determined by setting alpha to 0.01 (one-tailed, unc.) and removing deviant connection values ($SD > 3.5$). Five different functional communities (i.e., modules) were identified from the modularity analysis: occipital/temporal (yellow), fronto-parietal (red), parietal (green), medial temporal (light blue), and subcortical (dark blue). Note that the nodes comprising the retrieval assembly spanned all identified modules.

Table 4. Changes in hippocampal connectivity with nodes comprising the retrieval assembly. Nodes included in the assembly exhibited a significant increase in connectivity with the left hippocampus from forgotten to remembered retrieval networks ($p < 0.01$, one-sided, unc.). [Legend: L-CAU (left caudate), L-MFG (left middle frontal gyrus), L-SFGdor (left dorsal superior frontal gyrus), R-SFGdor (right dorsal superior frontal gyrus), R-MOG (right middle occipital gyrus), L-PostCG (left post-central gyrus), R-PostCG (right post-central gyrus), R-PreCG (right pre-central gyrus), L-pCUN (left precuneus), L-SMG (left supramarginal gyrus), R-SMG (right supramarginal gyrus), L-ITG (left inferior temporal gyrus)]

	Δ L. Hc Connectivity	
	Mean (SD)	<i>p</i> -value
L-CAU	0.14 (0.15)	0.0009
L-MFG	0.14 (0.17)	0.0021
L-SFGdor	0.16 (0.22)	0.0036
R-SFGdor	0.20 (0.21)	0.0007
R-MOG	0.15 (0.20)	0.0027
L-PostCG	0.12 (0.16)	0.0052
R-PostCG	0.12 (0.15)	0.0017
R-PreCG	0.11 (0.14)	0.0024
L-pCUN	0.12 (0.17)	0.0055
L-SMG	0.16 (0.25)	0.0094
R-SMG	0.15 (0.20)	0.0034
L-ITG	0.15 (0.22)	0.0085

Subsequent analyses using the permutation procedure developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index revealed that, on average, global efficiency and degree centrality were higher for nodes in the *retrieval assembly* for remembered compared to forgotten networks (global efficiency: $p = 0.03$; degree centrality: $p = 0.03$; Figure 17a). This was the case regardless of whether or not the left hippocampus was included as part of the assembly. In order to ensure that our finding was unique to nodes in the *retrieval assembly*, 1,000 random subsets of 13 network nodes were selected, and the average global efficiency of each randomly selected subset

of 13 nodes was computed. The *retrieval assembly* we identified did in fact exhibit a greater magnitude change in global efficiency between remembered and forgotten networks than any other randomly selected subset of 13 network nodes. This same procedure was implemented for degree centrality, but the observed change in degree centrality did not appear to be significantly greater than any other random subset of nodes in the network (rank 498 out of 1000). This suggests that while there was a significant change in degree centrality for the *retrieval assembly*, that there was a similar change with respect to other nodes in the network as well. Table 5 summarizes the global efficiency and degree centrality values obtained for each node in the *retrieval assembly* in both remembered and forgotten retrieval networks taken separately. Importantly, while it is possible that the observed increase in global efficiency or degree centrality for the retrieval assembly was driven by increased bivariate functional connectivity to the left hippocampus, such a finding would still underscore the importance of the left hippocampus in facilitating integrative processing via the assembly.

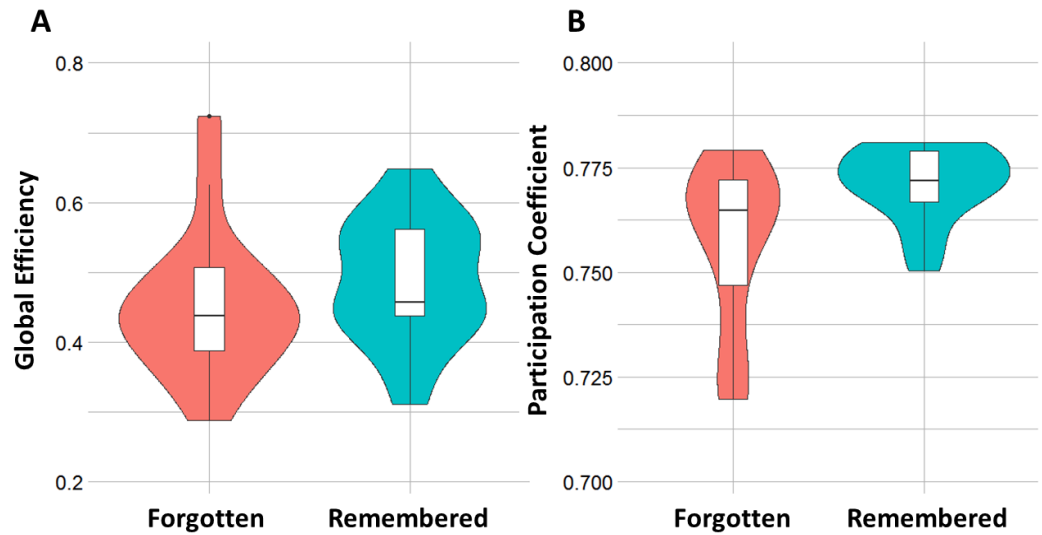


Figure 17. (A) Average global efficiency of all nodes in the retrieval assembly for both remembered and forgotten networks. The permutation framework developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index revealed that the global efficiency of the nodes comprising the retrieval assembly of the remembered network was greater than that of the forgotten network, regardless of whether or not the left hippocampus was included in the averaged global efficiency value ($p < 0.05$). (B) Average participation coefficient of all nodes in the retrieval assembly using the modular partition from the averaged remembered network as a template. The permutation framework developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index revealed that the average participation coefficient of nodes in the retrieval assembly in the remembered network was greater than that of the forgotten network, regardless of whether or not the left hippocampus is included in the average ($p < 0.01$).

Table 5. Global efficiency for all nodes included in the retrieval assembly. Statistical significance for changes in global efficiency between remembered and forgotten retrieval networks was assessed using non-parametric permutation tests using the framework developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index. [Legend: L-CAU (left caudate), L-MFG (left middle frontal gyrus), L-SFGdor (left dorsal superior frontal gyrus), R-SFGdor (right dorsal superior frontal gyrus), R-MOG (right middle occipital gyrus), L-PostCG (left post-central gyrus), R-PostCG (right post-central gyrus), R-PreCG (right pre-central gyrus), L-pCUN (left precuneus), L-SMG (left supramarginal gyrus), R-SMG (right supramarginal gyrus), L-ITG (left inferior temporal gyrus)]

	Global Efficiency: Median [IQR]			Degree Centrality: Median [IQR]		
	<i>Remembered</i>	<i>Forgotten</i>	<i>p-value</i>	<i>Remembered</i>	<i>Forgotten</i>	<i>p-value</i>
L-CAU	0.45 [0.38, 0.55]	0.39 [0.31, 0.46]	0.027	38 [28, 48]	31 [22, 38]	0.081
L-MFG	0.48 [0.40, 0.57]	0.45 [0.35, 0.49]	0.053	41 [32, 50]	37 [24, 43]	0.065
L-SFGdor	0.46 [0.44, 0.56]	0.44 [0.38, 0.51]	0.091	39 [33, 49]	37 [29, 43]	0.147
R-SFGdor	0.49 [0.46, 0.56]	0.43 [0.36, 0.52]	0.008	41 [38, 48]	35 [26, 44]	0.018
L-HC	0.44 [0.40, 0.46]	0.38 [0.30, 0.46]	0.021	36 [30, 41]	30 [21, 35]	0.039
R-MOG	0.43 [0.38, 0.61]	0.43 [0.36, 0.50]	0.308	36 [27, 54]	33 [25, 43]	0.524
L-PostCG	0.47 [0.43, 0.58]	0.44 [0.38, 0.53]	0.096	39 [34, 50]	32 [28, 47]	0.053
R-PostCG	0.53 [0.43, 0.57]	0.49 [0.39, 0.51]	0.080	43 [33, 51]	38 [28, 44]	0.139
R-PreCG	0.52 [0.41, 0.56]	0.48 [0.41, 0.52]	0.056	43 [33, 49]	36 [30, 43]	0.036
L-pCUN	0.53 [0.49, 0.63]	0.51 [0.45, 0.58]	0.259	44 [43, 56]	43 [36, 51]	0.090
L-SMG	0.45 [0.41, 0.54]	0.43 [0.35, 0.48]	0.098	36 [31, 47]	34 [25, 42]	0.064
R-SMG	0.46 [0.37, 0.51]	0.43 [0.34, 0.51]	0.352	37 [28, 46]	34 [21, 43]	0.256
L-ITG	0.53 [0.48, 0.60]	0.49 [0.41, 0.56]	0.038	46 [41, 53]	38 [32, 49]	0.024
Assembly	0.46 [0.44, 0.56]	0.44 [0.38, 0.51]	0.030	38 [36, 47]	35 [28, 42]	0.033

3.2.3.2.3 *Successful retrieval is associated with an increase in integration across the whole brain*

Our third hypothesis was that successful memory retrieval would be associated with increased integration (and a less segregated network architecture) across the entire functional brain network. To investigate this hypothesis, we computed modularity Q values for individual participants' brain networks. Q values represent the extent to which each network can be partitioned into densely interconnected communities of

functionally cooperating brain regions. Although it did not reach the threshold of significance, these modularity analyses revealed that the whole-brain network was somewhat more integrated (i.e., had smaller Q values) during remembered (Median = 0.11, IQR = 0.05) than forgotten trials (Median = 0.13, IQR = 0.08, $p = .07$), as assessed using the permutation procedure developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index (Figure 18a). While this difference in modularity did not reach significance, a subsequent individual-differences analysis revealed that the difference in the modularity Q statistic (remembered network Q minus forgotten network Q) was significantly and negatively related to memory performance assessed with d' ($r(15) = -0.72$, $p = 0.001$) (Figure 18b). Additionally, larger whole-brain modularity Q values in the forgotten networks were significantly related to better memory performance assessed with d' ($r(15) = .54$, $p = .03$), suggesting that individuals with a more modular forgotten network have better memory performance. However, whole-brain modularity Q values in the remembered networks were not significantly related to memory performance ($p > .10$).

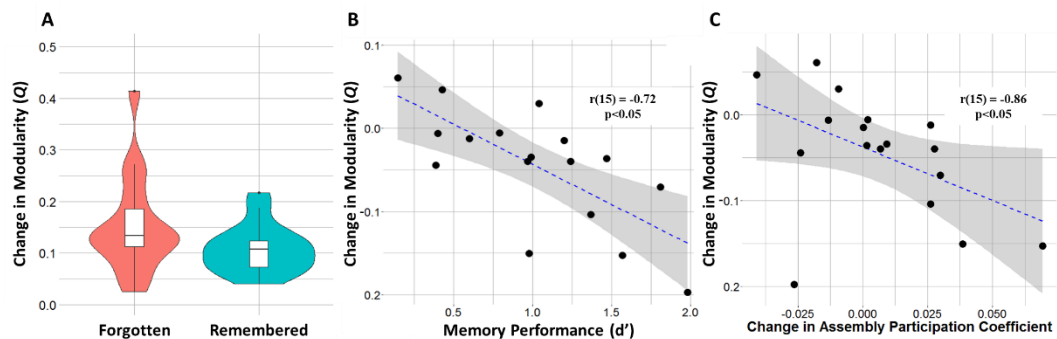


Figure 18. Provides a summary of the modularity findings. **A:** The permutation procedure developed by Simpson et al. [2013] in conjunction with the Jaccardized Czekanowski Index revealed that modularity Q values were somewhat greater for the forgotten network than for the remembered network ($P < 0.10$), but this did not reach significance in accordance with our set alpha level. **B:** The change in modularity between the remembered and forgotten networks was strongly correlated with individual variability in memory performance assessed with d' ($r(15) = -0.72$, $P = 0.001$). **C:** The relationship between the change in modularity (Q) and change in the retrieval assembly's PC for the remembered minus forgotten condition ($r(15) = -0.86$, $P < 0.05$).

In order to identify the best modular partition to serve as a template, the modularity algorithm was run 1,000 times on the *averaged* (across participants) *remembered network*. Of these 1,000 different runs, 90.3% had 5 modules and the remaining 9.7% had 4 modules. The JI revealed that the 5 module networks were more consistent between iterations in terms of the subsets of nodes comprising the different partitions than the 4 module networks (5 Module Network $M = 0.87$ ($SD = 0.12$); 4 Module Network $M = 0.82$ ($SD = 0.13$); $t(9310) = 20.1$; $p < 0.0001$). The maximum Q values were nearly identical between the 5 module and 4 module networks (5 Module Max $Q = 0.0623$; 4 Module Max $Q = 0.0624$). Given that the Q values were nearly identical

between 4 module and 5 module partitions, that the 5 module partition was identified more often than the 4 module partition, and that the 5 module networks was more consistently partitioned than the 4 module networks, the particular 5 module network with the highest Q value was used as the template partition for the *averaged remembered network*. While the *averaged remembered network* had a relatively low modularity, a series of random network rewirings (Rubinov & Sporns, 2010) revealed that modularity Q values in the observed networks were significantly greater than all Q values obtained from the corresponding randomized networks with equal density and degree distribution (1,000 permutations, range of Q values = [0.008, 0.0137]).

Interestingly, nodes comprising the *retrieval assembly* were distributed across all identified modules in the network (Figure 16). This suggests that the hippocampus better facilitates information transfer across multiple network modules to support the successful retrieval of words. While this observation is largely qualitative in nature, it is consistent with the notion that the role of the hippocampus in episodic memory retrieval is to facilitate information integration between diverse brain regions and to enhance the convergence and joint processing of information across different components of the network, each of which is presumably contributing particular functions that facilitate memory retrieval.

Using the modular partition identified from the *averaged* (across participants) *remembered retrieval network* as a fixed template, we computed the participation

coefficients of the nodes in the *retrieval assembly* for each of the individual participant's remembered and forgotten retrieval networks, respectively. While all nodes in the *retrieval assembly* had high participation coefficients in both remembered and forgotten retrieval networks, the permutation procedure revealed that the average participation coefficient of the assembly was higher in the remembered than the forgotten condition ($p = 0.008$; Figure 17b). This was the case regardless of whether or not the left hippocampus was removed from the *retrieval assembly* ($p = 0.005$ without the left hippocampus). While one might argue that this result arises from the fact that the modules were defined using the *averaged remembered network* as a template, this seems unlikely because the modularity algorithm is designed to minimize the connectivity between different modules; thus, one would expect that a node's between module connectivity (i.e. participation coefficient) would be *lower* in the network wherein the modular template was defined. A summary of differences in participation coefficients for nodes within the *retrieval assembly* between remembered and forgotten networks is presented in Table 6.

Table 6. Participation coefficients for all nodes included in the retrieval assembly. Statistical significance for changes in participation coefficients between remembered and forgotten retrieval networks was assessed using non-parametric permutation tests using the framework developed by Simpson et al. (2013) in conjunction with the Jaccardized Czekanowski Index. [Legend: L-CAU (left caudate), L-MFG (left middle frontal gyrus), L-SFGdor (left dorsal superior frontal gyrus), R-SFGdor (right dorsal superior frontal gyrus), R-MOG (right middle occipital gyrus), L-PostCG (left post-central gyrus), R-PostCG (right post-central gyrus), R-PreCG (right pre-central gyrus), L-pCUN (left precuneus), L-SMG (left supramarginal gyrus), R-SMG (right supramarginal gyrus), L-ITG (left inferior temporal gyrus)]

	Participation Coefficient: Median [IQR]		
	<i>Remembered</i>	<i>Forgotten</i>	<i>p-value</i>
L-CAU	0.786 [0.774, 0.789]	0.774 [0.762, 0.786]	0.047
L-MFG	0.759 [0.730, 0.765]	0.748 [0.669, 0.750]	0.001
L-SFGdor	0.767 [0.748, 0.771]	0.755 [0.739, 0.766]	0.074
R-SFGdor	0.767 [0.760, 0.777]	0.757 [0.744, 0.767]	0.035
L-HC	0.779 [0.767, 0.784]	0.771 [0.759, 0.783]	0.234
R-MOG	0.777 [0.764, 0.781]	0.776 [0.752, 0.778]	0.696
L-PostCG	0.781 [0.779, 0.785]	0.779 [0.757, 0.783]	0.006
R-PostCG	0.778 [0.776, 0.781]	0.776 [0.764, 0.778]	0.091
R-PreCG	0.780 [0.779, 0.783]	0.778 [0.766, 0.782]	0.057
L-pCUN	0.773 [0.767, 0.780]	0.773 [0.762, 0.777]	0.369
L-SMG	0.782 [0.774, 0.784]	0.774 [0.749, 0.781]	0.044
R-SMG	0.781 [0.769, 0.785]	0.771 [0.758, 0.778]	0.232
L-ITG	0.778 [0.769, 0.783]	0.773 [0.764, 0.777]	0.335
Assembly	0.772 [0.767, 0.779]	0.765 [0.747, 0.772]	0.008

Having found, for remembered as compared to forgotten items, that the modularity of the network decreased and that the *retrieval assembly*, which spanned all modules, exhibited a larger participation coefficient as a whole, we then hypothesized that these increased participation coefficient values for nodes in the *retrieval assembly* would be related to a decrease in modularity Q across the entire network. Accordingly, the *retrieval assembly's* change in participation coefficient was compared to the change in

the network modularity (Q) at the individual participant level. These measures were found to be closely related ($r(15) = -0.86, p < .001$; Figure 18c). Follow-up analyses were run to determine if this significant correlation was unique to the nodes comprising the *retrieval assembly*. 50,000 random subsets of 13 nodes were selected from the full network, and the change between remembered and forgotten retrieval networks in the averaged participation coefficient of each set of 13 nodes was compared to the change between remembered and forgotten retrieval networks in the modularity of the entire network. Thus, we obtained 50,000 correlation coefficient values, one for each subset of 13 nodes randomly selected. The observed retrieval assembly participation coefficient was among the most closely related to the change in modularity (top 5%).

3.2.3.2.4 Univariate

Univariate results were assessed for ROIs, as defined by the AAL atlas, and the univariate results were compared to graph theoretical results. This was done in order to ensure that graph theoretical results were not the product of differences in univariate activation. Within the *retrieval assembly*, no node exhibited a significant difference in univariate activation even before correction for multiple comparisons (Remembered minus Forgotten; $p > 0.10$). Additionally, none of the changes in graph theoretic measures were significantly correlated with changes in univariate activation or with each other even before correction for multiple comparisons ($p > 0.10$). Taken together, these results

suggest that the graph metrics captured unique features of the data above and beyond what can be gleaned from examining univariate activity.

3.2.4 Discussion

The purpose of this study was to investigate functional brain network architecture underlying episodic memory retrieval using graph theory measures at nodal, assembly, and whole-brain topological levels. This study yielded three main findings. First, the hippocampus displayed an increase in global efficiency and degree centrality for remembered compared to forgotten items, and it significantly reorganized its set of direct connections (first-step reorganization, FSR) between remembered and forgotten retrieval networks. Second, for remembered compared to forgotten trials, the hippocampus was more strongly connected with several network nodes comprising a retrieval assembly. This retrieval assembly included many regions previously linked to retrieval success, such as the prefrontal cortex and ventral parietal regions. Collectively, the retrieval assembly exhibited greater integration with rest of the functional brain network for remembered compared to forgotten trials. Third, at the level of the whole brain, modularity analyses revealed that the entire network exhibited a less segregated modular organization for remembered compared to forgotten items, suggesting that integrative processing throughout entire functional network supports the successful remembering of words. These three findings are discussed in separate sections below.

3.2.4.1 Hippocampal integration: increased global efficiency and degree centrality for remembered compared to forgotten items

Our results support our first hypothesis that the hippocampus would have a greater capacity for efficient integrative processing during remembered compared to forgotten conditions. The left lateralization of this effect builds upon results from previous univariate analyses showing that the left hippocampus tends to be more sensitive to the retrieval of verbal stimuli while the right hippocampus is more sensitive to pictorial stimuli (Papanicolaou et al., 2002). Our findings also complement recent multivariate functional connectivity results (Geib et al., 2017) showing that the right hippocampus exhibits a greater capacity for efficient information integration throughout the network for the vivid retrieval of scenes as compared to the dim retrieval of scenes.

The importance of segregated functioning during cognitive tasks has been extensively investigated using univariate activation methodologies. However, localized functional specialization alone cannot fully account for most aspects of brain function (Sporns, 2013). Integrated and distributed processes that subservise cognition may instead benefit from a higher capacity for efficient information transfer across the brain as a whole, especially for more complex cognitive tasks (Sporns, 2013; Stanley et al., 2015; van den Heuvel et al., 2009). Geib et al. (2017) demonstrated that the right hippocampus exhibits a greater capacity for global information integration to support the retrieval of scenes from memory. Similarly, our results here show that the left hippocampus exhibits a greater capacity for effective information integration to support the retrieval of

individual words from memory. To further clarify why the hippocampus exhibits changes in global efficiency and degree centrality when items are remembered as opposed to forgotten, we computed first-step reorganization (*FSR*), which is a recently developed measure for investigating the extent to which an observed change in nodal graph theory measures might be dependent upon a reorganization of direct functional connections in the network (Geib et al., 2017). There was a significant difference in *FSR* for the left hippocampus, indicating that successful memory retrieval involves a substantial reorganization of direct functional connections of the left hippocampus. This suggests that hippocampal changes in global efficiency and degree centrality are at least partly driven by the hippocampus reorganizing its set of direct functional connections.

Univariate activation changes were uncorrelated with all graph theory measures. However, changes in degree centrality and changes in global efficiency between remembered and forgotten networks were found to be highly correlated for all subjects (Spearman's $r(88) = [0.71, 0.98]$, all $p < .001$). While these measures do capture different topological properties of complex networks, it is clear that they are explaining much of the same variance. Future studies will investigate when and why these measures diverge in functional and structural brain networks.

3.2.4.2 Assembly organization: Successful memory is supported by increased hippocampal connectivity with a *retrieval assembly*

A retrieval assembly centered on the left hippocampus, which consisted of nodes with stronger connections to the left hippocampus for remembered compared to

forgotten conditions, contained many brain regions previously shown to be involved in memory retrieval operations using different methodologies, such as the PFC (Lundstrom, Ingvar, & Petersson, 2005; Wagner et al., 2005) and VPC (Cabeza et al., 2008). Additionally, the nodes identified as comprising the retrieval assembly were located in spatially disparate regions of the brain, complementing prior work showing that long-distance connections may be particularly important for facilitating healthy cognitive functioning across diverse tasks (Cohen, Sreenivasan, & D'Esposito, 2014; Liu et al., 2014; Schedlbauer et al., 2014; Wang, Li, Metzak, He, & Woodward, 2010).

The importance of the PFC for episodic memory retrieval has been well-established using diverse methodologies. PFC damage is associated with impaired episodic memory (Davidson, Troyer, & Moscovitch, 2006), and PFC activations are frequently observed during episodic retrieval and have been attributed to memory search and monitoring processes (Mitchell & Johnson, 2009; Reas & Brewer, 2013; Rugg & Vilberg, 2013; Wagner et al., 2005). In the current study, the bilateral superior frontal gyri showed increased bivariate functional connectivity and increased global efficiency for remembered compared to forgotten conditions. These increases in connectivity are likely due to the recruitment of these regions for memory search and monitoring to successfully retrieve items from memory.

Like the PFC, the VPC is commonly activated during episodic memory retrieval. According to one view, VPC activations reflect the capture of bottom-up attention by

recovered memories (Cabeza, 2008), whereas another account holds that these activations reflect the maintenance of recovered multimodal information within working memory (Vilberg & Rugg, 2008). The observed increase in functional connectivity between the hippocampus and the VPC for high-confidence hits is consistent with both views.

There is some limited existing evidence from different methodologies suggesting that the caudate is involved in memory retrieval processes. Hart et al. (2013) have suggested that semantic memory retrieval is dependent upon a pre-SMA, caudate, thalamus circuit. Additionally, this circuit has been shown to be active in tasks involving episodic memory retrieval in both fMRI (Bastin et al., 2012) and PET (Wiggs, Weisberg, & Martin, 1999) studies, and a circuit that just includes the thalamus and caudate has been shown to play a role in the retrieval of autobiographical memories (Burianova & Grady, 2007). Damage to the caudate has also been associated with poorer declarative memory for verbal information (Mizuta & Motomura, 2006). Our results suggest that the caudate may play an additional role in episodic memory retrieval by also serving a critical integrative function within the network, as evidenced by increased global efficiency, and a higher participation coefficient for remembered compared to forgotten conditions. These results indicate that the caudate has a greater capacity for efficient information transfer throughout the network when successfully retrieving memories. Additionally, given the known anatomical connections between the hippocampus and

caudate (Robinson et al., 2012) and between the caudate and frontal cortex (Lehericy et al., 2004), the caudate might serve as an intermediary node between the hippocampus and prefrontal nodes involved in retrieval control processes.

3.2.4.3 Assembly integration: Global efficiency increases and participation coefficient increases for remembered compared to forgotten items

For the subset of nodes comprising the *retrieval assembly*, global efficiency and degree centrality was higher, on average, for remembered than for forgotten items. This suggests that increased integration within the assembly contributes to memory success. While previous studies have reported that average whole brain global efficiency is higher during the retrieval of vivid compared to dim memories of scenes (Geib et al., 2017), here we narrowed our analyses to a level of organization between the nodal and whole-brain levels to show that the integrative properties of the *retrieval assembly* also increase from forgotten to remembered conditions. In the case of global efficiency, this increase was greater in magnitude than any other randomly selected subset of nodes.

Furthermore, we found that the average participation coefficient of nodes in the *retrieval assembly* was greater for the remembered than for the forgotten trials. This finding suggests that the nodes in the assembly are increasing their extra-modular functional connections. This kind of change in functional network architecture between remembered and forgotten conditions presumably facilitates integrative processing across many different functional communities in the network. Taken in conjunction with the hippocampal *FSR* result, this suggests that the overall hippocampal increases in

integration may be driven by changes both in direct functional connections (e.g. *FSR*) and in indirect functional connections via the *retrieval assembly*.

3.2.4.4 Whole brain integration: Decreased modularity for remembered compared to forgotten items

Modularity algorithms take into account the relative balance between integrative and segregated topological features within a network (Blondel et al., 2008; Newman, 2006). While the *retrieval assembly* results show that increased integrative processing among a small subset of nodes in the network was important for successful memory retrieval, modularity algorithms provide a statistic describing the extent to which the entire network exhibits a more integrative or segregated functional architecture. Our results demonstrate that whole-brain modularity decreased from forgotten to remembered retrieval networks, suggesting that successfully remembering items was associated with a relatively more integrated than segregated functional network architecture. Additionally, the differences in modularity Q scores (*remembered network Q* minus *forgotten network Q*) across individuals were inversely correlated with their memory performance. This corroborates and extends findings from previous work showing that better cognitive performance, especially for more complex tasks, is associated with a more globally integrated network architecture in diverse cognitive domains, including: working memory (Stanley et al., 2014; Stevens et al., 2012), odor recognition memory (Meunier et al., 2014), episodic memory for visual scenes (Geib et al., 2017; Westphal, Wang, & Rissman, 2017), and cognitive control (Braun et al., 2015).

Furthermore, we found that the nodes identified as comprising the hippocampal *retrieval assembly* span multiple, different modules. This suggests that the hippocampus is responsible for coordinating large-scale network interactions between differentially specialized and distributed subsystems. In fact, rapidly accumulating evidence has shown that the hippocampus coordinates with a diverse set of brain regions that support complex representations, including: encoding retrieval similarity in the occipitotemporal cortex (Wing et al., 2015), predictive coding in the extrastriate cortex (Hindy, Ng, & Turk-Browne, 2016), temporal sequences in the lateral PFC and medial PFC (Schapiro, Turk-Browne, Norman, & Botvinick, 2016), and attentional orientation in the retrosplenial cortex (Aly & Turk-Browne, 2015). Accordingly, these results suggest that the hippocampus alters its connectivity profile from forgotten to remembered retrieval conditions to better interconnect nodes in different modules. In this way, the hippocampus appears to facilitate the communication and the distribution of information between modules when items are successfully remembered as opposed to forgotten.

Post-hoc analyses also revealed that nodes within the *retrieval assembly* exhibited increased participation coefficients in accordance with an observed decrease in modularity (Q) for remembered compared to forgotten items. This suggests that the observed decrease in modularity Q for the entire network is closely related to the fact that nodes within the *retrieval assembly* increase their relative between module

connectivity. Previous work (Geib et al., 2017) has provided evidence that both the hippocampus and whole-brain become more integrated during successful memory retrieval. However, this work did not link these phenomena at a more intermediate level of network analysis. Here, we suggest that the *retrieval assembly* serves as the explanatory link between these phenomena. The *retrieval assembly* shows both increased integrative processing during successful memory retrieval and additionally facilitates between module communication across the network. Moreover, the extent to which the *retrieval assembly* increases its relative between module communication is closely related to the change in whole-brain integrative processing.

Investigating the properties of retrieval assemblies provides viable avenue for conducting future research on the neural instantiation of memory processes. Future work will investigate which components of the retrieval assembly are common across different kinds of retrieval tasks, which kinds of information are being shared between modules, and the temporal sequence of events using EEG data (e.g. does the *retrieval assembly* or hippocampus increase its efficiency first?). It is only through the discovery of this intermediate level of organization that we can fully parse out how network interactions influence memory retrieval.

3.2.4.5 Conclusions

Recently, there has been growing interest in identifying network nodes that facilitate the integration of information between segregated, specialized communities of

brain regions in order to support more complex cognitive processes (Bertolero, Yeo, & D'Esposito, 2015; Fornito, Zalesky, & Breakspear, 2015; Stanley et al., 2014). Network nodes that interconnect many different network modules play a key role in subserving more complex cognitive tasks by allowing groups of nodes to perform specific, specialized functions yet still effectively interact with other brain regions performing other kinds of functions. Here, we incorporated a novel analysis method to show how a critical network node can change between two conditions of interest to better facilitate the integration of information between modules. Using this approach, we showed that the hippocampus occupies a more topologically critical position in the network to facilitate the flow of information between network modules for successful compared to unsuccessful memory retrieval. To facilitate the successful retrieval of items from memory, the hippocampus became more strongly connected with sensory motor regions, attention regions, frontal control regions, higher order visual regions, subcortical regions and other default mode regions such as the precuneus. In general, to support successful episodic memory retrieval of semantic information, the hippocampus is likely to be strongly connected with prefrontal regions that mediate retrieval control process (Rugg & Vilberg, 2013; Wagner et al., 2005), ventral parietal regions serving attention functions (Cabeza et al., 2008), higher-order visual regions supporting sensory reactivation (Danker & Anderson, 2010), and the precuneus supporting internal mentation (Kim, 2016). It is through the complex but coordinated interactions involving

these regions that we are able to retrieve information from memory. Although our sample size is relatively small, our results are highly consistent with and complementary to this prior literature investigating the neural bases of memory using diverse methods.

Taken together, these findings suggest that memory retrieval is dependent upon a broadly distributed yet interconnected set of brain regions, as opposed to a single encapsulated region or set of encapsulated regions that can be identified using univariate activation analyses. Notably, evidence was presented for increased integrative processing at the hippocampal, retrieval assembly, and whole-brain topological levels to support the successful remembering of words. While we maintain that the hippocampus is a critical structure for memory retrieval, the functionality of the hippocampus itself is only partly responsible for our ability to successfully retrieve items from memory. The hippocampus does not function exclusively on its own as an isolated, encapsulated brain structure. Identifying the ways in which the hippocampus is directly and indirectly functionally interacting with other brain regions to facilitate episodic memory retrieval provides a more complete account of the neural processes underlying this critical cognitive function.

4. Single-item representations

4.1 The science of the singular: single-item decoding with multivariate pattern analysis

Over the past decade, multivariate pattern analysis (MVPA) has become a widely used and highly useful analysis approach in fMRI. In most MVPA applications, a machine-learning classifier is trained to differentiate between the fMRI responses to two classes of stimuli (e.g., faces and places), each consisting of dozens of individual exemplars. Here, we propose a novel MVPA approach based on logistic regression that can reliably decode the fMRI responses to single items, presented only once, even when that single item is in a dataset consisting of hundreds of items. In our analyses, we demonstrate two key findings related to this approach: [1] We demonstrate reliable single-item classification within scenes, objects, and words and additionally demonstrate that it is possible to classify single items from among highly similar items, e.g., discriminating a cardinal from among other bird exemplars. We also provide preliminary evidence of cross-modal decoding, e.g., distinguishing object-word stimuli using a classifier trained on corresponding object-picture stimuli (and vice-versa). [2] Using representational similarity analysis (RSA), we demonstrate that the classification evidence for a to-be-decoded item is typically greater for items within the same category, as opposed to those in different categories. The greater the degree of this increased within category similarity, the lower the within-category classification evidence, suggesting that the presence of gist-based information is detrimental to robust

single-item decoding. By demonstrating the ability to classify individual items, as opposed to categories, we greatly increase the specificity of MVPA and in doing so permit the design of more naturalistic paradigms with more diverse stimulus sets. As such, we believe this technique will be highly useful for a wide-array of cognitive neuroscience applications.

4.1.1 Introduction

Since its initial application, multivariate pattern analysis (MVPA) has burgeoned in popularity, quickly become a staple fMRI analysis technique (Haxby, Connolly, & Guntupalli, 2014; Norman, Polyn, Detre, & Haxby, 2006) across a wide range of cognitive neuroscience realms, including cognitive control (Bhandari, Gagne, & Badre, 2018), memory (Kuhl, Rissman, & Wagner, 2012; Kuhl, Rissman, Chun, & Wagner, 2011), language (Musz & Thompson-Schill, 2017), emotion (Kragel & LaBar, 2014), working memory (Rose, LaRocque, & Postle, 2017) and attention (Sprague & Serences, 2013). A large reason for this technique's popularity is the ability to distinguish brain responses to different types of stimuli, which many researchers have utilized to draw inferences about the neural mechanisms underlying cognitive functions. Up until now, however, MVPA has largely been limited to distinguishing between *categories* of stimuli (e.g., faces vs. places), as opposed to *individual* stimuli, which precludes nearly all analyses that require sensitivity to variation in basic-level concepts (e.g., variability between places). In this analysis, we present an approach that provides greater specificity by extending

the framework of MVPA to individual stimuli, allowing researchers to probe stimulus representations and their underlying cognitive processes with higher specificity.

In general, MVPA has most often been applied within-subject, as classification accuracy is generally higher when brain structure is constant across a comparisons (Haxby et al., 2011). While between-subject classification is also possible by using a leave-one-person-out framework (LOPO)(Poldrack, Halchenko, & Jos, 2009), classification accuracy is often lower, as it is hindered by between-subject brain variability. While hyperalignment procedures can improve between-subject alignment (Haxby et al., 2011), this procedure requires additional scanning sessions, and typically relies on consistent and balanced sample of discrete classes of items (Guntupalli et al., 2016). For both within-subject and between-subject decoding, classification of single-items is at the category-level (e.g. faces versus places). For example, individual places such as the Grand Canyon, Eiffel Tower, etc., are all regarded as location exemplars from the overarching category of locations. This classification framework is feasible when there are many exemplars, which are cleanly divisible between categories and/or when items are presented multiple times. However, also note, that this framework emphasizes the commonality of items (e.g. different places) within a category, as opposed to focusing on the individuality of specific items (e.g. Grand Canyon).

Researchers are often, however, interested in the individuality of items – for example – when studying human memory or language comprehension. Other times, the

data lacks a clear categorical structure (e.g. places versus faces) and/or a sufficient number of within class-stimuli for standard categorical MVPA analyses. Our goal was to address a specific but important version of this case, one where the aim is to be able to decode briefly presented individual items based upon a single-presentation. Previous work in this domain has focused on movie clips, which are perceptually very rich in detail and unfold over long temporal durations (Zadbood, Chen, Leong, Norman, & Hasson, 2017). In contrast, here we focused on event-related data from two paradigms consisting of words and pictures of either scenes or objects. In doing so, our goal was to demonstrate the feasibility of single-item decoding for event-related designs that utilize common discrete stimuli (i.e., pictures and words).

Because most MVPA analyses address binary- or tertiary-class problems, misclassifications are of little interest, as the misclassification is *always* associated with the other left-out class. However, our classification problem requires a distinction between hundreds of stimuli, and thus we were interested in the shape of the misclassifications as represented by a confusion matrix. A confusion matrix represents the probabilities of a to-be-decoded item in relation to all other items in the test set, and thus in multi-class problems it provides information on *how* the classifier both succeeds and fails. For example, when classifying emotional states (Kragel & LaBar, 2014), the confusion matrix contains information pertaining to how frequently the decoding for different emotions are confused with one another – say, for example, amusement and

contentment. *If* the decoding for amusement and contentment emotional states are frequently confused with one another, it suggests the representations of the emotions are difficult to differentiate (i.e. are similar), whereas if they are never confused with one another it suggests they are easy to differentiate. In this way, the confusion matrix is illuminating with respect to how the classifier differentiates between different stimuli.

There are a variety of algorithms available for MVPA, and neuroimagers have utilized many of these, including support-vector machines (Cox & Savoy, 2003), linear-discriminant analysis (Carlson, Schrater, & He, 2003; O'Toole, Jiang, Abdi, & Haxby, 2005), and artificial neural networks (Polyn et al., 2005). For our analyses, we opted for a logistic regression with a leave-one-person-out framework (**LR-LOPO**) due to its linear nature and the ability to convert classification estimates to posterior probabilities. We found the convertibility to posterior probabilities particularly important, as our analyses have hundreds of stimuli and it would therefore be quite difficult for the classifier to select the *exact* target stimulus. However, we do expect that the classifier should have the capacity to provide a good guess, such that, even if the first guess is wrong, the second has a good chance of being correct, which in a numerically large-class problem is still highly significant, given that chance probability can be very low (e.g. 1/300).

For our analyses, we formulated two primary goals. Our first goal was to investigate the possibility of decoding individual items, both **intra-modally** and **cross-modally**, even when the to-be-decoded item is amongst other highly similar items. Intra-

modal classification is decoding which proceeds within-modality, e.g. decoding pictures of landscapes that subjects have seen based upon other pictures of landscapes that they have seen – this is the procedure most commonly utilized with standard MVPA. Cross-modal decoding is decoding across-modality, e.g. decoding one stimulus format (e.g. scene pictures) from a parallel format capturing the same concept (e.g. descriptive text labels). This style of classification is substantially more difficult. However, evidence for cross-modal representations in the brain has been reported (Devereux, Clarke, Marouchos, & Tyler, 2013; Fairhall & Caramazza, 2013), and we hypothesized that the LR-LOPO algorithm utilized would also be capable of detecting these cross-modal properties. The distinction of a single item from among highly similar items is also *critical* in accessing our ability to decode *individual* items. An inability to do this suggests that the classifier is decoding categorical level details (e.g. a cardinal is a bird) rendering it little better than a categorical classifier, whereas the ability to decode within-category suggests the algorithm is deciphering item-specific details (e.g., a cardinal is a cardinal but not a bluebird).

Our second goal was to investigate whether the structure of LR-LOPO misclassifications bears significant resemblance to hypothesized theoretically-derived stimulus similarity matrix. More specifically, we tested whether the confusion matrix derived from the LR-LOPO analysis resembles a stimulus-similarity matrix, and whether the degree of this similarity reduces the overall classification accuracy of a to-

be-decoded item. The ability of the classifier to discriminate between the representations of highly similar items (e.g. cardinal vs. other birds) is critical; however, we also thought it likely the stimulus structure of the data would influence classification accuracy.

Notably, we hypothesized that misclassifications would be highest for within-category items as opposed to across category items - that is, when classifying a cardinal, we expected overall evidence to be higher for other birds as opposed to say mammals. In conjunction with this, we conversely hypothesized that greater within-category evidence might also hinder the classification of the to-be-decoded item. For example, if there is high evidence for birds, the classifier might have difficulty in identifying the to-be-decoded item as a cardinal, as the classifier might be deciphering more generic categorical-level features. As such, we use representational similarity analyses to determine how the inherent similarity structure of the data influences single-item decoding accuracy.

4.1.2 Methods

4.1.2.1 Paradigms

The current approach utilized data from two previously scanned memory-reactivation paradigms (Figure 19). In the **scene-memory paradigm** (Wing et al., 2015), during encoding 21 participants (12 female) were shown 96 individual scenes, with scene labels, and asked to assess the typicality of each scene on a four-point scale. During the memory retrieval phase of this paradigm, participants were presented with

scene-labels alone and asked how vividly they recalled each scene on a four-point scale. In the **object-memory paradigm** (Davis, Geib, Wing, Wang, Hovhannisyan, et al. 2020²), 19 participants (14 female) were shown 360 unique objects. During memory encoding, each object was preceded by a brief letter, with the task to indicate whether the letter matched the object's name, which was the case for 300 of the objects. The 60 remaining objects were preceded by mismatching letters and were not included in these analyses. During the memory retrieval phase of this paradigm, participants were shown 400 object names, 300 matching the encoded objects and 100 new-object names, with the task to indicate whether the associated word was old or new and how confident they were of that assessment (definitely old, probably old, probably new, definitely new). While the aforementioned papers (Davis et al. 2020 & Wing et al. 2015) examined memory-related effects, for the current analysis, no memory effects were examined, as the objective of this analysis was to assess the feasibility of single-item decoding across participants.

² Currently under review, for further details pertaining to experimental design and preprocessing see Appendix A

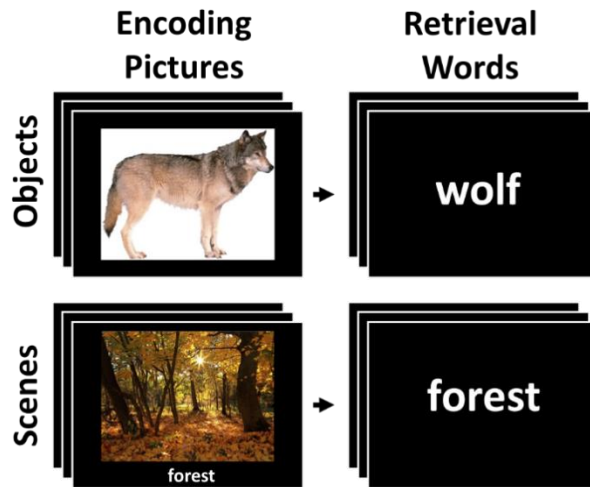


Figure 19. Format of the two studies included in the analyses

4.1.2.2 Stimulus-Similarity Matrices

In the **scene paradigm**, a representational **scene-similarity matrix** was created by assessing the dominant features (e.g., Beach = {sand, ocean, sun, etc.}) of each individual scene and then computing the Jaccard overlap between all scene pairs (i.e. $A \cap B / A \cup B$). A **semantic-similarity matrix** was created for the object study by assessing the feature overlap across all 300 items using a previously collected database of semantic feature norms, obtained in a separate normative study (see Davis 2020 et al., for details). In the scene-similarity and semantic-similarity matrix, similarity values are continuous, ranging from 0 to 1, with a 0 indicating that two items share no features and a 1 indicating that two items share all features. In addition to this **continuous similarity matrix**, a **categorical similarity matrix** was created containing only values 0 and 1, with a 1 indicating a pair of items belong to the same category, and a zero indicating a pair of items belong to different categories. For the scene study, this was based upon the

category of the scene (indoor/outdoor), whereas for the object study, this was based upon the object-category (e.g., birds, buildings, vehicles) of which there were 12.

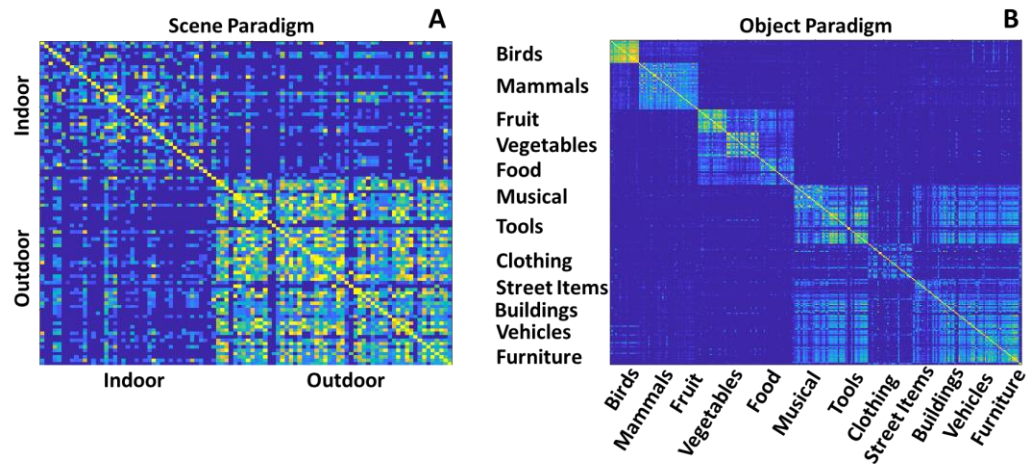


Figure 20. Similarity structure of the two studies. (A: Left) Feature similarity for the scene paradigm and (B: Right) Semantic similarity for the object paradigm. Note the clear categorical structure inherent to both studies.

4.1.2.3 fMRI Acquisition and Processing

For both paradigms, imaging data was collected using a 3T GE scanner.

Following a head-localizer scan, functional images were acquired using a SENSE spiral-in sequence (time repetition= 2000ms, time echo = 30 ms, field of view= 24 cm, 34 oblique slices with voxel dimensions of $3.75 \times 3.75 \times 3.8$ mm). Presented stimuli were projected onto a mirror at the back of the scanner bore, and responses were recorded using a 4-button fiber optic response box. Scanner noise was reduced with earplugs, and head motion was minimized with foam pads. A high-resolution anatomical image (96 axial slices parallel to the anterior commissure-posterior commissure plane with voxel dimensions of $0.9 \times 0.9 \times 1.9$ mm) was collected following functional scanning.

Scene Paradigm: Functional data were collected during six task runs of equal length (6 minutes and 40 seconds) during the performance of the scene paradigm described above with an average inter-trial interval of 12 seconds. A resting-state scan lasting for 5 minutes was collected following the third run but was not used in the current analyses.

Object Paradigm: Functional data for the object paradigm were collected during two task runs of equal length (14 minutes and 20 seconds) during the object paradigm described above, with an average intertrial interval of 4.5 seconds. Resting state scans lasting 5 minutes were collected before and after the picture presentation runs, but are not used in the current analyses.

Preprocessing was done using SPM software (www.fil.ion.ucl.ac.uk/spm). All the data were high-pass filtered and motion corrected, and each individual's brain map was registered to standard MNI space. During first-level general linear modeling (GLM) for the scene paradigm, scenes and words were coded as independent events (Rissman et al., 2004). The object paradigm utilized a fast event-related design, and as such we ran a family of first-level models (GLMs) such that each model contained the stimulus of interest and all other items (Mumford et al., 2012) - i.e., when modeling the encoding data, for each subject we ran 360 GLMs, one for each stimulus in the encoding set with all other items being included in a nuisance regressor. The resulting beta-estimates from each of these studies were then converted to t-values based upon the model residuals.

The resulting t-values, one for each subject and each stimulus, served as the basis for classification analyses. For both studies, analyses were restricted to a set of brain-volumes comprised of the 90 cortical ROIs in the AAL atlas (i.e., excluding the cerebellum and the CSF).

Further details pertaining to scanning parameters and experimental design can be found in the original publications (Davis et al. 2019 & Wing et al. 2015).

4.1.2.4 Logistic Regression (L-2 norm)

MVPA analyses utilized logistic regression (LR) as LR models output probabilistic values - i.e. the output might suggest there is a 15% chance the to-be-decoded stimulus is a cardinal, 7% chance a bluebird and so forth, such that the sum of all probabilities equals 100%. As is typical when running LR models, we utilized independent training and testing sets, in a leave-one-person-out framework (Poldrack et al., 2009) (i.e., we trained on all subjects, except one, and then tested our trained classifier on the left out subject) and a one vs. all classification schema (i.e., we trained on one to-be-decoded stimulus vs. all other stimuli). A one vs. all format was utilized, as a one vs. one framework (training and testing between *all* pairs of items) is computationally prohibitive - e.g., a 300-class problem requires $300^2 - 300 = 89,700$ models per left-out subject. Furthermore, we utilized data from the whole-brain when running the classification analyses and counted on the L2-norm to minimize the influence of regions that did not code for the stimulus of interest. Finally, we ran four models within

each paradigm: comparing pictures to pictures, pictures to words, words to pictures, and words to words.

It is also important to note that LR models contain a single hyperparameter, lambda, which determines the degree to which misclassifications are penalized: minimal penalization (high lambda value) prevents overfitting, whereas high penalization (low lambda value) prevents underfitting. Neuroimaging studies have used a wide range of lambda values – ranging from 0.01 to 100 [0.01; (Gordon et al., 2014), 0.05; (Thakral, Wang, & Rugg, 2017), 25; (Emrich, Riggall, LaRocque, & Postle, 2013); 100; (Kuhl et al., 2012; Kuhl et al., 2011)]. In our analyses we utilized values towards the ends of this range (0.05 and 100) in order to provide insight into what hyperparameter values would lead to the highest classification accuracy. Ideally, a validation set (independent of the training and testing set) should be used to determine lambda; however this is difficult when the number of observations is small (i.e., one per subject), as is the current case.

In addition to varying lambda, we also tested the effect of normalizing the data prior to LR training and prior to LR testing. Normalization is a common procedure in MVPA analyses e.g. (Kuhl, Bainbridge, & Chun, 2012), and we wanted to determine which procedure would lead to the best classification performance. For our normalization procedure, for each subject, we created a surrogate dataset with the same mean and variance across trials and voxels, henceforth described as a bias (see Figure 26 for more details). We then subtracted this bias from the subject's real data, which results

in a new matrix where the mean across all trials and voxels is zero. This procedure is mathematically akin to de-meaning the data simultaneously across 2-dimensions while also preserving the variance associated with each voxel and trial.

In order to find the best decoding model, we varied the lambda parameter and normalization in each train/test comparison (e.g., pictures to pictures) – resulting in 32 total models (4 train/test comparisons {picture-to-picture, picture-to-word, word-to-picture, word-to-word} x 2 lambda parameters {0.05, 100} x 2 training normalization {yes, no} x 2 testing normalization {yes, no}) for each paradigm. All subsequent analyses utilized the resulting confusion matrix from these LR-LOPO models.

4.1.2.5 Confusion Matrix

In MVPA analyses, misclassifications are often represented in a 2-dimensional confusion matrix, where one dimension consists of the class labels from the training data and the other dimension consists of the to-be-decoded items (i.e. test data). If the classifier has perfect decoding accuracy, then the off-diagonal should be zero, i.e., no items are misclassified; as such, the off-diagonal values typically reflect the frequency of misclassifications between items and can provide information on how confusable different classes are. In our analysis, we classified individual items, and thus we have no measure of frequency. However, because we use an LR-model, we have continuous probability measures for all to-be-decoded items - e.g., a 2% likelihood of item one, a 3% likelihood item two, etc.. When applied to all to-be-decoded items, this results in square

matrix, where the number of rows and columns is equal to the number of the to-be-decoded items. Each cell in this matrix then represents the likelihood of a to-be-decoded item to another item e.g., the likelihood that the to-be-decoded item (e.g., cardinal) is a bluebird.

4.1.2.6 Analyses for Decoding Individual Items

4.1.2.6.1 Determining classification accuracy

Multiclass problems of this magnitude can be evaluated in a variety of ways. We chose two metrics to evaluate classifier performance. Our preferred method considers how the confusion matrix ranks each to-be-decoded item with respect to a specified level of chance. For example, if we consider a chance threshold to be 50%, and the decoded item is ranked 200 out of 300 in terms of likelihood, we would consider its above-chance accuracy to be 17% (i.e., $200/300$ minus 0.50), thus being a correct classification [value = 1] whereas if its above-chance accuracy is negative we consider the classification incorrect [value = 0]. If ranks were tied (e.g., multiple items had a probability of 1%), we computed an average rank for the item. For this measure, we utilized a chance threshold of 50% and refer to this threshold throughout the manuscript with respect to accuracy (i.e. **Accuracy > 50%**), which represents the proportion of items correctly decoded based upon this chance-value.

In addition to this measure, we also examined the posterior probability of the to-be-decoded item as compared to chance. For example, if the posterior probability of the to-be-decoded item was 0.02 and we use a 300-class classifier, then the above-chance

classification accuracy is 0.017 (i.e., 0.02 minus 1/300). This is a more conservative measure of decoding accuracy as it ignores scenarios in which a to-be-decoded item is slightly below chance, but still far above all other items in the set. While we primarily utilized Accuracy >50% to be chance threshold in this manuscript, we also include analogous results from this **Post-prob > Chance** measure when appropriate.

4.1.2.6.2 *Decoding between highly similar stimuli*

The known categorical structure of the stimuli used in the current analysis allows us to ask whether we can decode between highly similar items within a category. For example, while the classifier might succeed in the 300-class problem, it might be doing so because it deciphers gross categorical structure—e.g., the classifier identifies the to-be-decoded item is a mammal, but it cannot reliably distinguish it from any of the other mammals in the dataset. To verify that we could reliably distinguish between like stimuli, we ran additional analyses where we considered **Accuracy >50%** *within* each category independently. For example, in the case of mammals, we would compute this measure by determining the item's rank in comparison to other mammals -- e.g., if the item rank is 15 and there are 20 mammals then Accuracy >50% would be 25% (i.e., 15/20 minus 0.50).

4.1.2.6.3 *Cross-modal Decoding*

To test for cross-modal classification we trained the classifier on one modality (e.g., words) and tested on other modality (e.g., pictures) and vice-versa. The assessment of the classification accuracy used the two measures previously discussed: Accuracy >50% and Post-prob > Chance.

4.1.2.7 **Analyses for Measuring Similarity Effects**

4.1.2.7.1 *Representational Similarity Analyses*

Confusion matrices can be thought of as representing the pair-wise similarity between all items in the stimulus set, e.g., the likelihood that a cardinal is a bluebird. As such, a confusion matrix is akin to a representational similarity matrix and is thus suitable for RSA (Kriegeskorte et al., 2008). RSA analyses compare a data-derived similarity matrix (the confusion matrix) to a hypothesized theoretically-derived similarity structure (e.g., a stimulus-similarity matrix based on semantic features). If these matrices are similar, it suggests that hypothesized similarity structure is similar to the derived similarity structure. In our case, such a positive relationship would suggest that classifier misclassifications are most typically to highly similar items, as opposed to random items. As mentioned, the stimuli in our studies have both a categorical (Scene paradigm: indoor or outdoor; Object paradigm: birds, mammals, street items, etc.) and continuous (Scene paradigm: feature; Object paradigm: semantic) structure. As such, to

test our hypothesis that the data-derived confusion matrices are similar to the hypothesized categorical and continuous matrices, we computed the 2nd order correlation (Spearman) between the **confusion matrix** and either the **continuous** or **categorical** similarity matrices for each subject. These analyses critically ignored the identity diagonal, such that a positive relationship between data-derived similarity, and hypothesized theoretically-derived similarity were driven only by misclassifications and not inflated by correct classifications. The resulting values from this operation were then Fischer transformed and entered into a t-test.

4.1.2.7.2 *How similarity influences decoding accuracy*

The final analysis sought to determine how, at the single-trial level, similarity structure influences decoding accuracy. Expanding on the initial hypothesis, we hypothesized that a strong categorical structure in the confusion matrix would boost *overall* classification evidence – that is, the ability to decode a cardinal should be higher *if* there is strong evidence that the to-be-decoded item is a bird. The suggestion here is that if the classifier deciphers the categorical structure in the data, it is likely that the to-be-decoded item is highly ranked with respect to other items. Alternatively, higher overall category evidence might make individual items from within the category less differentiable, such that high category evidence hinders within-category classification as well. In this case, strong evidence that the to-be-decoded item is a bird, might make it

more difficult for the classifier to discriminate what exact bird the to-be-decoded item is, e.g., a cardinal. The following analysis directly addresses these specific questions.

When examining how similarity structure influences decoding accuracy we utilized a regression framework (Equation 1) to model how **probability above chance** is influenced by the 2nd order similarity between the confusion matrix and the stimulus-similarity matrix. The variable probability above chance, is *not* identical to Accuracy > 50%. When computing Accuracy > 50% the data was binarized, whereas, for the single-item analysis, we retained the actual position - i.e., the to-be-decoded item was in the top 5% of guesses. To try and make these results more comparable to the previous Accuracy > 50%, we transformed this value $[(1 - \text{proportion}) - 0.50]$ such that if an item was in the top 5%, its associated value in the regression equation would be $[(1 - 0.05) - 0.50] = 0.45$ (e.g., 45% higher than a chance threshold of 50%). This same analysis was run for within-category classification, to examine how the stimulus similarity structure affects the ability to decode within category. When running this additional analysis, a category variable was added to Equation 1 [i.e. + (1 | Category)]. For both these trial-wise regression analyses, the 2nd order correlation was computed in a trial-dependent fashion by only including one column of the similarity matrix and one column of the confusion matrix, such that the 2nd order correlation reflects how well that individual trial fits the similarity matrix.

Equation 1. Addresses how similarity between the stimulus-similarity matrices and the confusion matrix influences classification accuracy. CM refers to the data-driven confusion matrix, whereas SSM refers to the hypothesized theoretically-derived stimulus similarity matrix – with Cont. referring to the continuous matrix and Cat. referring to the categorical matrix.

$$Prob. \text{ above Chance} \sim Corr(CM, Cont. SSM) + Corr(CM, Cat. SSM) + (1|Subject)$$

4.1.2.8 Single Item Decoding Code Repository

All LR-LOPO analyses were conducted in MATLAB and associated code can be freely accessed <https://github.com/brg015/SingleItemClassification>. The repository also includes example confusion matrices and a short document explaining the key functions and their implementation. Additional metadata are available upon request.

4.1.3 Results

The current analysis describes a novel application of the MVPA framework to single-item decoding. Our initial analyses suggested that classification accuracy was above chance regardless of which logistic-regression leave-one-person out (LR-LOPO) hyperparameters (i.e. lambda), train/test normalization, and train/test comparisons (e.g. pictures-to-pictures) were used (supplementary materials). While above-chance classification was possible regardless of these factors, the results qualitatively suggested that a high-lambda value and both training and testing normalization led to higher classification accuracy. Moreover, the results also suggest that picture-to-picture decoding was by far the best for all single-item classifications (in either **scene** or **object** paradigms; see Table 8, Figure 25), which is unsurprising given previous single-item-decoding results with respect to movie clips (Zadbood et al., 2017) and the shared

perceptual richness of pictures as compared to words. Accordingly, the results herein focus on models with the qualitatively best decoding parameters, while additionally presenting all picture-to-picture decoding results first, with word-to-word, word-to-picture, and picture-to-word results following. For further discussion pertaining to all the LR-LOPO models run, and the results from all models, please see the supplementary materials (Error! Reference source not found. and Table 11).

4.1.3.1 Decoding Individual Items (picture-to-picture)

To quantitatively assess the decoding-accuracy we relied upon two primary variables, namely **Accuracy > 50%** and **Post-prob > Chance**. Recall that **Accuracy > 50%** was based upon the premise that the rank of the to-be-decoded item was within the top 50% - i.e. if we had 100 items and 50 guesses, we would likely correctly guess the to-be-decoded item. This chance threshold can be set to any value, and Figure 21 shows the decoding results with the chance threshold ranging from 5% to 50% and the associated classification accuracy for the to-be-decoded items to *all items* (e.g. cardinal compared to all items) and to *within category* items (e.g. cardinal compared to birds). Statistical results from **Accuracy > 50%** can be found in Table 7.

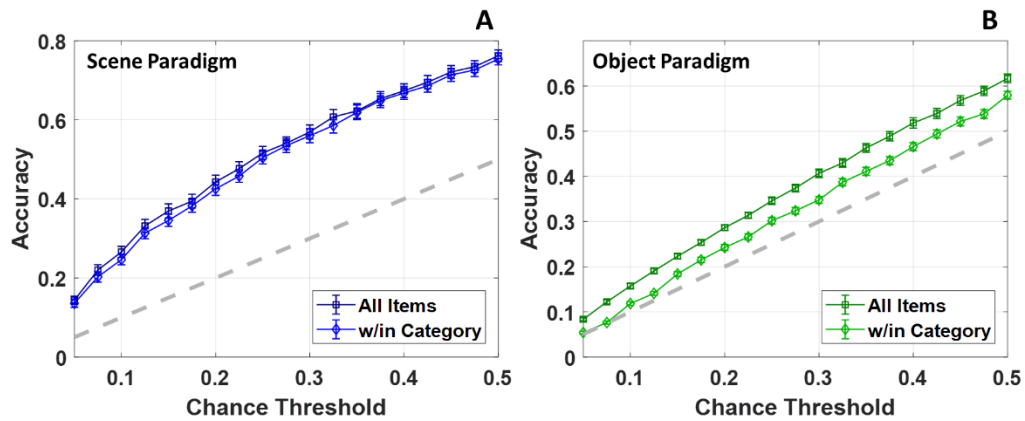


Figure 21. displays above-chance classification accuracy for the scene-paradigm (A: Blue) and the object-paradigm (B: Green) with within category accuracy being shown in a lighter color shade with error bars representing standard error. Regardless of the chance threshold chosen, classification accuracy was always above chance with within category classification accuracy being nearly as high. As is apparent, classification accuracy was higher in the scene-study.

We also plotted the classification evidence for correct classification as compared to incorrect classification as a way to visualize the **Post-prob > Chance** results (Figure 22). “Correct” in this context refers to the average posterior-probability for when the classifier was correct (on-diagonal of the confusion matrix), whereas incorrect refers to the average posterior-probability for when the classifier was incorrect (off-diagonal of the confusion matrix). For within category classification, this was conditionally based upon category, i.e., the off-diagonal only includes within category values. Also note, that for statistical testing (Table 7), we did not test correct > incorrect, as these parameters were effectively yoked (sum of posterior probabilities = 1), in that greater evidence for correct means there was less evidence for incorrect. As such, all tests for **Post-Prob > Chance** compared evidence for correct to theoretical chance levels (e.g., 1/96 or 1/300).

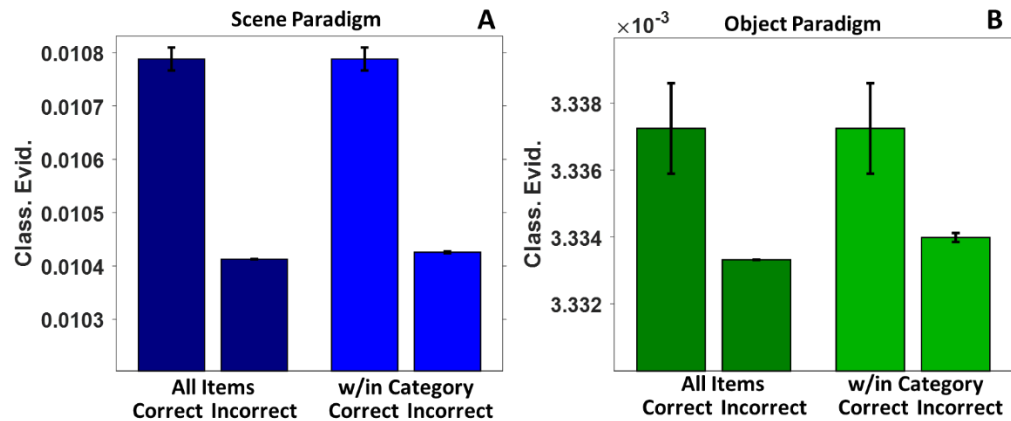


Figure 22. displays above-chance classification accuracy for the scene paradigm (A: Blue) and the object paradigm (B: Green), with the within category evidence being shown in a lighter color shade, and with error bars representing standard error. As would be expected given the Accuracy > 50% results, classification evidence for correct items is higher than classification evidence for incorrect items.

4.1.3.2 Effects of Similarity (picture-to-picture)

We hypothesized that the similarity structure of the confusion matrix should be similar to the hypothesized theoretically-derived stimulus similarity structure, such that when the classifier misclassifies items, it should misclassify to similar items (e.g., classifying cardinal as bluebird). While we relied upon RSA to explicitly test this hypothesis for both continuous and categorical similarity matrices (Table 7), we also sought to visualize this result as well. *If* there is a significant 2nd order correlation between the categorical similarity matrix and the confusion matrix, it suggests that, in general, classification evidence is higher within category as opposed to between categories. As such, if we average the classification evidence both within and between categories, we should see a blocky diagonal reflecting this fact. Figure 23 shows the

result of this manipulation, and such a blocky diagonal is indeed visually present. As with other visualizations, statistics related to this effect are found in Table 7.

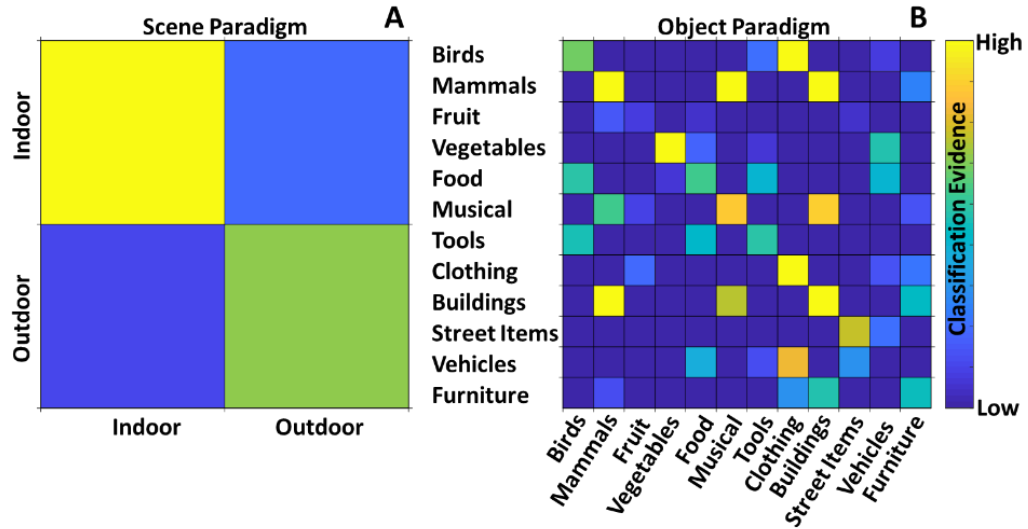


Figure 23. displays the average classification evidence within (on diagonal) and between (off diagonal) categories for the scene paradigm (A) and object paradigm (B). Note that for both studies there was clear trend such that within category evidence was higher than between category evidence.

While the RSA analysis provides evidence that the confusion matrix parallels the similarity structure of the data, it does not indicate how the similarity structure influences the overall decoding accuracy. The exact results of the regression analyses related to this (Equation 1) are included in the supplementary materials (Table 12), while Figure 24 provides a graphical interpretation of the regression results. The values of the probability above chance displayed (y-axis) are based upon the average accuracy (vertical dark grey bar), with the colored bars indicating how this changes in response to the similarity between the confusion matrix and the stimulus similarity matrix. To

estimate these colored bars, we first computed the mean and standard deviation of the 2nd order similarity values. Next, we estimated the probability above chance based upon the regression model. For example, for *Cat. Low* (low category fit), this computation is: probability above chance = average accuracy + beta(for category)*[mean(2nd order correlation) – 2*std(2nd order correlation)]. As such, this value represents how probability above chance changes in response to a low categorical fit (2 standard deviations below average) with respect to the data fit to the hypothesized theoretically-derived stimulus similarity matrix.

For the object study, the categorical similarity appears to have had profound effects on single-item decoding. High categorical similarity appears to have increased decoding accuracy with respect to *all items*, at the cost of significantly decreasing *within category decoding*. The effect of category with respect to the scene-paradigm is similar, with high-category evidence being associated with worse within category decoding.

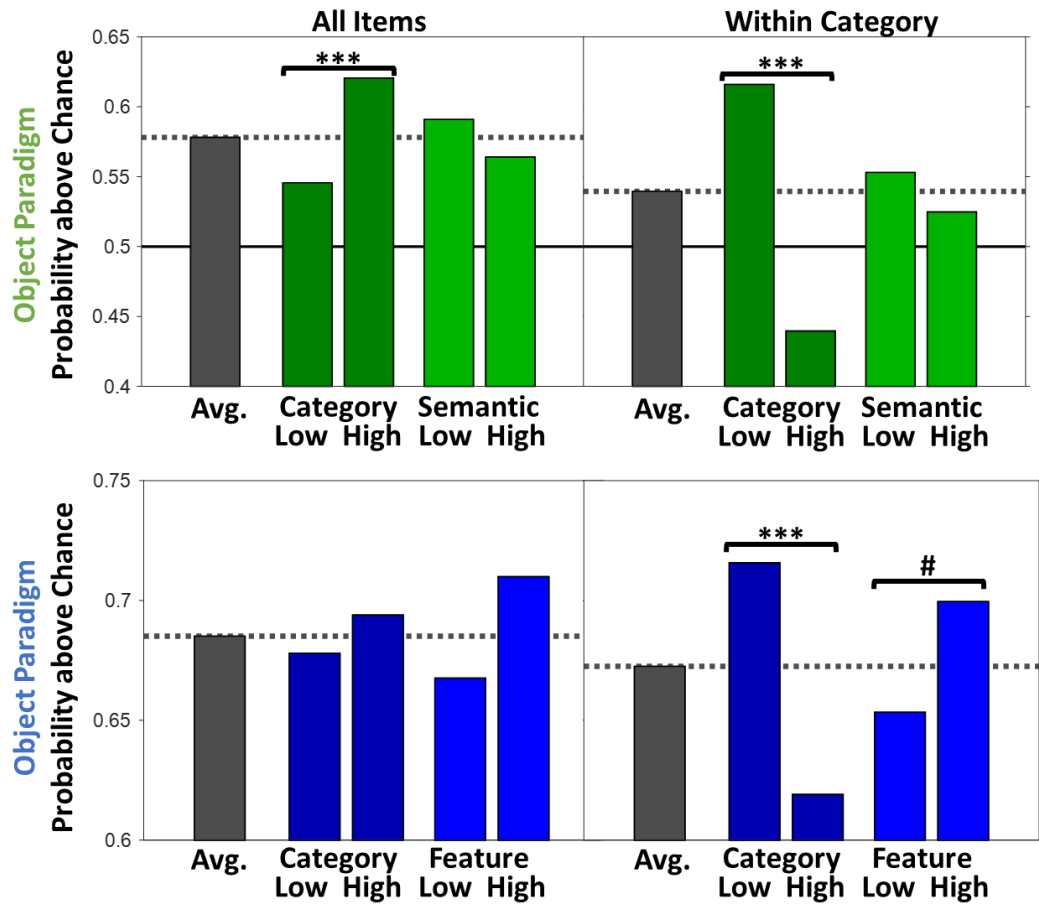


Figure 24. Regression results from equation 1 (see Table S3 for exact beta values) for picture-to-picture decoding. For visualization purposes, we set a baseline to be 0.50 (i.e., we added 0.50 to all estimated values). Category and semantic or feature refer to the hypothesized theoretical-derived similarity matrix, whereas low and high refer to low and high correlation values, with low being 2 standard deviations below average and high being 2 standard deviations above average. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, # $p < 0.10$ and are based upon the regression analysis.

4.1.3.3 Word-to-word, picture-to-word, word-to-picture results

Having graphically presented single-item classification data for the picture-to-picture analyses, we now present the associated statistics in conjunction with those for word-to-word, picture-to-word, and word-to-picture decoding (Table 7). The rows (blue:

scene paradigm, green: object paradigm) indicate training vs. testing comparisons. The first set of columns (Accuracy > 50%) provides statistics pertaining to the ability to classify items (both all and within category) above a chance threshold of 50% - this is visually presented as the right-most value in Figure 21. The average value here (u) was the accuracy above chance, e.g., 0.26 pertains to 76% accuracy when the chance threshold is 50%. Note that for all items, for all training vs. test comparisons, we successfully classified individual items above chance. For within category classification, we were able to classify above chance intra-modally, but failed to classify above chance cross-modally (Picture-to-Word) in the object paradigm. The next column (Post-Prob > Chance) refers to classification evidence. While intra-modal decoding is significantly above chance for both paradigms, for all items and within category, we could only decode between modality in the scene paradigm. Though also note that within category word-to-word decoding in the object paradigm was only trending ($p < 0.10$). The final column pertains to the RSA results for the continuous and categorical similarity matrices with respect to confusion matrices. The averages presented here (u) pertain to the Fischer-transformed 2nd order correlation values from the RSA analyses. The results from this column (categorical similarity) are visually depicted for picture-to-picture decoding in Figure 23. While there appeared to be a stimulus similarity structure to the confusion matrix for intra-modal decoding (picture-to-picture and word-word) most of

the time, the results for cross-modal decoding were less consistent, with the exception of the object paradigm, where there was always a clear categorical structure.

Table 7. Summarized decoding results: rows in blue are from the scene study and rows in green are from the object study. All measures are derived from t-tests within each condition (e.g., Picture-to-Picture), with the degrees of freedom equal to the number of the participants in the test. ***p<0.001, **p<0.01, *p<0.05, #p<0.10

	Accuracy > 50%						Post-Prob > Chance (x10,000, x100,000)						Representational Similarity Analysis					
	All Items			Within Category			All Items			Within Category			Continuous			Categorical		
	u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat
Picture-to-Picture	0.26	0.01	18.08***	0.25	0.07	17.94***	3.71	0.99	17.20***	3.62	0.99	16.75***	0.034	0.013	12.08***	0.024	0.015	7.40***
Picture-to-Word	0.06	0.01	5.27***	0.07	0.06	5.33***	0.85	0.74	5.28***	0.88	0.76	5.29***	0.007	0.009	3.95***	-0.001	0.010	-0.59
Word-to-Picture	0.06	0.01	6.41***	0.06	0.04	6.53***	0.69	0.44	7.15***	0.70	0.45	7.20***	0.007	0.009	3.49**	-0.001	0.011	-0.41
Word-to-Word	0.06	0.01	5.31***	0.07	0.05	6.08***	0.74	0.66	5.13***	0.75	0.65	5.25***	0.003	0.007	2.22*	-0.001	0.010	-0.46
Picture-to-Picture	0.12	0.01	12.29***	0.08	0.04	9.41***	0.39	0.59	2.89**	0.32	0.53	2.66*	0.003	0.006	2.70*	0.015	0.005	14.15***
Picture-to-Word	0.03	0.01	4.21***	0.01	0.02	1.06	0.00	0.20	-0.04	-0.02	0.25	-0.43	0.002	0.005	2.15*	0.005	0.004	4.81***
Word-to-Picture	0.04	0.01	4.54***	0.02	0.03	2.20*	0.00	0.12	-0.02	-0.01	0.15	-0.37	0.002	0.005	1.37	0.006	0.004	7.48***
Word-to-Word	0.10	0.03	3.22**	0.08	0.14	2.39*	0.08	0.16	2.13*	0.07	0.17	1.90#	0.004	0.005	3.44**	0.005	0.004	5.38***

4.1.4 Discussion

The current study had two main goals. Our first goal was to investigate the possibility of decoding the fMRI responses to individual items, both intra-modally (e.g., object-picture to object-picture) and cross-modally (e.g., object-picture to object-word), even when the to-be-decoded item is amongst other highly similar items. The results confirmed that our approach was able to reliably decode the fMRI responses to single items, from single presentations, even when subjects viewed hundreds of unique stimuli. This finding replicated across different stimulus types including scenes, object pictures, and words. Importantly, intra-modally, the method can decode single items amongst highly similar items (e.g., a cardinal amongst other birds). We could additionally classify stimuli cross-modally (e.g., decode object-words using a classifier trained on object-pictures) – though here the classifier struggled to distinguish items within the same category. Our second goal was to investigate whether classification accuracy is a function of stimulus similarity. Using representational similarity analysis (RSA), we confirmed that classification evidence varied with the similarity between to-be-decoded item and the stimulus similarity structure. As a result, the algorithms tended to produce false positive for items that were similar to the target, with more false positives leading to worse within category classification performance. In summary, our results extend the current MVPA literature by demonstrating the reliability of decoding individual items as opposed to groups or categories of items.

4.1.4.1 Decoding Individual Items

Our first finding was that our single-item decoding technique was able to reliably identify individual stimuli across two disparate datasets. Within-modality classification accuracy was highest for scenes, then for objects, and lastly for words. This pattern is highly consistent with the notion that decoding is easiest when the sensory input is highly detailed, such that individual differences related to the stimuli are heavily constrained. Furthermore, the strength of our approach suggests a new means of both a more fine-grained classification scheme, as well as capturing the stimulus-level variance in neural responses (Westfall, Nichols, & Yarkoni, 2017). Previous evidence has suggested that individual movie clips, for example, can be reliably decoded between participants (Zadbood et al., 2017), however, relatively few researchers currently utilize such dynamic stimuli and the applicability of this approach is limited. Here, we provide a rigorous framework to demonstrate the ability to decode individual static items presented only once, and in doing so provide researchers with a valuable methodology applicable to a wide range of studies.

The strength of our classification technique, as noted above, was largely influenced by two critical factors: the categorical structure of the stimulus set, and the modality in which the items were viewed. The first trend is interesting because it suggests that the observed uniqueness of an item's representations is influenced by its reference to a category exemplar or centroid. A related and potential critique of the

employed methodology is that our single-item classifier approach might have been capable of only detecting gross categorical-level features and thus incapable of distinguishing between highly similar items. To address this potential shortfall we demonstrated that, for picture-to-picture and word-to-word intra-modal decoding, we could reliably decode within a category that consisted of similar exemplars (e.g., decoding cardinal amongst other birds). Critically, the within-category decoding accuracy was only slightly worse than the classification accuracy across all items, strongly suggesting that the classifier analysis does not rely on gross categorical features but is instead capable of detecting subtle differences between even highly similar items. As such, these results suggest that even though item decoding approach are dominated by the success of classifying broad categories of items (e.g., faces vs. houses), a more fine-grained dissociation of such biological (or non-biological) classes is possible, in line with other approaches using more constrained stimuli sets (Connolly et al., 2012; Mack, Love, & Preston, 2016)

The second factor affecting our item decoding result was that cross-modal decoding, as expected, was more difficult than intra-modal decoding. As our classifier took advantage of whole-brain data, and cross-modal representations for objects and words have been reported in several cortical regions, including the left intraparietal sulcus (Devereux et al., 2013) and the precuneus (Fairhall & Caramazza, 2013), we hypothesized that the classifier would be able to successfully classify cross-modally as

well. Indeed, we found that, in the scene paradigm, cross-modal classification accuracy was significantly above chance for *all items* and *within category* for the Accuracy > 50% and Post-prob > chance measures. However, in the object paradigm, we only achieved above-chance cross-modal decoding based upon the Accuracy > 50% measure for *all items*. This suggests that when operating cross-modally, the classifier might mainly capture categorical level properties (e.g., the classifier correctly identifies the stimulus as a bird, but cannot identify it as cardinal). A potential reason for the inability to classify within category in the object paradigm might be the strong categorical structure, which made it harder to identify specific stimuli, whereas the scene paradigm has only two major categories, making it easier to distinguish within category. Another potential explanation is that the scene stimuli were more detailed (perceptually richer) than the object stimuli, thereby also enabling easier within-category classification. Regardless, we again note that intra-modal, within-category classification was significantly above chance for both paradigms for both Post-prob > chance and accuracy > 50% (the only exception being within category word-to-word decoding on the Post-prob > chance measure in the object paradigm – which was only trending $p < 0.10$).

Above these two critical design factors affecting the reliability of individual-item decoding, it should also be noted, that there are both benefits and costs of utilizing cross-subject decoding as opposed to within-subject decoding, as the inferences drawn from these results can be different (Clithero, Smith, Carter, & Huettel, 2011). With regard to

the nature of individual stimuli, presented only once, the current study was forced to rely upon cross-subject decoding, and, in doing so, we captured what is common across subjects, as opposed to what is unique. This might have drawbacks in some circumstances in which you are attempting to capture how a particular factor specifically affects an individual participant. Despite this limitation, our single-item approach addresses a more fundamental problem in the fMRI literature, namely the “stimulus-as-fixed-effect fallacy”, such that most fMRI decoding approaches fail to model stimulus variability in a manner that allows one to generalize a) to a new set of similar objects or images, or b) across other members of the population not viewing the exact same sets of stimuli. Furthermore, this approach also has benefits with respect to generalizing common effects across participants, given the cross-subject approach mentioned above.

4.1.4.2 Effects of Similarity

While our results demonstrate the ability to differentiate individual stimuli from highly similar stimuli, we had also hypothesized that classification evidence would be highest for similar items. Accordingly, we predicted that the confusion matrix from each subject would be highly similar to a stimulus-similarity matrix. The current results were consistent with this prediction. We found that the categorical similarity matrices were highly similar both intra-modally and cross-modally in the object paradigm, whereas it was only statistically significant for picture-to-picture decoding in the scene study. The

robust results from the object study were likely partly driven by the fact that the categorical structure in that paradigm was purposefully imposed - i.e. the study was explicitly designed to have twelve very distinct categories, whereas the scene study was sorted *post hoc* into only two categories, namely indoor/outdoor.

We did note a number of differences between our scene- and object-memory paradigms worthy of note. In our scene-memory paradigm the continuous similarity matrices (reflecting the overlap of scene features) fit the data better than a categorical model (reflecting indoor/outdoor status), while the opposite was true for the object paradigm. Potential explanation for this effect include the fact while an indoor/outdoor is well represented in very localized regions of cortex (Henderson, Larson, & Zhu, 2007), this information alone is not useful in decoding an individual item's identity. In contrast, the more complex category structure of the object-memory paradigm (12 separate, evenly-balanced categories) may better capture the granularity of itemwise differences in neural pattern similarity. A second explanation is that the categorical structure for the scene paradigm was based upon *post-hoc* distinctions, and thus some scenes did not cleanly fit into one category or the other (e.g., storefront). Thus, the continuous similarity structure was better able to capture the subtleties in the confusion matrix. However, for the object paradigm, the categories were *a priori* defined, and participants likely noticed that the stimuli fit into a discrete number of bins. This increased attention to the categorical structure might have warped the neural representations of the

categorical items, leading to the classifier more clearly deciphering the categorical structure (Çukur, Nishimoto, Huth, & Gallant, 2013).

Although similarity between the confusion matrices and the stimulus-similarity matrices provides evidence that misclassifications were typically occurring for highly similar items, this similarity does not demonstrate how these misclassifications impacted overall classification performance. In the object- and scene-paradigm, the subsequent regression analyses demonstrated a link between these measures by showing that greater fits between the confusion matrix and the categorical stimulus-similarity matrix led to worse within-category classification (but better classification with respect to *all items* in the case of the object-paradigm). This suggests that one cause of low within-category classification accuracy was that the classifier successfully identified categorical content (i.e., the to-be-decoded item is a bird) but had difficulty deciphering the exact exemplar within the category (e.g., which type of bird). Unsurprisingly though, the ability of the classifier to identify the category of the stimulus (e.g, a bird) improved decoding accuracy with respect to *all* items.

4.1.4.3 Limitations and Caveats

A major limitation of this analysis is that both studies utilized memory reactivation paradigms, and as such, the picture and word presentations were linked by their mnemonic strength, given that during both studies, during the presentation of the word phase, participants were actively attempting to retrieve representations of specific

pictures previously viewed. This characteristic of these studies thus confounds the across-modality classification analysis. Regardless, it is quite clear that picture-to-picture single-item decoding is far above chance, even when decoding within category.

4.1.4.4 General Conclusion

In this manuscript, across two paradigms using different stimulus sets, we demonstrated the feasibility of intra-modal and cross-modal single-item decoding, while also suggesting optimal preprocessing and hyperparameter settings to maximize classification accuracy. We demonstrated reliable intra-modal decoding even when the to-be-decoded item was highly similar to other comparison items (i.e. within-category classification), a finding which highlights the value of taking into account stimulus variability even within categorically parcellated stimulus sets. Additionally, we showed that misclassifications exhibited a systematic structure based upon the similarity of the to-be-decoded items and that the degree of this similarity negatively influenced the ability to classify single-items. In total, we have demonstrated the ability to reliably classify individual items and thus provide researchers with a very useful way to ask representational questions about single-items as opposed to just categories of items.

4.1.3 Supplementary Materials

4.1.3.1 Determining whether single-items can be decoded

To determine whether single-items can be decoded we ran a series of regression analyses (Equation 2). In this model we included categorical variables for all training vs.

test scenarios (e.g., picture-to-picture, picture-to-word etc.) and paradigm (object or scene). We further sorted the training vs. test scenarios such that the lowest accuracy term was absorbed by the intercept term, and as such we listed the intercept term as the Training vs. Test scenario that had the worst performance (though also note the study term is also partly reflected in this term as well). If the intercept term was positive and other beta terms were significant, it suggests that we could decode above chance. We also ran additional models in which we examined within-category Accuracy > 50% for each study independently, without a study term. Note that we do not consider these to be a rigorous statistical assessment of the viability of our models, but rather as a qualitative assessment of the likelihood of our technique.

Equation 2. Addresses whether the classification approach in general achieves above chance classification accuracy.

$$\text{Accuracy} > 50\% \sim \text{Train vs. Test} + \text{Scene Study}$$

The results from this analysis (Table 8 and Figure 25) provide robust evidence that regardless of normalization or the setting of lambda, single items can be decoded both modally (e.g., picture-to-picture) and cross-modally (e.g., picture-to-word). Additionally, the results show that classification performance is best for picture-to-picture, then word-to-word, then word-to-picture and then picture-to-word. The results also suggest that, in general, classification accuracy was higher in the scene-study than the object-study. The regression-parameter estimate (Est.) refers to accuracy above chance, and thus the results (Table 8A) suggest that the worst classification was for

picture-to-word in the object study, whereas the best classification was for picture-to-picture in the scene study. Following similar logic, within category classification (Table 8B) was above chance in the scene and object paradigm for all training vs. test comparisons. Figure 25 summarizes the results of the regression model in a graphical form.

Table 8. Results from equation 1. A.Left. Result from combined analysis, across the scene and object study {df=1280=32 models*40 participants} B.Right(Blue) Results from scene study within category {df=1344=32 models*21 participants*2 categories} B.Left(Green) Results from object study within category {df=7296=32 models*19 participants*12 categories }. ***p<0.001, **p<0.01, *p<0.05, #p<0.10

A								B			
<i>Accuracy > 50% ~ Train vs. Test + Scene Study</i>				<i>Accuracy > 50% ~ Train vs. Test (within category)</i>							
	Est.	SE	tStat		Est.	SE	tStat		Est.	SE	tStat
Pictures to Words (PtoW)	0.019	0.004	4.4 ***	Pictures to Words (PtoW)	0.006	0.003	2.1 *	Pictures to Words (PtoW)	0.046	0.004	10.7 ***
Words to Pictures (WtoP)	0.004	0.005	0.8	Words to Pictures (WtoP)	0.001	0.004	0.4	Words to Pictures (WtoP)	0.007	0.006	1.1
Words to Words (WtoW)	0.019	0.005	3.5 ***	Words to Words (WtoW)	0.051	0.004	13.4 ***	Words to Words (WtoW)	-0.007	0.006	-1.2
Pictures to Pictures (PtoP)	0.117	0.005	22.0 ***	Pictures to Pictures (PtoP)	0.054	0.004	14.0 ***	Pictures to Pictures (PtoP)	0.152	0.006	24.9 ***
Scene Paradigm	0.033	0.004	8.7 ***								

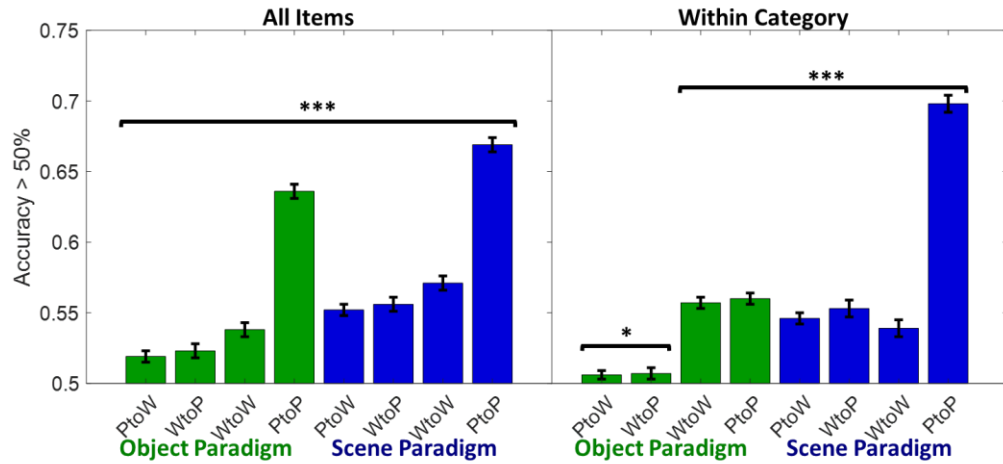


Figure 25. Results from equation 1 plotted as a function of beta estimates from the associated linear model. The results indicate that the LR-LOPO models, in general, are capable of above chance decoding both modally and cross-modally. With respect to the x-axis, P indicates picture and W indicates word, such that PtoW is the LR-LOPO model associated with training on pictures and testing on words.

***p<0.001, **p<0.01, *p<0.05, #p<0.10

4.1.3.2 Determining optimal LR-LOPO decoding parameters

To determine optimal decoding parameters we again ran a family of regression analyses for all training vs. test comparisons, both for all items and within category for each study, in which we examined the influence of the lambda and normalization parameters, while including subject as a random effect (Equation 3). When testing for within category accuracy > 50%, an additional categorical variable was included for category.

Equation 3. Addresses how the parameters of interest (lambda and normalization) influence classification accuracy.

$$\text{Accuracy} > 50\% \sim \text{High Lambda} + \text{Normalize Training} + \text{Normalize Testing} + (1|\text{Subject})$$

The results from these regression analyses suggest that a high lambda value (100) and normalization both prior to training and testing maximized Accuracy>50% (Table 9). Note that although these settings did not *always* lead to better Accuracy>50%, they almost never led to statistically worse performance (i.e. associated parameters are non-significant). The only exception to this was picture-to-picture normalization prior to training in the scene paradigm, in which case no normalization led to higher accuracy.

Table 9. Results from equation 2 {df=320=8 models*40 participants for Object & Scene; 8*21=168 for Scene Study and 8*19=152 for Object Study}. Each column of the table indicates weather the associated beta-value was significant, and if so in which direction (+ indicates positive, - indicates negative).

Accuracy > 50% ~ High Lambda + Normalize Training + Normalize Testing + (1|Subject)

	Accuracy > 50%					Accuracy > 50% (w/in category)									
	Object & Scene Paradigm					Object Paradigm					Scene Paradigm				
	PtoP	PtoW	WtoP	WtoW	All	PtoP	PtoW	WtoP	WtoW	All	PtoP	PtoW	WtoP	WtoW	All
High Lambda	<0.05+	n.s.	<0.10+	<0.05+	<0.05+	<0.05+	n.s.	n.s.	n.s.	n.s.	<0.05+	n.s.	n.s.	<0.05+	<0.05+
Normalize Training	<0.05+	n.s.	<0.05+	<0.05+	<0.05+	<0.05+	n.s.	<0.05+	n.s.	<0.05+	<0.05-	n.s.	n.s.	<0.10+	n.s.
Normalize Testing	<0.05+	<0.05+	<0.05+	<0.05+	<0.05+	<0.05+	n.s.	n.s.	n.s.	<0.05+	<0.05+	<0.05+	<0.05+	<0.05+	<0.05+

In general, normalization had profound benefits that were easily visible in the data (Figure 26 and Figure 27) as without this procedure the confusion matrix ended up being heavily biased towards specific stimuli. A likely reason for this is that the voxels associated with these stimuli had intrinsically elevated activity in the left-out subject and thus caused the classifier to routinely false-alarm.

The finding that a high lambda increased accuracy was expected, as a high lambda value helps to minimize overfitting the data, which is a risk when dealing with low numbers of subjects (i.e., training examples). However, if a large number of participants is available, we would recommend the inclusion of a validation set to determine the optimal value of the lambda parameter, as a large number of samples likely reduces the risk of overfitting.

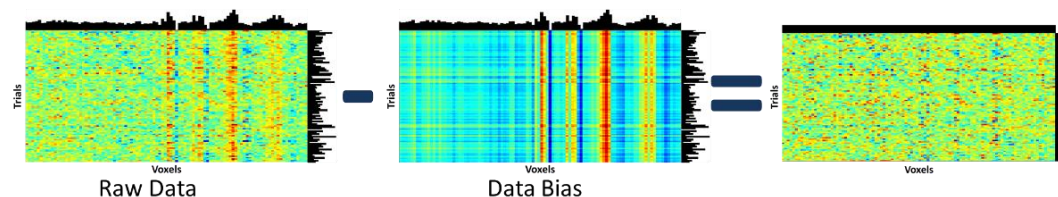


Figure 26. Effect of debiasing the data. (Left) displays raw data from a subset of trials and voxels from an example subject in the scene-study. The histograms on the edges represent summed activity for each specific voxel or trial. (Center) a surrogate dataset is created to have some of the same statistical properties as the actual dataset, in this case the surrogate dataset has the same trial-wise and voxel-wise mean and standard deviation – this is clearly evident via the histograms on the axes. (Right) To create normalized dataset, we subtract the bias from the raw data. The bias, essentially represents the expected value of each trial and voxel based upon all other trials and voxels, as such, the subtraction provides an indication of this trial-voxel value with respect to its expected value.

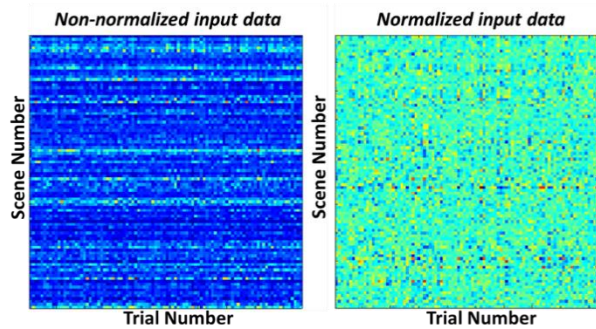


Figure 27. Effect of normalization on the confusion matrix (example subject from the scene study). As can be seen, the normalization procedure removes streaks in the data, which in this case are likely consistent false alarms to incorrect individual items.

We also include the results from all 32 models [training normalization (on/off), testing normalization (on/off), lambda (high/low) = 8 comparisons, for all train (picture/word) vs. test (picture/word) = $8 \times 4 = 32$] for both the scene and object paradigm (Error! Reference source not found.: Scene, Table 11: Object).

Table 10. Results from all scene-paradigm models. The format is identical to table S1 with the inclusion of additional

columns (left) pertaining to model parameters. ***p<0.001, **p<0.01, *p<0.05, #p<0.10

Lambda	Stimuli		Norm.		Accuracy > 50%						Post-Prob > Chance (x10,000)						Representational Similarity Analysis					
					All Items			Within Category			All Items			Within Category			Feature			Categorical		
					u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat
0.05	P	P	0	0	0.17	0.01	15.10***	0.16	0.05	13.26***	106.97	49.86	9.83***	105.37	49.73	9.71***	0.035	0.032	5.02***	0.017	0.018	4.37***
0.05	P	P	1	1	0.23	0.02	15.05***	0.22	0.07	13.90***	5.27	1.68	14.34***	5.14	1.73	13.66***	0.032	0.010	14.75***	0.021	0.013	7.73***
0.05	P	P	0	1	0.26	0.02	17.11***	0.25	0.06	18.16***	60.39	15.51	17.84***	58.88	15.52	17.39***	0.032	0.014	10.70***	0.030	0.015	9.34***
0.05	P	P	1	0	0.15	0.01	16.79***	0.13	0.04	14.71***	5.21	1.65	14.49***	5.09	1.69	13.80***	0.024	0.031	3.56**	0.007	0.016	1.93#
100	P	P	0	0	0.17	0.01	22.72***	0.17	0.04	21.37***	13.99	4.25	15.08***	13.64	4.28	14.59***	0.027	0.026	4.80***	0.014	0.013	4.87***
100	P	P	1	1	0.26	0.01	18.08***	0.25	0.07	17.94***	3.71	0.99	17.20***	3.62	0.99	16.75***	0.034	0.013	12.08***	0.024	0.015	7.40***
100	P	P	0	1	0.26	0.01	20.87***	0.26	0.06	19.11***	10.22	2.42	19.38***	9.95	2.44	18.68***	0.035	0.010	15.79***	0.028	0.015	8.42***
100	P	P	1	0	0.16	0.01	17.38***	0.15	0.04	15.68***	3.69	0.98	17.21***	3.59	0.98	16.74***	0.029	0.029	4.46***	0.007	0.012	2.59*
0.05	P	W	0	0	0.04	0.01	2.68*	0.03	0.06	2.36*	18.50	18.06	4.69***	18.71	17.99	4.76***	0.042	0.036	5.32***	0.009	0.015	2.65*
0.05	P	W	1	1	0.05	0.01	4.13***	0.05	0.06	3.62**	0.95	1.18	3.68**	0.96	1.22	3.61**	0.007	0.007	4.21***	0.000	0.009	-0.22
0.05	P	W	0	1	0.06	0.01	4.26***	0.05	0.06	3.66**	10.62	11.56	4.21***	10.87	11.31	4.40***	0.011	0.012	3.91***	0.001	0.013	0.43
0.05	P	W	1	0	0.03	0.01	2.39*	0.03	0.05	2.48*	0.96	1.18	3.70**	0.97	1.22	3.63**	0.029	0.027	4.87***	0.003	0.015	0.88
100	P	W	0	0	0.04	0.01	3.38**	0.04	0.06	3.50**	2.63	2.82	4.27***	2.73	2.98	4.21***	0.042	0.030	6.45***	0.005	0.013	1.75#
100	P	W	1	1	0.06	0.01	5.27***	0.07	0.06	5.33***	0.85	0.74	5.28***	0.88	0.76	5.29***	0.007	0.009	3.95***	-0.001	0.010	-0.59
100	P	W	0	1	0.06	0.01	4.41***	0.06	0.07	3.79**	2.13	1.93	5.05***	2.16	1.97	5.04***	0.010	0.009	5.01***	0.000	0.010	-0.21
100	P	W	1	0	0.04	0.01	3.52**	0.04	0.05	3.22**	0.85	0.74	5.29***	0.87	0.75	5.31***	0.031	0.023	6.25***	-0.002	0.013	-0.73
0.05	W	P	0	0	0.03	0.01	3.17**	0.03	0.04	3.89***	1.24	15.45	0.37	1.28	15.52	0.38	0.023	0.041	2.60*	0.003	0.018	0.66
0.05	W	P	1	1	0.07	0.01	5.52***	0.07	0.06	5.41***	1.02	0.67	6.99***	1.01	0.69	6.73***	0.009	0.012	3.37**	0.002	0.010	0.90
0.05	W	P	0	1	0.06	0.01	4.35***	0.05	0.06	3.71**	7.33	14.32	2.35*	7.06	14.57	2.22*	0.002	0.016	0.65	0.002	0.014	0.60
0.05	W	P	1	0	0.04	0.01	4.43***	0.04	0.05	4.35***	1.02	0.67	6.96***	1.01	0.69	6.72***	-0.003	0.024	-0.47	-0.003	0.011	-1.43
100	W	P	0	0	0.05	0.01	6.38***	0.05	0.04	5.04***	2.76	2.15	5.89***	2.83	2.13	6.10***	-0.006	0.024	-1.23	-0.003	0.016	-0.78
100	W	P	1	1	0.06	0.01	6.41***	0.06	0.04	6.53***	0.69	0.44	7.15***	0.70	0.45	7.20***	0.007	0.009	3.49**	-0.001	0.011	-0.41
100	W	P	0	1	0.08	0.01	7.36***	0.08	0.05	6.84***	1.97	1.07	8.42***	1.99	1.08	8.47***	0.010	0.011	4.11***	-0.001	0.014	-0.41
100	W	P	1	0	0.05	0.01	6.53***	0.05	0.03	6.73***	0.68	0.44	7.09***	0.69	0.44	7.17***	-0.012	0.020	-2.73*	-0.009	0.009	-4.78***
0.05	W	W	0	0	0.00	0.01	0.45	0.01	0.04	1.06	22.42	44.56	2.31*	22.38	45.20	2.27*	-0.005	0.026	-0.82	-0.001	0.017	-0.18
0.05	W	W	1	1	0.04	0.01	3.50**	0.04	0.05	3.84**	0.66	0.91	3.30**	0.67	0.93	3.26**	0.002	0.010	0.93	-0.002	0.012	-0.86
0.05	W	W	0	1	0.03	0.01	2.72*	0.03	0.05	2.97**	12.41	17.01	3.34**	12.42	17.41	3.27**	0.003	0.016	0.95	0.003	0.013	1.19
0.05	W	W	1	0	0.03	0.01	2.51*	0.03	0.05	2.83*	0.65	0.89	3.34**	0.66	0.91	3.30**	-0.005	0.024	-1.03	-0.009	0.016	-2.67*
100	W	W	0	0	0.04	0.01	3.21**	0.03	0.05	2.69*	1.86	2.28	3.73**	1.85	2.44	3.48**	0.005	0.020	1.08	-0.002	0.008	-1.11
100	W	W	1	1	0.06	0.01	5.31***	0.07	0.05	6.08***	0.74	0.66	5.13***	0.75	0.65	5.25***	0.003	0.007	2.22*	-0.001	0.010	-0.46
100	W	W	0	1	0.06	0.01	4.35***	0.06	0.06	4.55***	1.68	1.59	4.84***	1.69	1.61	4.82***	0.003	0.009	1.60	0.000	0.011	-0.16
100	W	W	1	0	0.04	0.01	3.50**	0.04	0.05	3.58**	0.74	0.65	5.17***	0.75	0.65	5.29***	-0.012	0.016	-3.57**	-0.011	0.011	-4.51***

Table 11. Results from all object-paradigm models. The format is identical to Error! Reference source not found. with the inclusion of additional columns (left) pertaining to model parameters. *p<0.001, **p<0.01, *p<0.05, #p<0.10**

Lambda	Stimuli		Norm.	Accuracy > 50%									Post-Prob > Chance (x100,000)						Representational Similarity Analysis					
	Trn	Tst		All Items			Within Category			All Items			Within Category			Semantic			Categorical					
				u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat	u	std	tStat			
0.05	P	P	0	0	0.06	0.01	8.61***	0.03	0.03	3.84**	30.59	22.20	6.01***	19.77	25.95	3.32**	0.001	0.006	1.04	0.011	0.004	14.31***		
0.05	P	P	1	1	0.11	0.01	13.93***	0.08	0.04	8.95***	0.31	0.38	3.53**	0.23	0.27	3.66**	0.002	0.007	1.32	0.016	0.006	12.05***		
0.05	P	P	0	1	0.07	0.01	10.10***	0.04	0.02	7.22***	24.14	13.63	7.72***	13.17	21.82	2.63*	0.004	0.004	3.99***	0.014	0.004	14.77***		
0.05	P	P	1	0	0.11	0.01	11.65***	0.07	0.04	7.59***	0.31	0.38	3.53**	0.23	0.27	3.66**	0.000	0.008	0.02	0.016	0.006	12.18***		
100	P	P	0	0	0.08	0.01	11.36***	0.04	0.03	7.09***	12.07	6.55	8.03***	8.95	6.06	6.44***	0.005	0.007	2.96**	0.017	0.005	15.53***		
100	P	P	1	1	0.12	0.01	12.29***	0.08	0.04	9.41***	0.39	0.59	2.89**	0.32	0.53	2.66*	0.003	0.006	2.70*	0.015	0.005	14.15***		
100	P	P	0	1	0.10	0.01	11.12***	0.06	0.04	7.28***	9.55	4.40	9.46***	7.06	4.06	7.58***	0.005	0.005	3.81**	0.019	0.005	15.74***		
100	P	P	1	0	0.11	0.01	12.00***	0.07	0.04	8.94***	0.39	0.59	2.89**	0.32	0.53	2.66*	0.002	0.006	1.43	0.015	0.005	14.16***		
0.05	P	W	0	0	0.01	0.01	2.20*	0.00	0.03	-0.23	2.65	12.86	0.90	-0.48	12.98	-0.16	0.003	0.007	1.58	0.005	0.003	7.04***		
0.05	P	W	1	1	0.03	0.01	4.08***	0.02	0.03	2.41*	0.02	0.12	0.85	-0.01	0.16	-0.20	0.004	0.007	2.59*	0.005	0.004	6.77***		
0.05	P	W	0	1	0.02	0.01	2.92**	0.00	0.03	0.06	6.21	38.51	0.70	3.63	39.64	0.40	0.002	0.004	1.83#	0.005	0.004	6.02***		
0.05	P	W	1	0	0.03	0.01	4.12***	0.01	0.03	1.29	0.02	0.12	0.85	-0.01	0.16	-0.20	0.003	0.006	2.25*	0.006	0.004	6.64***		
100	P	W	0	0	0.02	0.01	3.34**	0.01	0.03	0.83	2.29	2.79	3.58**	1.25	3.05	1.79#	0.002	0.005	2.03#	0.006	0.004	7.74***		
100	P	W	1	1	0.03	0.01	4.21***	0.01	0.02	1.06	0.00	0.20	-0.04	-0.02	0.25	-0.43	0.002	0.005	2.15*	0.005	0.004	4.81***		
100	P	W	0	1	0.02	0.01	4.07***	0.01	0.03	1.33	1.98	2.31	3.75**	1.16	2.29	2.20*	0.002	0.003	3.66**	0.006	0.003	7.84***		
100	P	W	1	0	0.02	0.01	4.34***	0.00	0.02	0.81	0.00	0.20	-0.04	-0.02	0.25	-0.43	0.003	0.005	2.04#	0.004	0.004	4.58***		
0.05	W	P	0	0	0.00	0.01	0.18	-0.01	0.02	-2.24*	12.52	28.97	1.88#	10.40	28.62	1.58	-0.001	0.008	-0.61	0.002	0.003	2.25*		
0.05	W	P	1	1	0.04	0.01	7.08***	0.02	0.03	3.35**	0.05	0.09	2.30*	0.04	0.08	1.93#	0.001	0.006	0.88	0.006	0.003	8.22***		
0.05	W	P	0	1	0.01	0.01	1.23	-0.01	0.03	-1.20	1.70	11.68	0.63	0.56	11.72	0.21	0.000	0.004	-0.04	0.002	0.003	3.23**		
0.05	W	P	1	0	0.04	0.01	6.29***	0.02	0.04	2.74*	0.05	0.09	2.31*	0.04	0.08	1.93#	-0.001	0.006	-0.78	0.006	0.003	7.95***		
100	W	P	0	0	0.01	0.00	2.83*	0.00	0.02	-0.50	2.11	2.96	3.11**	0.92	3.24	1.24	-0.002	0.010	-0.68	0.005	0.004	5.59***		
100	W	P	1	1	0.04	0.01	4.54***	0.02	0.03	2.20*	0.00	0.12	-0.02	-0.01	0.15	-0.37	0.002	0.005	1.37	0.006	0.004	7.48***		
100	W	P	0	1	0.01	0.01	2.39*	0.00	0.02	0.36	1.44	2.79	2.25*	0.38	2.79	0.60	0.001	0.004	0.64	0.006	0.005	5.45***		
100	W	P	1	0	0.04	0.01	5.33***	0.01	0.03	1.88#	0.00	0.12	-0.02	-0.01	0.15	-0.37	-0.001	0.005	-1.06	0.006	0.003	7.07***		
0.05	W	W	0	0	0.05	0.02	2.24*	0.03	0.09	1.58	31.10	68.84	1.97#	28.54	68.37	1.82#	0.003	0.011	1.26	0.003	0.004	3.28**		
0.05	W	W	1	1	0.09	0.03	3.03**	0.08	0.14	2.48*	0.10	0.15	3.02**	0.09	0.14	2.87*	0.004	0.006	2.48*	0.004	0.004	4.98***		
0.05	W	W	0	1	0.06	0.02	2.58*	0.05	0.09	2.11*	22.67	47.21	2.09#	20.62	47.31	1.90#	0.001	0.004	0.92	0.004	0.004	4.25***		
0.05	W	W	1	0	0.09	0.03	3.00**	0.08	0.13	2.50*	0.10	0.15	3.02**	0.09	0.14	2.86*	0.001	0.006	0.97	0.004	0.004	4.46***		
100	W	W	0	0	0.05	0.02	2.15*	0.04	0.10	1.63	10.15	22.81	1.94#	9.23	22.75	1.77#	0.003	0.006	1.74#	0.003	0.004	4.00***		
100	W	W	1	1	0.10	0.03	3.22**	0.08	0.14	2.39*	0.08	0.16	2.13*	0.07	0.17	1.90#	0.004	0.005	3.44**	0.005	0.004	5.38***		
100	W	W	0	1	0.06	0.02	2.28*	0.04	0.11	1.66	7.69	18.69	1.79#	6.88	18.64	1.61	0.002	0.004	2.42*	0.005	0.004	5.57***		
100	W	W	1	0	0.09	0.03	3.06**	0.07	0.14	2.18*	0.08	0.16	2.13*	0.07	0.17	1.90#	0.001	0.004	1.49	0.005	0.004	5.45***		

4.1.3.3 Effects of Similarity

The Est. parameter indicates how classification accuracy above chance changes as a function of 2nd order similarity to the stimulus similarity matrix (see Equation 1). See Figure 24 for a graphical interpretation of these regression results.

Table 12. Results for items {df=2016=96*21 for the scene paradigm (blue) and df=5700=300*19 for the object paradigm (green)} *p<0.001, **p<0.01, *p<0.05, #p<0.10**

		All Items			Within Category		
		Est.	SE	tStat	Est.	SE	tStat
Object Paradigm	Categorical	0.27	0.06	4.72 ***	-0.64	0.06	-11.11 ***
	Continuous	-0.11	0.06	-1.70 #	-0.12	0.06	-1.78 #
Scene Paradigm	Categorical	0.03	0.06	0.61	-0.21	0.06	-3.73 ***
	Continuous	0.09	0.06	1.63	0.10	0.06	1.78 #

5. Conclusions and Future Directions

Episodic memory, as a cognitive construct, exists only in relation to those other cognitive constructs that reference it. The primary goal of this dissertation was to demonstrate, across a range of studies, how various cognitive constructs facilitate episodic memory. In Chapter 2, I demonstrate how attentional processes, and perhaps cognitive control processes, aid the successful encoding of information. In Chapter 3, I demonstrate how large-scale network interactions across the whole brain aid episodic memory retrieval. And in Chapter 4 I develop a machine learning framework to decode individual events, with the goal of examining event-specific traces with respect to episodic memory encoding and retrieval. In doing so, I demonstrate how cognitive processes not often studied with respect to episodic memory, are in fact critical to our overall understanding of episodic memory as a cognitive phenomenon.

5.1 Linking Episodic Memory to Other Cognitive Functions

One of my overarching research goals is to link successful episodic memory encoding to other cognitive functions. A primary reason for my interest in this domain pertains to the ambiguity associated with many subsequent memory results. This draws from the fact, that at its core, a subsequent memory result *only* asserts that activity at encoding is different between later remembered items and later forgotten items. As such, these types of observations can fall prey to reverse inference, as most researchers hope to explain why differences between later remembered and later forgotten items

occur. Accordingly, I hoped to design a study that minimized reverse inference, while additionally providing more specific information/insight into *why* observed subsequent memory effects occur.

In Chapter 2, I utilized EEG to address this perceived explanatory failure by designing a novel subsequent memory (Paller et al., 1987)(reviewed in: (Paller & Wagner, 2002)) experimental task. While there are myriad of ways to potentially explain subsequent memory effects, I incorporated a visuo-spatial attentional task, as this type of cognitive manipulation evokes a cascade of associated well-established EEG markers of specific attentional and cognitive processes.

5.1.1 Summary

I utilized a modified visual-search task which incorporated a subsequent memory test to examine how processes associated with visuo-spatial markers of attention might contribute to subsequent memory. I additionally examined mid-frontal theta power which has been linked to cognitive control-processes and subsequent memory effects (reviewed in: (Hsieh & Ranganath, 2014)), as well as the standard “difference due to memory effect” (DM)(reviewed in: (Paller & Wagner, 2002)) which has been frequently observed in deep-encoding tasks such as this one, where participants were required to make a judgement as to whether the item was a known or unknown object, which required conceptual-content related processing.

Consistent with my hypothesis, all of the spatio-attentional effects of interest (N2pc, SPCN, and contralateral reductions in alpha power (CRA)) were found to also predict subsequent memory in the exact manner as expected. For example, when attentionally orienting towards a target-item (within a colored annulus) the hallmark N2pc was evoked, however, importantly, the magnitude of this component was larger when items were later remembered, suggesting that *stronger* attentional orienting is associated with better subsequent memory. In general, These findings established a clear link between distinct visuo-spatial attentional processes and subsequent memory.

In addition to demonstrating links between attentional processes and memory encoding, we also found that the N2pc and SPCN subsequent-memory ERP effects were found to be unrelated to those on the subsequent-memory modulations of the attention-related alpha-oscillation changes (CRA), suggesting that these attentional contributions provide separable benefits to memory encoding. I also note that the SPCN might be more reflective of a contralateral delay activity (CDA)(Luria et al., 2016), which is spatially and temporally highly similar to the SPCN, but more generally associated with working memory processing in certain task designs.

In addition to demonstrating how specific attentional processes (as indexed by the N2pc, SPCN, and CRA) aid memory encoding, I also observed that all of these attention-related effects were unrelated to oft reported DM effect. This suggests that enhanced attentional processing of stimulus items and conceptual-content related

processing of those items provide mostly independent contributions to memory encoding.

Finally, I observed that the magnitude of midfrontal theta power, both pre- and post-stimulus item, predicted subsequent memory. Additionally, pre-DM theta power predicted the magnitude of the longer-latency DM, and mediation analyses suggested that the DM mediates the relationship between early theta power and subsequent memory success. The interpretation of the theta effect is complex, as the theta power differences occurred both pre- and post- stimulus. More importantly, our task was not designed to manipulate mid-frontal power and as such differences in evoked power are necessarily more speculative. However, I also note that the task was quite difficult, requiring a high degree of cognitive control. As such, I believe that the observed theta differences might be related to trial-to-trial variability in cognitive control, which in turn modulates the effectiveness of the memory formation.

5.1.2 Future Directions

The key finding from this study was that there are unique attention processes, conceptual processes, and perhaps cognitive control processes that all contribute independently to successful memory encoding. Future directions follow two avenues: (1) involves the design of other studies, which evoke well-characterized neural markers of specific cognitive processes, to identify their links to memory encoding and (2)

involves disentangling what aspects of memory these attention and conceptual processes relate to.

The design of other studies that evoke well-characterized neural markers of specific cognitive processes is motivated by a desire to find further links between those cognitive processes and encoding success. As an example, one could design a study based on the N400 (reviewed in: (Kutas & Federmeier, 2011)), which is evoked when a semantically unexpected word arises in a sentence (e.g. *I take coffee with cream and **dog***). A standard design, which then incorporates a subsequent memory test, could determine how the processing of semantic incongruity relates to encoding success.

Another important future direction will entail disentangling what aspects of memory these attention and conceptual processes are related to. In the current design, only a memory recognition test was run. As such, all of the observed memory effects are related to recognition memory. However, other memory test could be used, for example those related to the conceptual content or the visual content of the memory item. With regards to conceptual content, one could present object words at retrieval, instead of object images, whereas for the visual content, one could present identical objects images from the encoding phase, as well as lure objects, forcing subjects to concretely identify what exact visual object they saw during encoding. Perhaps surprisingly, test of conceptual memory and visual memory can be uncorrelated (Bahrick & Boucher, 1968). As such, subsequent memory effects could be focused on memory for visual and

conceptual aspects. A simple prediction is that subsequent conceptual memory effects would be related to content processing (i.e., and thus to the DM and perhaps mid-frontal theta), whereas subsequent visual memory would be more related to the visuo-spatial attentional processes (reflected by the N2pc, SPCN, CRA). More fine-grained test might even be able to more clearly distinguish which aspects of memory these visuo-spatial attentional processes contribute to.

5.2 Linking Episodic Memory to Large-Scale Network Interactions

Another goal of my dissertation was to demonstrate that successful memory retrieval is dependent on the processing in large scale networks, rather than just one or several specific areas. Across two studies utilizing visual and verbal stimuli, I demonstrate that the hippocampus is significantly more integrated with the rest of the brain when individuals remember an item as opposed to when they forget item. Additional findings are discussed in more detail below.

5.2.1 Summary

In addition to finding that the hippocampus is more integrated with the rest of the brain when you remember something, there were numerous additional findings across the two studies as well. One was that the hippocampus significantly reorganizes its sets of connections when items are remembered (i.e. first-step reorganization). This suggests that one of the ways in which the hippocampus increases its integration is by

fundamentally changing its connectivity profile. Another common finding across both studies was that the whole brain becomes more integrated when items are remembered.

In the verbal memory study, which followed the visual memory study, I also examined the integrative properties of nodes that exhibited the strongest bivariate changes in connectivity with the hippocampus. This was motivated by the hippocampal page rank centrality from the visual study, as increased page rank centrality suggests that the nodes, a given node is connected to, are also highly integrated. And as hypothesized, I found that the nodes with strongest bivariate changes in connectivity to the hippocampus also exhibit increased integrative processing.

An additional finding from this verbal memory study was that participants who had a higher drop in modularity (from forgotten to remembered) had the best overall memory. This result was driven by the modularity in the forgotten condition -- i.e. individuals with less modular (more integrated) forgotten networks had the worst memory, whereas those with the most modular (least integrated) forgotten networks had the best memory. This is curious, as these individuals show a relatively more integrated forgotten network, which we associate with successful memory retrieval, while at the same time doing poorly on the memory task; which provides some evidence that an integrated network alone is likely insufficient for retrieval success.

5.2.2 Future Directions

Methodologically, network science and its application to neuroimaging is quite young, and the work on large-scale task-based networks in this dissertation is some of the earliest. As such, an obvious direction related to future directions would be directed toward using the most modern methods within this domain. A clear deficit of the studies presented here was the 90-node AAL atlas, which is extremely coarse. In the future, more parcellated atlases would be better to use (among other things).

Another direction is the role of the hippocampus with regards to the network. Perhaps most importantly is the diversity of processing e.g. (Ranganath, 2010) and connectivity (Frank, Bowman, & Zeithamova, 2018) present along the posterior to anterior hippocampal gradient – which my analysis, which examines only the left and right hippocampus, does not consider. Verily, more recent network analyses have focused on this gradient (Cooper & Ritchey, 2019) and have demonstrated that integration in the posterior hippocampus better supports spatial memory (i.e. location), whereas integration in the anterior hippocampus better supports item-specific memory (i.e. color).

A key finding from my results is that large-scale hippocampal interactions support memory retrieval. Thus, while the hippocampus is important, the regions it interacts with are as well. As such, in addition to examining the role of the hippocampus, another avenue of research is examining the role of hippocampal

assemblies (i.e. those regions showing the strongest interactions with the hippocampus for remembered > forgotten). For example, in Chapter 2, I demonstrate how lateralized attentional processing facilitates memory encoding, which should in turn relate to lateralized retrieval reactivation processes e.g. (Slotnick & Thakral, 2011). As consequence, I would hypothesize that the retrieval assemblies associated with remembering laterally presented items would also be lateralized, but that in both cases, the hippocampus would be more integrated than in a forgotten condition.

Manipulations, such as this one, would permit stronger claims with respect to the role of the nodes in the hippocampal assembly and how and why their increased integration aids memory retrieval.

5.3 Linking Episodic Memory to Individual Memory Traces

This methodological chapter focused on using established machine-learning techniques applied in a novel way to determine whether single-items, viewed only once, can be decoded across participants. This analysis focused on two studies, one with scenes and words and the other with objects and words. Classification was applied within modality (i.e. picture-to-picture) and across modality (i.e. picture-to-word). Both within and across modality decoding was significantly above chance, with picture-to-picture decoding being the most robust. More detailed conclusions follow.

5.3.1 Summary

While above-chance classification of single-stimuli was the primary goal of the analysis, additional analyses focused on subtler aspects of this procedure to confirm that classification of individual items was not artifactual. To this end, I wanted to ensure two aspects: (1) that the resultant confusion matrices had a logical structure to them with regards to the stimulus structure and (2) that the classifier was truly reflecting stimulus-specific information and was *not* operating on categorical level information. I also wanted to examine how category classification related to item classification.

By running a representational similarity-style analysis (Kriegeskorte et al., 2008) I demonstrated the stimulus-similarity structure was similar to the confusion matrix. This demonstrates that when the classifier is *wrong* it usually false-alarms to a similar item (e.g. classifying a dog as cat – as opposed to classifying a dog as car).

To ensure that the classifier was not relying on category information, I also tested for item-level classification within category. For example, when classifying dog, I compared classifier output to other mammals – i.e., compared to all the stimuli in the set. Even with this limitation, classification was generally above chance.

Given that I could classify individual items, even within category, and that the stimulus similarity structure was similar to the confusion matrices, I wanted to see how

the stimulus similarity affected decoding accuracy for picture-to-picture decoding³.

Here, I demonstrate that high similarity between the confusion matrix and the stimulus similarity matrix relates to lower accuracy. One interpretation of this is that when the classifier is 'confused' – it detects the category of the to-be-decoded stimulus but lacks the fidelity to detect the exact item. Put another way, if classification evidence for a category is high, classification evidence for the particular item within the category is low.

These findings have some important ramifications for tasks that rely upon categorical classification, where high category evidence is often attributed as being a positive effect, as my finding suggests that high category evidence is also related to low item-specific evidence. Therefore, conclusions suggesting that high-category evidence = high-item evidence might be misleading. It may be important to note, however, that the procedure I have followed across subjects here is quite different from how classification have usually been applied, and thus this deduction might not be applicable.

³ Within-category classification accuracy is discussed here, as I believe this better targets the goal of single-item classification.

5.3.2 Future Directions

The clear cognitive-neuroscience future direction of this finding is to examine how measures of memory relate to classification accuracy⁴. Naturally, there are fields beyond episodic memory where single-item decoding methods could be applied as well. There are also additional directions that could be pursued in further development of this methodological approach.

The algorithm utilized whole-brain, logistic regression, with a leave-one-person out (LR-LOPO) algorithm is a clear example of a brute force, black-box approach. However, the primary reason for its development and application here was to prove that single-items can be classified using across subject information. And while the results were significant and positive across modality and word-to-word, the methodology is clearly only robust with visual stimuli. As such my subsequent primary goal would be to design a more transparent and more robust algorithm.

With regards to this goal, I've designed a newer algorithm that is dependent on univariate activity alone, while also requiring no machine-learning. The resultant algorithm is more consistent with the framework utilized for encoding and decoding algorithms (Naselaris, Kay, Nishimoto, & Gallant, 2011). It currently performs just as

⁴ Yes, I did examine how classification accuracy relates to memory, however there were no positive results. I attribute this to the coarseness of the whole-brain algorithm used. As I've developed newer methods (described briefly at the end of this document), I've not investigated this question, with this method, with increased detail.

well as the LR-LOPO described herein while operating in a more transparent manner – as it requires a user-defined encoding model.

5.4 Overarching Conclusions

Episodic memory, as a cognitive construct, exists only in relation to those other cognitive constructs that reference it. Herein I have worked to demonstrate the importance of these referents - from demonstrating how specific attentional processes aid episodic memory encoding, to showing how hippocampal interactions with the whole-brain facilitate episodic memory retrieval, to developing a methodology to examine stimulus-specific memory traces – all with goal of illuminating the importance of these oft neglected referents in relation to episodic memory.

Appendix A

Note: We expect the memory encoding results from the object paradigm to be published around the same time of the submitted manuscript. As the associated manuscript (Davis et al. 2020) is still in preparation, the pertinent methodological details have been copied here. Note that two additional subjects were removed from the below analyses due to technical errors on Day 2, as such the submitted manuscript only contains data from 19 participants from the object paradigm. *If* the below manuscript (Davis et al. 2020) is *not* published by the time the submitted manuscript is accepted (i.e. Chapter 4), the information below will be incorporated into the main text⁵.

Materials & Methods

Participants

Twenty-seven healthy younger adults were recruited for this study (all native English speakers; 14 females; age mean +/- SD, 20.4 +/- 2.4 years; range 18-26 years) and participated for monetary compensation; informed consent was obtained from all participants under a protocol approved by the Duke Medical School IRB. All procedures and analyses were performed in accordance with IRB guidelines and regulations for experimental testing. Participants had no history of psychiatric or neurological disorders and were not using psychoactive drugs. Of the 27 participants tested, three participants

⁵ The following section was *not* written by me, it was written by Dr. Simon Davis and is provided only as reference. Though also note, with respect to the project, that it was a collaboration and I am a co-author on this manuscript.

were excluded due to poor performance/drowsiness during Day 1, one subject suffered a fainting episode within the MR scanner on Day 2, and two participants were subsequently removed from the analysis due to excessive motion, leaving 21 participants in the final analysis.

Stimuli

A total of 360 objects were used, from one of 12 object categories (*living* objects: 36 mammals, 22 birds, 20 fruits, 22 vegetables; *nonliving-small* objects: 29 tools, 29 clothing items, 23 food items, and 22 musical instruments; *nonliving-large* objects: 31 vehicles, 25 furniture items, 18 buildings, and 23 street-level objects); additionally, 60 catch-trial items evenly distributed from these 12 categories were also included in the design, but not included in the analysis. Isolated colored objects were shown in the center of a white background and normalized to a maximum visual angle of 7.5°.

Semantic feature norms. Semantic feature norms for the semantic feature RDM (described below) were collected online via Amazon Mechanical Turk and based on a larger database of 946 natural objects (www.cabecalab.org/cabecalabobjects). One hundred sixty-two members of the Amazon Mechanical Turk subject pool (18-62 years of age, self-reported native speakers of American English) took part in the feature norm data collection sessions, in which AMT workers were presented with a target object and space to add their features. The participants were asked to add a relation word chosen from a drop-down menu, with presets for *is*, *has*, *does*, *made of*, and “...” (participants

were instructed to use “...” as they wished for some other relationship. Participants were able to take part in repeat sessions and completed between 1 and 5 sessions. Sessions were designed to last about an hour, with 30 concepts presented within each session, and data from 30 participants per concept were used to create a feature x concept production frequency matrix.

Feature responses underwent various stages of processing before their use in the semantic RDM creation, following the procedures used by McRae et al. (2005) and Devereaux et al. (2013). These steps included: 1) removal of adverbs, such as *really* and *very*, 2) feature-splitting, for example a feature such as has a round face was rewritten as has a round face and has a face, 3) synonym mapping, which involves identifying synonyms both within and across each concept; for example “*does travel in groups*” and “*does travel in packs*” and “*does travel in a flock*” were collapsed to “*does travel in groups*”, 4) correction of spelling mistakes, 5) morphological mapping, for example “*is used in cooking*” and “*is used by cooks*” were collapsed together as “*is used in cooking*”, 6) removal of plural forms, and 7) removal of features not present in at least two concepts. At all stages in the development and use of these processing rules, results were checked manually and corrected if necessary, to guard against overgeneralized modification of the features. In the last stage of feature processing feature labels from the McCrae database: visual-color (“*is blue*”), visual-form (“*is round*”), visual-motion (“*does spin*”), smell, sound, taste, tactile, function (“*is for sleeping*”), taxonomic (“*is a mammal*”), and

encyclopedic (“lives in India”). These feature categories were used to classify features by five independent raters with strong inter-rater reliability of features across each raters (ICC > 0.8). After processing, a feature x concept production frequency matrix was created to describe the normalized frequency with a given feature is reported for a given object.

Procedure

Encoding. Each Encoding trial consisted of an initial fixation cross lasting 500 ms, followed immediately by a single letter probe for 250 ms, immediately followed by an object presented for 500 ms, followed by a blank response screen lasting between 2 and 7 s. Participants we instructed to covertly name each object (e.g., “stork”, “hammer”). In order to ensure they did so, they had to confirm that the letter probe presented before each object matched the object’s name, which was not the case in a small percentage of “catch trials.” If participants could not remember the objects’ name, they pressed a “don’t know” key. Catch trials (10%) and “don’t know” trials (mean=8%) were excluded from the analyses. The object presentation order was counterbalanced across participants, although a constant category proportion was maintained ensuring a relatively even distribution of object categories across the block. This category ordering ensures objects from the 12 different categories do not cluster in time, avoiding potential category clustering as a consequence of temporal proximity. The presentation and

timing of stimuli were controlled with Presentation (Psychology Software Tools), and naming accuracy was recorded by the experimenter during acquisition.

Retrieval. Memory was tested approximately 24 hrs later (range = 20-28 hrs, SD = 1.4 hrs) with (1) a Conceptual Memory test, which comprised a name recognition presented within the scanner, and (2) a Perceptual Memory test, which comprised a picture recognition test with similar distractors (not scanned).

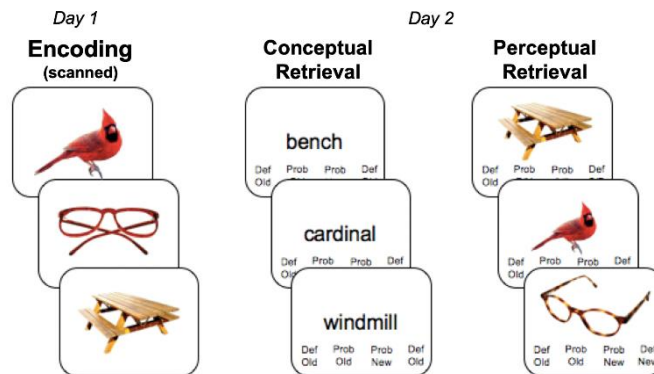


Figure 28. Task Paradigm.

MRI Acquisition

An outline of all data acquisition events is depicted in **Figure 1**. Scanning was divided between two days, 1 day apart, with the first day comprised of an Encoding session, while Day 2 comprised a Retrieval session split between a Conceptual Memory task (scanned) and Perceptual Memory task (not scanned, due to time constraints). All procedures were completed on a GE MR 750 3-Tesla scanner (General Electric 3.0 tesla Signa Excite HD short-bore scanner, equipped with an 8-channel head coil). Coplanar

functional images were acquired with an 8-channel head coil using an inverse spiral sequence with the following imaging parameters: flip angle = 77° , TR = 2000ms, TE = 31ms, FOV = 24.0 mm², and a slice thickness of 3.8mm, for 37 slices. The diffusion-weighted imaging (DWI) dataset was based on a single-shot EPI sequence (TR = 1700 ms, 50 contiguous slices of 2.0 mm thickness, FOV = 256 × 256 mm², matrix size 128 × 128, voxel size 2 × 2 × 2 mm, b-value = 1000 s/mm², 25 diffusion-sensitizing directions, total scan time ~5 min). The anatomical MRI was acquired using a 3D T1-weighted echo-planar sequence (256 × 256 matrix, TR = 12 ms, TE = 5 ms, FOV = 24 cm, 68 slices, 1.9 mm slice thickness). Scanner noise was reduced with earplugs and head motion was minimized with foam pads. Behavioral responses were recorded with a four-key fiber optic response box (Resonance Technology), and when necessary, vision was corrected using MRI-compatible lenses that matched the distance prescription used by the participant.

Functional preprocessing and data analysis were performed using SPM12 (Wellcome Department of Cognitive Neurology, London, UK) and custom MATLAB scripts. Images were corrected for slice acquisition timing, motion, and linear trend; motion correction was performed by estimating 6 motion parameters and regressing these out of each functional voxel using standard linear regression. Images were then temporally smoothed with a high-pass filter using a 190s cutoff, and normalized to the Montreal Neurological Institute (MNI) stereotaxic space. White matter and CSF signals

were also removed from the data, using WM/CSF masks and regressed from the functional data using the same method as the motion parameters. Event-related BOLD responses for correct trials were analyzed using a modified general linear model (Worsley & Friston, 1995) and RSA modelling (described below). Brain images were visualized using the FSLEyes toolbox (fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLEyes) and Surfice (www.nitrc.org/projects/surface/).

References

- Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe : A simple network model, *91*(July), 7041–7045.
- Aly, M., & Turk-browne, N. B. (2015). Attention promotes episodic encoding by stabilizing hippocampal representations. *PNAS*.
<https://doi.org/10.1073/pnas.1518931113>
- Anderson, M. C., Bunce, J. G., & Barbas, H. (2016). Neurobiology of Learning and Memory Prefrontal – hippocampal pathways underlying inhibitory control over memory. *Neurobiology of Learning and Memory*, *134*, 145–161.
<https://doi.org/10.1016/j.nlm.2015.11.008>
- Backus, A. R., Schoffelen, J. M., Szebényi, S., Hanslmayr, S., & Doeller, C. F. (2016). Hippocampal-prefrontal theta oscillations support memory integration. *Current Biology*, *26*(4), 450–457. <https://doi.org/10.1016/j.cub.2015.12.048>
- Bae, G.-Y., & Luck, S. J. (2017). Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *The Journal of Neuroscience*, *38*(2), 2860–17. <https://doi.org/10.1523/JNEUROSCI.2860-17.2017>
- Bahrack, H. P., & Boucher, B. (1968). Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, *78*(3), 417–422.
- Bai, F., Zhang, Z., Watson, D. R., Yu, H., Shi, Y., Yuan, Y., ... Qian, Y. (2009). Abnormal Functional Connectivity of Hippocampus During Episodic Memory Retrieval Processing Network in Amnesic Mild Cognitive Impairment. *Biological Psychiatry*, *65*(11), 951–958. <https://doi.org/10.1016/j.biopsych.2008.10.017>
- Bassett, D. S., & Bullmore, E. D. (2006). Small-World Brain Networks. *The Neuroscientist*, *12*(6). <https://doi.org/10.1177/1073858406293182>
- Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2010). Dynamic reconfiguration of human brain networks during learning. *Learning*, *108*(18), 7641–7646. <https://doi.org/10.1073/pnas.1018985108>
- Bastin, C., Feyers, D., Majerus, S., Balteau, E., Degueldre, C., Luxen, A., ... Collette, F. (2012). The neural substrates of memory suppression: a fMRI exploration of directed forgetting. *PloS One*, *7*(1), e29905.
<https://doi.org/10.1371/journal.pone.0029905>
- Battaglia, F. P., Benchenane, K., Sirota, A., Pennartz, C. M. a., & Wiener, S. I. (2011). The hippocampus: hub of brain network communication for memory. *Trends in Cognitive Sciences*, (July 2015). <https://doi.org/10.1016/j.tics.2011.05.008>
- Bertolero, M. A., Yeo, B. T. T., & D'Esposito, M. (2015). The modular and integrative

- functional architecture of the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 112(49), E6798-807.
<https://doi.org/10.1073/pnas.1510619112>
- Bhandari, A., Gagne, C., & Badre, D. (2018). Just above chance: Is it Harder to Decode Information from Prefrontal Cortex Hemodynamic Activity Patterns? *Journal of Cognitive Neuroscience*, 30(10), 1473–1498. <https://doi.org/10.1162/jocn>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/p10008>
- Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal Cortex and Long-Term Memory Encoding: An Integrative Review of Findings from Neuropsychology and Neuroimaging. *The Neuroscientist*, 13(3), 280–291.
<https://doi.org/10.1177/1073858407299290>
- Boldi, P., Santini, M., & Vigna, S. (2009). PageRank: Functional Dependencies. *ACM Trans. Inf. Syst.*, 27(4). <https://doi.org/10.1145/1629096.1629097>
- Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., ... Bassett, D. S. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 112(37), 11678–11683.
<https://doi.org/10.1073/pnas.1422487112>
- Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition : emerging methods and principles. *Trends in Cognitive Sciences*, 14(6), 277–290.
<https://doi.org/10.1016/j.tics.2010.04.004>
- Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., ... Johnson, K. A. (2009). Cortical Hubs Revealed by Intrinsic Functional Connectivity : Mapping , Assessment of Stability , and Relation to Alzheimer ' s Disease, 29(6), 1860–1873. <https://doi.org/10.1523/JNEUROSCI.5062-08.2009>
- Bullmore, E., & Sporns, O. (2009). Complex brain networks : graph theoretical analysis of structural and functional systems. *Nature Reviews: Neuroscience*, 10.
<https://doi.org/10.1038/nrn2575>
- Burianova, H., & Grady, C. L. (2007). Common and unique neural activations in autobiographical, episodic, and semantic retrieval. *Journal of Cognitive Neuroscience*, 19(9), 1520–1534. <https://doi.org/10.1162/jocn.2007.19.9.1520>
- Bussey, T. J., Saksida, L. M., & Elisabeth, A. (2005). The perceptual-mnemonic / feature conjunction model of perirhinal cortex function. *Journal of Experimental Psychology*, (March 2014), 37–41.

- Cabeza, R. (2008). Role of the parietal regions in episodic memory retrieval: the dual attentional processes hypothesis. *Neuropsychologia*, 46(7), 1813–1827.
- Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews: Neuroscience*, 9(8), 613–625. <https://doi.org/10.1038/nrn2459>
- Cabeza, R., & Nyberg, L. (2000). Neural bases of learning and memory : functional neuroimaging evidence. *Current Opinion in Neurobiology*, (13), 415–421.
- Cao, M., Wang, J.-H., Dai, Z.-J., Cao, X.-Y., Jiang, L.-L., Fan, F.-M., ... He, Y. (2014). Topological organization of the human brain functional connectome across the lifespan. *Developmental Cognitive Neuroscience*, 7(16), 76–93. <https://doi.org/10.1016/j.dcn.2013.11.004>
- Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience*, 15(5), 704–717. <https://doi.org/10.1162/089892903322307429>
- Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, 18(8), 414–421. <https://doi.org/10.1016/j.tics.2014.04.012>
- Chastelaine, M. De, & Rugg, M. D. (2014). The relationship between task-related and subsequent memory effects. *Human Brain Mapping*, 35(8), 3687–3700. <https://doi.org/10.1002/hbm.22430>
- Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: a hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, 46(7), 1828–1851. <https://doi.org/10.1016/j.neuropsychologia.2008.03.022>
- Cisler, J. M., Bush, K., & Steele, J. S. (2014). A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. *NeuroImage*, 84, 1042–1052. <https://doi.org/10.1016/j.neuroimage.2013.09.018>
- Clithero, J. A., Smith, D. V., Carter, R. M., & Huettel, S. A. (2011). Within- and cross-participant classifiers reveal different neural coding of information. *NeuroImage*, 56(2), 699–708. <https://doi.org/10.1016/j.neuroimage.2010.03.057>
- Cohen, J. R., Sreenivasan, K. K., & D’Esposito, M. (2014). Correspondence between stimulus encoding- and maintenance-related neural processes underlies successful working memory. *Cerebral Cortex*, 24(3), 593–599. <https://doi.org/10.1093/cercor/bhs339>
- Connolly, A. C., Swaroop Guntupalli, J., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., ... Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, 32(8), 2608–2618.

<https://doi.org/10.1523/JNEUROSCI.5547-11.2012>

- Cooper, R. A., & Ritchey, M. (2019). Cortico-hippocampal network connections support the multidimensional quality of episodic memory. *ELife*, 1–22.
- Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, 19(2 Pt 1), 261–270. [https://doi.org/10.1016/s1053-8119\(03\)00049-1](https://doi.org/10.1016/s1053-8119(03)00049-1)
- Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, 16(6), 763–770. <https://doi.org/10.1038/nn.3381>
- Damasio, A. R. (1989). The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones. *Neural Computation*, 1(1), 123–132. <https://doi.org/10.1162/neco.1989.1.1.123>
- Danker, J. F., & Anderson, J. R. (2010). The Ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, 136(1), 87–102. <https://doi.org/10.1037/a0017937>.The
- Davidson, P. S. R., Troyer, A. K., & Moscovitch, M. (2006). Frontal lobe contributions to recognition and recall: linking basic research with clinical evaluation and remediation. *Journal of the International Neuropsychological Society : JINS*, 12(2), 210–223. <https://doi.org/10.1017/S1355617706060334>
- De Vries, I., van Driel, J., & Olivers, C. N. L. (2017). Posterior alpha EEG Dynamics Dissociate Current from Future Goals in Working Memory-Guided Visual Search. *Journal of Neuroscience*. <https://doi.org/10.1523/JNEUROSCI.2945-16.2016>
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 33(48), 18906–18916. <https://doi.org/10.1523/JNEUROSCI.3809-13.2013>
- Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11), 2229–2243. <https://doi.org/10.1002/asi.21171>
- Donohue, S., Hopf, J.-M., Bartsch, M., Schoenfeld, M., Heinze, H.-J., & Woldorff, M. (2013). The Rapid Capture of Attention by Rewarded Objects. *J. of Cog. Neuro.*, 26(3),

- 194–198. <https://doi.org/10.1162/jocn>
- Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., ... Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *PNAS*, *104*(26), 11073–11078.
- Eichenbaum, H., Yonelinas, A. R., & Ranganath, C. (2007). The Medial Temporal Lobe and Recognition Memory. *Annu Rev Neurosci.*, *30*, 130–134.
<https://doi.org/10.1016/j.pestbp.2011.02.012>.Investigations
- Eimer, M., & Kiss, M. (2007). Attentional capture by task-irrelevant fearful faces is revealed by the N2pc component. *Biological Psychology*, *74*(1), 108–112.
<https://doi.org/https://doi.org/10.1016/j.biopsycho.2006.06.008>
- Emrich, S. M., Riggall, A. C., LaRocque, J. J., & Postle, B. R. (2013). Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. *Journal of Neuroscience*, *33*(15), 6516–6523.
<https://doi.org/10.1523/JNEUROSCI.5732-12.2013>
- Fairhall, S. L., & Caramazza, A. (2013). Brain Regions That Represent Amodal Conceptual Knowledge. *Journal of Neuroscience*, *33*(25), 10552–10558.
<https://doi.org/10.1523/JNEUROSCI.0051-13.2013>
- Fornito, A., Yoon, J., Zalesky, A., Bullmore, E. T., & Carter, C. S. (2011). General and specific functional connectivity disturbances in first-episode schizophrenia during cognitive control performance. *Biological Psychiatry*, *70*(1), 64–72.
<https://doi.org/10.1016/j.biopsych.2011.02.019>
- Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews. Neuroscience*, *16*(3), 159–172. <https://doi.org/10.1038/nrn3901>
- Frank, L. E., Bowman, C. R., & Zeithamova, D. (2018). Differential Functional Connectivity along the Long Axis of the Hippocampus Aligns with Differential Role in Memory Specificity and Generalization. *Journal of Cognitive Neuroscience*, (2015), 1–18.
- Fukuda, K., Mance, I., & Vogel, E. K. (2015). Power Modulation and Event-Related Slow Wave Provide Dissociable Correlates of Visual Working Memory. *Journal of Neuroscience*, *35*(41), 14009–14016. <https://doi.org/10.1523/JNEUROSCI.5003-14.2015>
- Garrido, M. I., Barnes, G. R., Kumaran, D., Maguire, E. A., & Dolan, R. J. (2015). Ventromedial prefrontal cortex drives hippocampal theta oscillations induced by mismatch computations. *NeuroImage*, *120*, 362–370.
<https://doi.org/10.1016/j.neuroimage.2015.07.016>
- Geib, B. R., Stanley, M. L., Wing, E. A., Laurienti, P. J., & Cabeza, R. (2017). Hippocampal Contributions to the Large-Scale Episodic Memory Network Predict Vivid Visual

- Memories. *Cerebral Cortex*, 27, 680–693. <https://doi.org/10.1093/cercor/bhv272>
- Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C., & Beaulieu, C. (2008). Mapping Anatomical Connectivity Patterns of Human Cerebral Cortex Using In Vivo Diffusion Tensor Imaging Tractography. *Cerebral Cortex*, 19(3), 524–536. <https://doi.org/10.1093/cercor/bhn102>
- Gordon, A. M., Rissman, J., Kiani, R., & Wagner, A. D. (2014). Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cerebral Cortex*, 24(12), 3350–3364. <https://doi.org/10.1093/cercor/bht194>
- Grent-T-Jong, T., Boehler, C. N., Kenemans, J. L., & Woldorff, M. G. (2011). Differential functional roles of slow-wave and oscillatory-alpha activity in visual sensory cortex during anticipatory visual-spatial attention. *Cerebral Cortex*, 21(10), 2204–2216. <https://doi.org/10.1093/cercor/bhq279>
- Guimerà, R., & Nunes Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *Nature*, 433(7028), 895–900. <https://doi.org/10.1038/nature03288>
- Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex*, bhw068. <https://doi.org/10.1093/cercor/bhw068>
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, 6(7), e159. <https://doi.org/10.1371/journal.pbio.0060159>
- Handwerker, D. a., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21(4), 1639–1651. <https://doi.org/10.1016/j.neuroimage.2003.11.029>
- Hart, J., Maguire, M. J., Motes, M., Mudar, R. A., Chiang, H.-S., Womack, K. B., & Kraut, M. a. (2013). Semantic memory retrieval circuit: role of pre-SMA, caudate, and thalamus. *Brain and Language*, 126(1), 89–98. <https://doi.org/10.1016/j.bandl.2012.08.002>
- Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404–416. <https://doi.org/10.1016/j.neuron.2011.08.026>
- Haxby, J. V, Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience*. <https://doi.org/10.1146/annurev-neuro-062012-170325>

- Hayasaka, S., & Laurienti, P. J. (2010). Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. *NeuroImage*, *50*(2), 499–508. <https://doi.org/10.1016/j.neuroimage.2009.12.051>
- He, Y., Wang, J., Wang, L., Chen, Z. J., Yan, C., Yang, H., ... Evans, A. C. (2009). Uncovering Intrinsic Modular Organization of Spontaneous Brain Activity in Humans. *PLOS ONE*, *4*(4), e5226. Retrieved from <https://doi.org/10.1371/journal.pone.0005226>
- Henderson, J. M., Larson, C. L., & Zhu, D. C. (2007). Cortical activation to indoor versus outdoor scenes: An fMRI study. *Experimental Brain Research*, *179*(1), 75–84. <https://doi.org/10.1007/s00221-006-0766-2>
- Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, 1–7. <https://doi.org/10.1038/nn.4284>
- Hsieh, L.-T., & Ranganath, C. (2014). Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *NeuroImage*, *85 Pt 2*, 721–729. <https://doi.org/10.1016/j.neuroimage.2013.08.003>
- Huijbers, W., Pennartz, C. M. A., Cabeza, R., & Daselaar, S. M. (2011). The hippocampus is coupled with the default network during memory retrieval but not during memory encoding. *PLoS ONE*, *6*(4). <https://doi.org/10.1371/journal.pone.0017463>
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, (532), 453–458. <https://doi.org/10.1038/nature>
- Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*, *76*(6), 1210–1224. <https://doi.org/10.1016/j.neuron.2012.10.014>
- Jolicoeur, P., Brisson, B., & Robitaille, N. (2008). Dissociation of the N2pc and sustained posterior contralateral negativity in a choice response task. *Brain Research*, *1215*, 160–172. <https://doi.org/10.1016/j.brainres.2008.03.059>
- Joyce, K. E., Laurienti, P. J., Burdette, J. H., & Hayasaka, S. (2010). A New Measure of Centrality for Brain Networks. *PLOS ONE*, *5*(8), e12200. Retrieved from <https://doi.org/10.1371/journal.pone.0012200>
- Khader, P., Ranganath, C., Seemüller, A., & Rösler, F. (2007). Working memory maintenance contributes to long-term memory formation: Evidence from slow event-related brain potentials. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 212–224. <https://doi.org/10.3758/CABN.7.3.212>

- Kim, H. (2013). Differential neural activity in the recognition of old versus new events: An Activation Likelihood Estimation Meta-Analysis. *Human Brain Mapping, 34*, 814–836. <https://doi.org/10.1002/hbm.21474>
- Kim, H. (2015). Encoding and Retrieval Along the Long Axis of the Hippocampus and Their Relationships With Dorsal Attention and Default Mode Networks : The HERNET Model. *Hippocampus, 25*, 500–510. <https://doi.org/10.1002/hipo.22387>
- Kim, H. (2016). Default network activation during episodic and semantic memory retrieval: A selective meta-analytic comparison. *Neuropsychologia, 80*, 35–46. <https://doi.org/10.1016/j.neuropsychologia.2015.11.006>
- Kim, J., Kim, Y.-H., & Lee, J.-H. (2013). Hippocampus-precuneus functional connectivity as an early sign of Alzheimer’s disease: a preliminary study using structural and functional magnetic resonance imaging data. *Brain Research, 1495*, 18–29. <https://doi.org/10.1016/j.brainres.2012.12.011>
- King, D. R., de Chastelaine, M., Elward, R. L., Wang, T. H., & Rugg, M. D. (2015). Recollection-Related Increases in Functional Connectivity Predict Individual Differences in Memory Accuracy. *Journal of Neuroscience, 35*(4), 1763–1772. <https://doi.org/10.1523/JNEUROSCI.3219-14.2015>
- Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology, 45*(2), 240–249. <https://doi.org/10.1111/j.1469-8986.2007.00611.x>
- Kragel, P. A., & LaBar, K. S. (2014). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience, 10*(11), 1437–1448. <https://doi.org/10.1093/scan/nsv032>
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*(November), 4. <https://doi.org/10.3389/neuro.06.004.2008>
- Kuhl, B. A., & Chun, M. M. (2014). Successful Remembering Elicits Event-Specific Activity Patterns in Lateral Parietal Cortex. *J Neurosci., 34*(23), 8051–8060. <https://doi.org/10.1523/JNEUROSCI.4328-13.2014>
- Kuhl, B. A., Bainbridge, W., & Chun, M. (2012). Neural reactivation reveals mechanisms for updating memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 32*(10), 3453–3461. <https://doi.org/10.1523/JNEUROSCI.5846-11.2012>
- Kuhl, B. A., Rissman, J., & Wagner, A. D. (2012). Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia, 50*, 458–469. <https://doi.org/10.1016/j.neuropsychologia.2011.09.002>

- Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *PNAS*, *108*(14), 5903–5908. <https://doi.org/10.1073/pnas.1016939108>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(August), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Laufs, H., Kleinschmidt, A., Beyerle, A., Eger, E., Salek-Haddadi, A., Preibisch, C., & Krakow, K. (2003). EEG-correlated fMRI of human alpha activity. *NeuroImage*, *19*(4), 1463–1476.
- Lee, H., Chun, M. M., & Kuhl, B. A. (2016). Lower Parietal Encoding Activation Is Associated with Sharper Information and Better Memory. *Cerebral Cortex (New York, N.Y. : 1991)*, bhw097. <https://doi.org/10.1093/cercor/bhw097>
- Lehericy, S., Ducros, M., Van de Moortele, P.-F., Francois, C., Thivard, L., Poupon, C., ... Kim, D.-S. (2004). Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Annals of Neurology*, *55*(4), 522–529. <https://doi.org/10.1002/ana.20030>
- Liu, Y., Yu, C., Zhang, X., Liu, J., Duan, Y., Alexander-Bloch, A. F., ... Bullmore, E. (2014). Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease. *Cerebral Cortex (New York, N.Y. : 1991)*, *24*(6), 1422–1435. <https://doi.org/10.1093/cercor/bhs410>
- Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience)*. The MIT Press.
- Luck, S. J., & Hillyard, S. (1994). Spatial filtering during visual search: evidence from human electrophysiology. *J Exp Psychol Hum Percept Perform*. <https://doi.org/10.1037/0096-1523.20.5.1000>
- Lundstrom, B. N., Ingvar, M., & Petersson, K. M. (2005). The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *NeuroImage*, *27*(4), 824–834. <https://doi.org/10.1016/j.neuroimage.2005.05.008>
- Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience and Biobehavioral Reviews*, *62*, 100–108. <https://doi.org/10.1016/j.neubiorev.2016.01.003>
- Mack, M. L., Love, B. C., & Preston, A. R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *PNAS*. <https://doi.org/10.1073/pnas.1614048113>
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are

- complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*. US: American Psychological Association.
<https://doi.org/10.1037/0033-295X.102.3.419>
- McCormick, C., Moscovitch, M., Protzner, A. B., Huber, C. G., & McAndrews, M. P. (2010). Hippocampal-neocortical networks differ during encoding and retrieval of relational memory: Functional and effective connectivity analyses. *Neuropsychologia*, 48(11), 3272–3281. <https://doi.org/10.1016/j.neuropsychologia.2010.07.010>
- McIntosh, A. R., Nyberg, L., Bookstein, F. L., & Tulving, E. (1997). Differential functional connectivity of prefrontal and medial temporal cortices during episodic memory retrieval. *Human Brain Mapping*, 5(4), 323–327. [https://doi.org/10.1002/\(SICI\)1097-0193\(1997\)5:4<323::AID-HBM20>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-0193(1997)5:4<323::AID-HBM20>3.0.CO;2-D)
- McLelland, V., Chan, D., Ferber, S., & Barense, M. (2014). Stimulus familiarity modulates functional connectivity of the perirhinal cortex and anterior hippocampus during visual discrimination of faces and objects. *Frontiers in Human Neuroscience*, 8, 117. <https://doi.org/10.3389/fnhum.2014.00117>
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4), 547–559. <https://doi.org/10.3758/BF03192726>
- Medaglia, J. D., Lynall, M.-E., & Bassett, D. S. (2015). Cognitive Network Neuroscience. *Journal of Cognitive Neuroscience*, 27(8), 1471–1491. https://doi.org/10.1162/jocn_a_00810
- Meunier, D., Fonlupt, P., Saive, A.-L., Plailly, J., Ravel, N., & Royet, J.-P. (2014). Modular structure of functional networks in olfactory memory. *NeuroImage*, 95, 264–275. <https://doi.org/10.1016/j.neuroimage.2014.03.041>
- Mišić, B., Goñi, J., Betzel, R. F., Sporns, O., & McIntosh, A. R. (2014). A Network Convergence Zone in the Hippocampus. *PLoS Computational Biology*, 10(12), e1003982. <https://doi.org/10.1371/journal.pcbi.1003982>
- Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, 135(4), 638–677. <https://doi.org/10.1037/a0015849>
- Mizuta, H., & Motomura, N. (2006). Memory dysfunction in caudate infarction caused by Heubner's recurring artery occlusion. *Brain and Cognition*, 61(2), 133–138. <https://doi.org/10.1016/j.bandc.2005.11.002>
- Moll, M., & Miikkulainen, R. (1997). Convergence-Zone episodic memory: Analysis and simulations. *Neural Networks*. Netherlands: Elsevier Science.

[https://doi.org/10.1016/S0893-6080\(97\)00016-6](https://doi.org/10.1016/S0893-6080(97)00016-6)

- Moussa, Malaak N, Steen, M. R., Laurienti, P. J., & Hayasaka, S. (2012). Consistency of Network Modules in Resting-State fMRI Connectome Data. *PLOS ONE*, 7(8), e44428. Retrieved from <https://doi.org/10.1371/journal.pone.0044428>
- Moussa, Malaak Nasser, Vechlekar, C. D., Burdette, J. H., Steen, M. R., Hugenschmidt, C. E., & Laurienti, P. J. (2011). Changes in cognitive state alter human functional brain networks. *Frontiers in Human Neuroscience*, 5(August), 83. <https://doi.org/10.3389/fnhum.2011.00083>
- Moussa, Malaak Nasser, Wesley, M. J., Porrino, L. J., Hayasaka, S., Bechara, A., Burdette, J. H., & Laurienti, P. J. (2014). Age-related differences in advantageous decision making are associated with distinct differences in functional community structure. *Brain Connectivity*, 4(3), 193–202. <https://doi.org/10.1089/brain.2013.0184>
- Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, 59, 2636–2643. <https://doi.org/10.1016/j.neuroimage.2011.08.076>
- Musz, E., & Thompson-Schill, S. L. (2017). Tracking competition and cognitive control during language comprehension with multi-voxel pattern analysis. *Brain and Language*, 165, 21–32. <https://doi.org/10.1016/j.bandl.2016.11.002>
- Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, 56(2), 400–410. <https://doi.org/10.1016/j.neuroimage.2010.07.073>
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 8577 LP – 8582. <https://doi.org/10.1073/pnas.0601602103>
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2), 26113. <https://doi.org/10.1103/PhysRevE.69.026113>
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cogn Sci.*, 10(9), 424–430. <https://doi.org/10.1016/j.tics.2006.07.005>
- Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain the neurophysics of EEG* (2nd ed.). Oxford ; New York: Oxford University Press. Retrieved from http://getit@duke.library.duke.edu/?sid=sersol%7B&%7DSS%7B_%7Djc=TC0000086519%7B&%7Dtitle=Electric
- O’Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17(4), 580–590. <https://doi.org/10.1162/0898929053467550>

- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*, 2011, 156869. <https://doi.org/10.1155/2011/156869>
- Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an identical learning paradigm. *EEG and Clinical Neurophysiology*, 67, 360–371.
- Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cog Sci.*, 6(2), 93–102.
- Papanicolaou, A. C., Simos, P. G., Castillo, E. M., Breier, J. I., Katz, J. S., & Wright, A. A. (2002). The Hippocampus and Memory of Verbal and Pictorial Material. *Learning & Memory*. <https://doi.org/10.1101/lm.44302>.Squire
- Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, 72(5), 692–697. <https://doi.org/10.1016/j.neuron.2011.11.001>.Inferring
- Poldrack, R. A., Halchenko, Y., & Jos, S. (2009). Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, 1–16.
- Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(23), 1963–1966. <https://doi.org/10.1126/science.1117645>
- Power, J. D., Fair, D. A., Schlaggar, B. L., & Petersen, S. E. (2010). The development of Human Functional Brain Networks. *Neuron*, 67(5), 735–748. <https://doi.org/10.1016/j.neuron.2010.08.017>.The
- Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current Biology*, 23(17), R764–R773. <https://doi.org/10.1016/j.cub.2013.05.041>
- Qi, S., Zeng, Q., Ding, C., & Li, H. (2013). Neural correlates of reward-driven attentional capture in visual search. *Brain Research*, 1532, 32–43. <https://doi.org/10.1016/j.brainres.2013.07.044>
- Ranganath, C. (2010). Binding items and contexts: The cognitive neuroscience of episodic memory. *Current Directions in Psychological Science*, 19(3), 131–137. <https://doi.org/10.1177/0963721410368805>
- Ranganath, C., Heller, A., Cohen, M. X., Brozinsky, C. J., & Rissman, J. (2005). Functional connectivity with the hippocampus during successful memory formation. *Hippocampus*, 15(8), 997–1005. <https://doi.org/10.1002/hipo.20141>
- Reas, E. T., & Brewer, J. B. (2013). Retrieval search and strength evoke dissociable brain

- activity during episodic memory recall. *Journal of Cognitive Neuroscience*, 25(2), 219–233. https://doi.org/10.1162/jocn_a_00335
- Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, 23(2), 752–763. <https://doi.org/10.1016/j.neuroimage.2004.06.035>
- Ritchey, M., Wing, E. A., Labar, K. S., & Cabeza, R. (2013). Neural Similarity Between Encoding and Retrieval is Related to Memory Via Hippocampal Interactions. *Cerebral Cortex*, 23, 2818–2828. <https://doi.org/10.1093/cercor/bhs258>
- Robin, J., Hirshhorn, M., Rosenbaum, R. S., Winocur, G., Moscovitch, M., & Grady, C. L. (2015). Functional Connectivity of Hippocampal and Prefrontal Networks During Episodic and Spatial Memory Based on Real-World Environments. *Hippocampus*, 93, 81–93. <https://doi.org/10.1002/hipo.22352>
- Robinson, J. L., Laird, A. R., Glahn, D. C., Blangero, J., Sanghera, M. K., Pessoa, L., ... Fox, P. T. (2012). The functional connectivity of the human caudate: an application of meta-analytic connectivity modeling with behavioral filtering. *NeuroImage*, 60(1), 117–129. <https://doi.org/10.1016/j.neuroimage.2011.12.010>
- Rolls, E. T. (2000). Hippocampo-Cortical and Cortico-Cortical Backprojections. *Hippocampus*, 10, 380–388.
- Rose, N., LaRocque, J., & Postle, B. (2017). Reactivation of latent WM with TMS. *Science*, 511–528. <https://doi.org/10.1002/9783527610174.ch4d>
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52(3), 1059–1069. <https://doi.org/10.1016/j.neuroimage.2009.10.003>
- Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, 23(2), 255–260. <https://doi.org/10.1016/j.conb.2012.11.005>
- Rugg, M. D., Vilberg, K. L., Mattson, J. T., Yu, S. S., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: fMRI evidence. *Neuropsychologia*, 50(13), 3070–3079. <https://doi.org/10.1016/j.neuropsychologia.2012.06.004>
- Ryan, L., Cardoza, J. A., Barense, M. D., Kawa, K. H., Wallentin-Flores, J., Arnold, W. T., & Alexander, G. E. (2012). Age-related impairment in a complex object discrimination task that engages perirhinal cortex. *Hippocampus*, 22(10), 1978–1989. <https://doi.org/10.1002/hipo.22069>
- Rzucidlo, J. K., Roseman, P. L., Laurienti, P. J., & Dagenbach, D. (2013). Stability of Whole Brain and Regional Network Topology within and between Resting and

- Cognitive States. *PLoS ONE*, 8(8), e70275.
<https://doi.org/10.1371/journal.pone.0070275>
- Sadeh, T., Shohamy, D., Levy, D. R., Reggev, N., & Maril, A. (2010). Cooperation between the Hippocampus and the Striatum during Episodic Encoding. *Journal of Cognitive Neuroscience*, 23(7), 1597–1608. <https://doi.org/10.1162/jocn.2010.21549>
- Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., & Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex*, 15(9), 1332–2342. <https://doi.org/10.1093/cercor/bhi016>
- Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, 26(1), 3–8. <https://doi.org/10.1002/hipo.22523>
- Schedlbauer, A. M., Copara, M. S., Watrous, A. J., & Ekstrom, A. D. (2014). Multiple interacting brain areas underlie successful spatiotemporal memory retrieval in humans. *Scientific Reports*, 4(6431), 1–9. <https://doi.org/10.1038/srep06431>
- Schott, B. H., Wüstenberg, T., Wimber, M., Fenker, D. B., Zierhut, K. C., Seidenbecher, C. I., ... Richardson-Klavehn, A. (2013). The relationship between level of processing and hippocampal-cortical functional connectivity during episodic memory formation in humans. *Human Brain Mapping*, 34(2), 407–424. <https://doi.org/10.1002/hbm.21435>
- Schubert, András. (2013). Measuring the Similarity between the Reference and Citation Distributions of Journals. *Scientometrics*, 96(1), 305–313. <https://doi.org/10.1007/s11192-012-0889-0>
- Schubert, Andras, & Telecs, A. (2014). A note on the Jaccardized Czekanowski similarity index. *Scientometrics*, 1397–1399. <https://doi.org/10.1007/s11192-013-1044-2>
- Simpson, S. L., Lyday, R. G., Hayasaka, S., Marsh, A. P., & Laurienti, P. J. (2013). A permutation testing framework to compare groups of brain networks. *Frontiers in Computational Neuroscience*, 7, 171. <https://doi.org/10.3389/fncom.2013.00171>
- Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia*, 45(10), 2163–2179. <https://doi.org/10.1016/j.neuropsychologia.2007.03.007>
- Slotnick, S. D., & Thakral, P. P. (2011). Memory for motion and spatial location is mediated by contralateral and ipsilateral motion processing cortex. *NeuroImage*, 55(2), 794–800. <https://doi.org/10.1016/j.neuroimage.2010.11.077>
- Spaniol, J., Davidson, P. S. R., Kim, A. S. N., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. *Neuropsychologia*, 47(8–9), 1765–

1779. <https://doi.org/10.1016/j.neuropsychologia.2009.02.028>
- Sporns, O. (2013). Network attributes for segregation and integration in the human brain. *Current Opinion in Neurobiology*, 23(2), 162–171. <https://doi.org/10.1016/j.conb.2012.11.015>
- Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, 16(12), 1879–1887. <https://doi.org/10.1038/nn.3574>
- Stankevich, B. A., & Geng, J. J. (2014). Reward associations and spatial probabilities produce additive effects on attentional selection. *Attention, Perception & Psychophysics*, (June). <https://doi.org/10.3758/s13414-014-0720-5>
- Stanley, M. L., Dagenbach, D., Lyday, R. G., Burdette, J. H., & Laurienti, P. J. (2014). Changes in global and regional modularity associated with increasing working memory load. *Front Hum Neurosci*, (December). <https://doi.org/10.3389/fnhum.2014.00954>
- Stanley, M. L., Moussa, M. N., Paolini, B. M., Lyday, R. G., Burdette, J. H., & Laurienti, P. J. (2013). Defining nodes in complex brain networks. *Frontiers in Cognitive Neuroscience*, 7(November), 1–14. <https://doi.org/10.3389/fncom.2013.00169>
- Stanley, M. L., Simpson, S. L., Dagenbach, D., Lyday, R. G., Burdette, J. H., & Laurienti, P. J. (2015). Changes in Brain Network Efficiency and Working Memory Performance in Aging. *PLOS ONE*, 10(4), e0123950. Retrieved from <https://doi.org/10.1371/journal.pone.0123950>
- Staresina, B. P., Cooper, E., & Henson, R. N. (2013). Reversible Information Flow across the Medial Temporal Lobe: The Hippocampus Links Cortical Modules during Memory Retrieval. *The Journal of Neuroscience*, 33(35), 14184–14192. <https://doi.org/10.1523/JNEUROSCI.1987-13.2013>
- Stevens, A. A., Tappon, S. C., Garg, A., & Fair, D. A. (2012). Functional Brain Network Modularity Captures Inter-and Intra-Individual Variation in Working Memory Capacity. *PLoS ONE*, 7(1), e30468. <https://doi.org/10.1371/journal.pone.0030468>
- Takahashi, E., Ohki, K., & Kim, D.-S. (2008). Dissociated pathways for successful memory retrieval from the human parietal cortex: anatomical and functional connectivity analyses. *Cerebral Cortex (New York, N.Y. : 1991)*, 18(8), 1771–1778. <https://doi.org/10.1093/cercor/bhm204>
- Telesford, Q. K., Simpson, S. L., Burdette, J. H., Hayasaka, S., & Laurienti, P. J. (2011). The brain as a complex system: using network science as a tool for understanding the brain. *Brain Connectivity*, 1(4), 295–308. <https://doi.org/10.1089/brain.2011.0055>

- Teyler, T. J., & Rudy, J. W. (2007). The hippocampal indexing theory and episodic memory: updating the index. *Hippocampus*, *17*(12), 1158–1169.
<https://doi.org/10.1002/hipo>
- Thakral, P. P., Wang, T. H., & Rugg, M. D. (2017). Decoding the content of recollection within the core recollection network and beyond. *Cortex*, *91*, 101–113.
<https://doi.org/10.1016/j.cortex.2016.12.011>
- Tomasi, D., & Volkow, N. D. (2010). Functional connectivity density mapping. *Proceedings of the National Academy of Sciences*, *107*(21), 9885 LP – 9890.
<https://doi.org/10.1073/pnas.1001414107>
- Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, *91*(11), 5033–5037.
<https://doi.org/10.1073/pnas.91.11.5033>
- Turk-Browne, N. B., Golomb, J. D., & Chun, M. M. (2013). Complementary attentional components of successful memory encoding. *NeuroImage*, *66*, 553–562.
<https://doi.org/10.1016/j.neuroimage.2012.10.053>
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. <https://doi.org/10.1006/nimg.2001.0978>
- Uncapher, M. R., Hutchinson, J. B., & Wagner, A. D. (2011). Dissociable Effects of Top-Down and Bottom-Up Attention during Episodic Encoding. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(35), 12613–12628.
<https://doi.org/10.1523/JNEUROSCI.0152-11.2011>
- van den Heuvel, M. P., Kahn, R. S., Goñi, J., & Sporns, O. (2012). High-cost, high-capacity backbone for global brain communication. *Proceedings of the National Academy of Sciences*, *109*(28), 11372 LP – 11377.
<https://doi.org/10.1073/pnas.1203593109>
- van den Heuvel, M. P., & Sporns, O. (2013). Network hubs in the human brain. *Trends in Cognitive Sciences*, *17*(12), 683–696. <https://doi.org/10.1016/j.tics.2013.09.012>
- van den Heuvel, M. P., Stam, C. J., Kahn, S., & Pol, H. E. H. (2009). Efficiency of Functional Brain Networks and Intellectual Performance. *The Journal of Neuroscience*, *29*(23), 7619–7624. <https://doi.org/10.1523/JNEUROSCI.1443-09.2009>
- van Diepen, R. M., Miller, L., Mazaheri, A., & Geng, J. J. (2016). The role of alpha activity in spatial and featured-based attention. *ENeuro*.
<https://doi.org/10.1523/ENEURO.0204-16.2016>

- Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: a review of evidence from a dual-process perspective. *Neuropsychologia*, *46*(7), 1787–1799. <https://doi.org/10.1016/j.pestbp.2011.02.012>.Investigations
- Vogel, A. C., Church, J. A., Power, J. D., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2013). Brain & Language Functional network architecture of reading-related regions across development. *Brain and Language*, *125*(2), 231–243. <https://doi.org/10.1016/j.bandl.2012.12.016>
- Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, *59*(6), 1037–1050. <https://doi.org/10.1016/j.neuron.2008.09.006>
- Wager, T. D., Waugh, C. E., Lindquist, M., Noll, D. C., Fredrickson, B. L., & Taylor, S. F. (2009). Brain mediators of cardiovascular responses to social threat. Part I: Reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. *NeuroImage*, *47*(3), 821–835. <https://doi.org/10.1016/j.neuroimage.2009.05.043>
- Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453. <https://doi.org/https://doi.org/10.1016/j.tics.2005.07.001>
- Wang, L., Li, Y., Metzak, P., He, Y., & Woodward, T. S. (2010). Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition. *NeuroImage*, *50*(3), 862–872. <https://doi.org/10.1016/j.neuroimage.2010.01.044>
- Watrous, A. J., & Ekstrom, A. D. (2014). The spectro-contextual encoding and retrieval theory of episodic memory. *Frontiers in Human Neuroscience*, *8*(February), 75. <https://doi.org/10.3389/fnhum.2014.00075>
- Westfall, J., Nichols, T. E., & Yarkoni, T. (2017). Fixing the stimulus-as-fixed-effect fallacy in task fMRI. *Wellcome Open Research*, *1*(May), 1–24. <https://doi.org/10.12688/wellcomeopenres.10298.1>
- Westphal, A. J., Wang, S., & Rissman, J. (2017). Episodic Memory Retrieval Benefits from a Less Modular Brain Network Organization. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(13), 3523–3531. <https://doi.org/10.1523/JNEUROSCI.2509-16.2017>
- Wiggs, C. L., Weisberg, J., & Martin, A. (1999). Neural correlates of semantic and episodic memory retrieval. *Neuropsychologia*, *37*(1), 103–118. [https://doi.org/10.1016/s0028-3932\(98\)00044-x](https://doi.org/10.1016/s0028-3932(98)00044-x)
- Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neuropsychologia Neural correlates of

- retrieval-based memory enhancement : An fMRI study of the testing effect.
Neuropsychologia, 51(12), 2360–2370.
<https://doi.org/10.1016/j.neuropsychologia.2013.04.004>
- Wing, E., Ritchey, M., & Cabeza, R. (2015). Reinstatement of Individual Past Events Revealed by the Similarity of Distributed Activation Patterns during Encoding and Retrieval. *Journal of Cognitive Neuroscience*, 26(3), 194–198.
<https://doi.org/10.1162/jocn>
- Woldorff, M. G., Liotti, M., Seabolt, M., Busse, L., Lancaster, J. L., & Fox, P. T. (2002). The temporal dynamics of the effects in occipital cortex of visual-spatial selective attention. *Cognitive Brain Research*, 15(1), 1–15.
- Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *J Neurosci*, 20(6), RC63. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10704517>
- Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited--again. *NeuroImage*, 2(3), 173–181. <https://doi.org/10.1006/nimg.1995.1023>
- Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How We Transmit Memories to Other Brains: Constructing Shared Neural Representations Via Communication. *Cerebral Cortex*, 27(10), 4988–5000.
<https://doi.org/10.1093/cercor/bhx202>
- Zuo, X.-N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F. X., Sporns, O., & Milham, M. P. (2011). Network Centrality in the Human Functional Connectome. *Cerebral Cortex*, 22(8), 1862–1875. <https://doi.org/10.1093/cercor/bhr269>

Biography

Benjamin Ryan Geib attended Bucknell University for college and graduated magna cum laude in 2011 with a Bachelor of Science in Biomedical Engineering.

Following graduation he worked for two years as Research Associate at the University California Davis under Dr. Cameron Carter. Following this, he matriculated into Duke University's Cognitive Neuroscience Admitting Program in 2013.

Benjamin has published two first author publications and co-authored numerous other scientific papers and chapters. His first author publications include: 1) Hippocampal contributions to the large-scale episodic memory network predict vivid visual memories in *Cerebral Cortex* in 2015 and 2) From hippocampus to whole brain: The role of integrative processing in episodic memory retrieval in *Human Brain Mapping* in 2017.

Benjamin was awarded a National Research Service Award (NRSA) for 2018-2020 titled *Spatiotemporal Dynamics of Memory Reactivation*.