Title

Decomposing complex cognitive processes to understand individual differences in behavior

Author

Walsh, Catherine

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Decomposing complex cognitive processes to understand individual differences in behavior

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy in Psychology

by

Catherine Ruth Elizabeth Walsh

2023

ABSTRACT OF THE DISSERTATION

Decomposing complex cognitive processes to understand individual differences in behavior

by

Catherine Ruth Elizabeth Walsh

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2023

Professor Jesse A. Rissman, Chair

Cognitive psychologists and neuroscientists traditionally take advantage of tightly controlled experimental studies to make claims about how the mind and the brain work – we design tasks and conditions to isolate a single component and use the results to make conclusions. However, it is becoming increasingly clear that as we move to understand more complex, naturalistic cognition, we need to consider the *integration* of cognitive processes. This dissertation probes how the integration of processes involved in human memory can help us better understand behavior. The first chapter begins by providing an overview of three domains of memory that are subserved by multiple components – declarative memory, memory for faces, and working memory. Chapter 2 investigates the bidirectional interactions of episodic and semantic memory by employing a paired associate learning task with semantically related and unrelated word pairs that additionally manipulates whether the word pairs are learned through either active retrieval

practice or passive restudying. This paradigm, along with a novel extension of a multi-arrangement task (Kriegeskorte & Mur, 2012) to index semantic space before and after learning, shows that prior knowledge asymmetrically reshapes semantic space to make cue words more predictive of target words and testing reduces potential interference by repelling moderately related lure pairs away from the to-be-learned pair in semantic space. Chapter 3 shifts to focus on face memory. In this chapter, we use a large battery of behavioral tasks to identify four latent cognitive processes – face perception, episodic long-term memory, general intelligence and working memory. We show that these factors linearly combine to predict performance on two measures of face memory, one of which we designed to be a more ecologically valid index of personal identity memory. We also identify two cognitive profiles that differentially use working memory and face perception to accomplish our face memory task, highlighting how individual differences in cognitive ability may impact how processes interact to support behavior. In Chapter 4, we sought to explain individual differences in working memory performance, working memory capacity (WMC), and psychiatric outcomes using a wide range of structural and functional MRI measures. Our results highlight how working memory performance and WMC may be predicted by different aspects of functional and structural MRI that reflect distinct underlying processes. Taken together, the findings presented in this dissertation underscore the complexity of cognition. Expanding how we study the mind and brain to take multiple processes of cognition into account will allow us to better characterize and understand complex, real-life behavior.

The dissertation of Catherine Ruth Elizabeth Walsh is approved.

David Clewett

Barbara Knowlton

Michal Ramot

Jesse A. Rissman, Chair

University of California, Los Angeles

2023

DEDICATION

This dissertation is dedicated to those who are guided by curiosity and compassion.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Parts of this research were supported by the National Science Foundation Graduate Research Fellowship Program (DGE - 2034835), the UCLA/Weizmann Collaboration in Neuroscience, and NIMH grant (NIMH R01-MH101478).

I am incredibly grateful to have had so much support through the course of my Ph.D at UCLA. I would first like to thank my advisor, Dr. Jesse Rissman. His mentorship, deep passion for neuroscience, and unconditional support through my graduate school career has shaped me as a researcher and a scientist. Thank you for letting me explore all the things that I was curious about, even when they seemed like a long shot, and guiding me to be the scientist I am today. Next, I like to thank Dr. Michal Ramot, who I have been lucky enough to work with both before and during graduate school. Your mentorship throughout the years has helped me push through worries about finding my place at UCLA and helped me develop the way I think about approaching problems. I would also like to thank the rest of my dissertation committee: Dr. Dave Clewett and Dr. Barbara Knowlton – your feedback and support over my graduate school career have been instrumental in my time at UCLA. I would like to thank Dr. Jean-Baptiste Pochon for his work on the initial pre-processing and analyses of the RDoC dataset, Dr. Craig Enders for his generosity with his time and insight into how to handle missing data, and the UCLA OARC and Department of Psychology Statistical Consulting Services for their support developing appropriate statistical analyses.

Next, I would like to thank my friends and family. I may have been far from home physically, but that did not mean they were any less supportive of me while I have been at UCLA. Thank you to my family – to my parents for never being more than a phone call away for teaching me to always be curious, to value knowledge, and to know that I can do anything I set

my mind to. Thank you to my siblings and their partners (Landis and Leslie, Myles and Carla, Megan and Travis, Fraser and Janelle, Rob and Jill, Lydia and Jordan) for never failing to make me smile and always listening to my science, even when you don't entirely understand it. Thank you to my loved ones in and out of science: Kristi Boazman, Emily Neer, Sarah Kalinowski, Dr. Savannah Lokey, Nicole Mlynaryk, Cameron Riddell, Jason Crutcher, Meghan Collins, Rachel Smith – you all helped keep me grounded, picked me up when I was down, and help me be a better person. Thank you to the members of the Rissman Memory Lab – Mouslim Cherkaoui, Mary Vitello, Samantha Walters, and Fleming Peck – I'm grateful to have co-workers who have become friends and to have learned from all of you on this wild ride of graduate school. Last but certainly not least, I would like to thank Noah Tomares. I can't put into words how grateful for your love and support – you make the hard days easier and the easy days better. Thank you for everything you do for me and Mabel.

VITA

EDUCATION

2016	**Bachelor of Arts, Neuroscience** Colgate University – Hamilton, NY
2020	**Master of Arts, Psychology** University of California, Los Angeles – Los Angeles, CA
2022	**Candidate in Philosophy, Psychology** University of California, Los Angeles – Los Angeles, CA

ACHIEVEMENTS AND AWARDS

2022	UCLA Brain Research Institute/Semel Institute Graduate Travel Award
2018 – 2023	National Science Foundation Graduate Research Fellowship
2018 – 2022	UCLA Graduate Dean's Scholar Award
2016	Phi Beta Kappa Honor Society
2015	Colgate Summer Undergraduate Research Fellowship
2013	Phi Eta Sigma First-Year Honor Society
2012 – 2016	Colgate University Dean's Award for Academic Excellence Recipient

PUBLICATIONS

1. **Walsh, C. R.** and Rissman, J. (2023). Behavioral representational similarity analysis reveals how episodic learning is influenced by and reshapes semantic memory. *In press at Nature Communications*.
2. Hennessee, J. P., Schorn, J.M., **Walsh, C. R.**, Castel, A.D., Knowlton, B.J. (2023). Age-Related Differences in Directed-Forgetting for Memory of To-Be-Forgotten Items and Task-Irrelevant Incidental Details. *Under revision at Aging, Neuropsychology and Cognition.*
3. Kadlec, J., **Walsh C. R.,** Sadé, U., Amir, A., Rissman, J., Ramot, M. (2023). Putting your tasks on trial: A measure of reliability convergence. *Under review at Communications – Psychology.*
4. Phillips, A., **Walsh, C. R.,** Grayson, K., Penney, C., Husain, F. (2022). Diversifying representations of female scientists on social media: a case study from the Women Doing Science Instagram. *Social Media and Society, 8*(3). https://doi.org/10.1177/20563051221113068
5. Reimann, G.E.., **Walsh, C.,** Csumitta, K.D., McClure, P., Pereira, F., Martin, A., Ramot, M. (2021). Gauging facial feature viewing preference as a stable individual trait in autism spectrum disorder. *Autism Research, 14(8).* https://doi.org/10.1002/aur.2540
6. Ramot, M., **Walsh, C.,** Reimann, G., Martin, A. (2020). Distinct neural mechanisms of social orienting and mentalizing revealed by independent measures of neural and eye movement typicality. *Communications Biology 3*(48). https://doi.org/10.1038/s42003-020-0771-1.
7. Ramot, M. **Walsh, C.,** Martin, A. (2019). Multifaceted integration - memory for faces is subserved by widespread connections between visual, memory, auditory and social

networks. *Journal of Neuroscience, 10(1523).* https://doi.org/10.1523/JNEUROSCI.0217-19.2019

SELECTED POSTERS AND PRESENTATIONS

1. **Walsh, C.,** Kadlec, J., Ramot, M., Rissman, J. Inadequate Reliability When Measuring Cognitive Traits Obscures Investigation of Complex Behavioral Tasks. Annual Meeting of the Society for Neuroscience, San Diego, CA. November, 2022.
2. Kadlec, J., **Walsh, C.,** Rissman, J., Ramot, M. Separation of face memory and face perception in humans. Annual Meeting of the Society for Neuroscience, San Diego, CA. November, 2022.
3. Rissman, J. and **Walsh, C.** How semantic space is sculpted by episodic associative learning: A behavioral representational similarity account. Memory Disorders Research Society, Philadelphia, PA. October, 2022.
4. **Walsh, C.,** and Rissman, J. A behavioral representational similarity approach reveals how pre-existing semantic associations are re-sculpted by episodic paired associate learning. Talk presented at Psychonomic Society 62nd Annual Meeting [held online due to COVID-19]. November 2021.
5. Stonehouse Salinas, W.*, **Walsh C**. and Rissman, J. Fusing fMRI and EEG Data to Investigate Individual Differences in Working Memory Capacity. UCLA Psychology Undergraduate Research Conference. Remote. May 2021.
6. **Walsh, C.** and Rissman J. A Behavioral Representational Similarity Approach for Characterizing the Mechanisms of the Testing Effect. Annual Meeting of the Cognitive Neuroscience Society. Remote. March 2021.
7. **Walsh, C.**, Pochon, J.B., Enriquez, K.D., Truong, H., Lenartowicz, A., Loo, S.K., Sugar, C.A., Bearden, C.E., Bilder, R.M, Rissman, J. Understanding the non-monotonic relationship between working memory capacity and maintenance-related fMRI activity. Talk presented at Bay Area Memory Meeting [held online due to COVID-19]. November 2020.
8. **Walsh, C.**, Pochon, J.B., Enriquez, K.D., Truong, H., Lenartowicz, A., Loo, S.K., Sugar, C.A., Bearden, C.E., Bilder, R.M, Rissman, J. A non-monotonic relationship between working memory capacity and load-related increases in brain activity. Annual Meeting of the Society for Neuroscience, Chicago, IL. October 2019.
9. **Walsh, C.**, Pochon, J.B., Enriquez, K.D., Truong, H., Lenartowicz, A., Loo, S.K., Sugar, C.A., Bearden, C.E., Bilder, R.M, Rissman, J. Characterizing the Relationship Between Working Memory Capacity and Load-Related Increases in fMRI Activity. Cognitive Neuroscience Annual Meeting, San Francisco, CA. March 2019.
10. Ramot, M., **Walsh, C.**, Martin, A. Face memory performance is predicted by the strength of resting state functional connectivity between task-defined face patched and medial temporal lobe structures. Cognitive Neuroscience Society Annual Meeting, Boston, MA. March 2018.

* Indicates undergraduate student mentee

Chapter 1: General Introduction

Cognitive psychologists and neuroscientists traditionally take advantage of tightly controlled experimental studies to make claims about how the mind and the brain work. Much of the literature has relied on the assumption of *pure insertion* – that the combination of two tasks is an additive process, where the addition of one task adds something new but does not fundamentally impact the original processes (Donders, 1868; Sternberg, 1969). This framework has inspired many studies that use *cognitive subtraction*, where one employs at least two conditions that theoretically differ only with the use of a single aspect, to isolate specific cognitive processes. For example, one early neuroimaging study presented subjects with a noun and required them to either read it aloud or generate a related verb (Petersen et al., 1988). This paradigm was intended to isolate the process of semantic retrieval, as the other aspects of the conditions (e.g., reading, visual content, physiological arousal, etc.) were assumed to be constant across conditions.

More recent work has used these assumptions to show complex tasks can be broken down into simpler components. One study used multivariate pattern analyses with functional magnetic resonance imaging (fMRI) to identify four states of processing (encoding, planning, solving ,and responding) during mathematical problem solving (Anderson, Pyke, et al., 2016). Other fMRI work has shown that it is possible to predict when subjects are following "compound rules" (i.e. rules that are composed of multiple stimulus-response contingencies) based solely on classifiers trained on prefrontal cortex activity during the simpler rules, suggesting that the rule information is compositional (Reverberi et al., 2012).

Despite the intrinsic appeal of pure insertion due to its simplicity, it is becoming increasingly evident that these kinds of hypotheses cannot reflect the true nature of cognition. For instance, these additive frameworks do not take into account the potential interactivity of cognitive

processes – that is, that the inclusion of a new cognitive process to a task may have an impact on the other parts of the task (Friston, Price, et al., 1996; Poldrack, 2010). Novel analytical approaches such as hidden semi-Markov models with multivariate pattern analysis have demonstrated that although there are distinct stages of processing during complex tasks like recognition memory (Anderson, Zhang, et al., 2016), the combination of these processes violates the assumption of pure insertion (Zhang et al., 2018). Additionally, experimental work has shown that individuals with low fluid intelligence struggle on complex reasoning tasks but are capable of performing well when the task is segmented into smaller, simpler parts (Duncan et al., 2017), suggesting that the integration of the simpler components that make up a complex task might be an important cognitive process itself.

In order to understand how the brain supports these complex tasks, therefore, it is not enough to study a few brain regions or cognitive processes in isolation. Instead, we must turn towards larger-scale interactions to explain complex behavior. The interactive nature of cognitive processes also begs the question: how does the brain integrate these processes? Can complex tasks be considered a composite of their component functions, or is the whole more than the sum of its parts? Are the combinations linear?

Human memory can provide a perfect case study for these questions. When considering the phenomenological experience of memory, it all might seem to be one experience, with few differences between the acts of recalling feelings, thoughts, events, or facts. Decades of psychological and neuroscientific research, however, has suggested that this is not the case. Memory can be conceptualized as a number of fundamentally different processes supported by different neurological structures (Squire, 2004). Below, I will provide an overview of three sub-

fields of memory where integration across cognitive domains and processes is crucial for successful performance: episodic and semantic memory, memory for faces and working memory.

Interactions of Episodic and Semantic Memory

According to early frameworks of memory, semantic memory, or the memory for facts and information separate from experiences, and episodic memory, memory for experiences and events, are two separate cognitive systems (Squire, 2004); studying them separately has proved useful for our understanding of them. Clever experimental designs have demonstrated that semantic memory is organized hierarchically (Collins & Quillian, 1969; Quillian, 1967) across conceptual and categorical dimensions that capture semantic similarity (Binder et al., 2016; Landauer & Dumais, 1997; Simmons & Estes, 2006). Although the hippocampus is important for the acquisition and retrieval of semantic knowledge (Gabrieli et al., 1988; Kapur & Brooks, 1999; Manns et al., 2003; Scoville & Milner, 1957), other work has suggested that semantic information is also supported by a more distributed processing code across the brain (Binder et al., 2009; Duff et al., 2020; Huth et al., 2012; Just et al., 2010; Martin, 2007; Pereira et al., 2016; Rissman & Wagner, 2012), with the anterior temporal lobe (ATL) acting as a central convergence zone that integrates semantic information (Coutanche & Thompson-Schill, 2015; Díez et al., 2017; Martin et al., 2018; Patterson et al., 2007; Peelen & Caramazza, 2012).

Separate lines of research have garnered information on the function and neuroanatomical correlates of episodic memory. Episodic memory can be considered the ability to recall the 'what,' 'when,' and 'where' of an event (Clayton & Dickinson, 1998). Tulving & Markowitsch (1998) further suggest that episodic memory necessitates autonoetic consciousness, or the ability to place oneself in one's own memories. Behavioral studies of patients with damage to the medial temporal

lobe (Scoville & Milner, 1957) and of healthy patients using fMRI (Eldridge et al., 2000; Greicius et al., 2003; Hayes et al., 2004) have linked episodic encoding and recall to the hippocampus. Multivariate analyses have separately showed that spatial patterns of encoding in the hippocampus are related to subsequent memory (Chadwick et al., 2010; Larocque et al., 2013). Other lines of research have shown that as memories are consolidated in the hippocampus, they are stored in the neocortex (McClelland et al., 1995; O'Reilly et al., 2014; O'Reilly & Norman, 2002). Specifically, episodic information has been shown to be encoded in category specific cortex (Polyn et al., 2005) and in parietal cortex (Favila et al., 2018; Jonker et al., 2018; Sestieri et al., 2017).

While studying these systems independently has certainly provided much insight into each system in isolation, there is also a body of work that suggests that there might be less of a distinction between the episodic and semantic memory systems as once was thought. Patients with semantic dementia have been shown to have relatively intact episodic memory (Graham et al., 2000; Hornberger & Piguet, 2012; Irish et al., 2016) and can use this episodic memory system to support some semantic knowledge, particularly when it was in an episodically relevant context (Graham et al., 1999; Snowden et al., 1996). Semantic information can also support better episodic recall in healthy older adults (Badham et al., 2012; Dennis et al., 2007; Jarjat et al., 2020; Loaiza & Srokova, 2020) and younger adults (Antony et al., 2017; Carpenter, 2011; Davelaar et al., 2006; Payne et al., 2012; Poirier & Saint-Aubin, 1995).

Neuroimaging work has highlighted that episodic and semantic memory tasks have both shared and unique aspects (Burianova et al., 2010; Burianova & Grady, 2007), which may be driven by flexible search and retrieval of linked information by the hippocampus (Ryan et al., 2008). Multiple trace theory suggests that episodic and semantic memory tasks are not entirely separable, with semantic memory showing some "episodic residue," such as the conditions the fact

was learned in, that results in hippocampal activation in both episodic and semantic tasks (Moscovitch et al., 2005). The hippocampus' role in pattern completion in episodic memory has been well established (Dimsdale-Zucker et al., 2018; McClelland et al., 1995); some have also theorized that a similar process is involved in semantic search, where concepts are represented by features, and when a subset are activated, the hippocampus uses a pattern completion mechanism to activate other semantic features and complete the retrieval (Solomon & Schapiro, 2020).

One perspective addressing this evidence for and against overlapping memory systems suggests that the episodic and semantic systems are not necessarily separate, but instead show gradients of activation depending on context and task demands. Irish and Vatansever (2020) proposed that both memory systems are anchored in the default mode network and noted that even the hippocampus and anterior temporal lobe, sometimes considered to be linked exclusively to episodic and semantic memory respectively, show gradients of function. In this framework, the overlap arises from the fundamentally shared processes necessary for each task: episodes are comprised of both general conceptual reinstatement and episode specific sensory processing, while recall of semantic memory often includes episodic information about when the information was acquired (Renoult et al., 2019).

While the evidence reviewed here has shown that interactions between the semantic and episodic memory systems exist, a larger question is *how* they interact to facilitate memory. One explanation highlights how semantically related information can bind and integrate information into a broader context that may help support later recall. Prior knowledge can modulate integration into memory; information that was congruent with prior knowledge showed more integration, as measured by more activation of the B stimulus during AC recall in an AB/AC design experiment (van Kesteren et al., 2020). Moreover, prior knowledge of a cue has been shown to cause

asymmetrical representational change in the left inferior frontal gyrus, such that the target representation becomes more similar to the cue, which shows very little representational change, suggesting the integration of new information into an already existing knowledge system (Bein et al., 2020). This integrated knowledge across events in memory networks can then be reinstated to support recall and extraction of novel information (Baldassano et al., 2017; Schlichting & Preston, 2015). Semantic information can facilitate this kind of integration; semantically congruent context information is recalled at a higher rate than semantically incongruent context information, suggesting that the context gets "appended" to the target to make the representation more elaborate and create additional retrieval cues to support memory (Bein et al., 2015).

One additional benefit of integration of semantically related materials is that it can set the stage for pattern separation and hippocampal differentiation. Computational modeling has suggested that strengthening and differentiation of overlapping neural representations shows a non-monotonic function to reduce competition between memories. In this framework, representations that are strongly co-activated get strengthened, representations that are moderately co-activated get weakened and representations that are weakly co-activated have relatively little change, resulting in both a strengthening of target representations paired with a reduction of noise from interfering representations (Ritvo et al., 2019). While some have argued that there is no evidence for this kind of pattern separation in the hippocampus (Quian Quiroga, 2020), other work has robustly shown the differentiation of overlapping representations in the hippocampus reducing interference and increasing later memory performance (Ballard et al., 2019; Chanales et al., 2017; Favila et al., 2016; Hulbert & Norman, 2015; Karlsson Wirebring et al., 2015; Kim et al., 2017; Larocque et al., 2013; Schapiro et al., 2012; Wing et al., 2020). Crucially, Koolschijn and colleagues (2019) propose that hippocampal pattern separation uses relational information, paired

with neo-cortical inhibition to prevent co-activation between similar memories. Unitization facilitated by semantic relatedness could increase the co-activation of related information associated with a goal. This would increase the strength of the memory while simultaneously making it more unique relative to competing pairs and decrease the strength of competing memories to support stronger, more specific memory for semantically related episodic content.

Memory for Faces and Personal Identity Information

Neuroimaging work has been particularly fruitful for understanding how the brain supports face memory. Early work has identified regions in the occipital and temporal lobes as crucial for the visual processing of faces, specifically the fusiform face area (FFA) (Clark et al., 1998; Kanwisher & Yovel, 2006). Together with regions in the posterior superior temporal sulcus (pSTS), the fusiform gyrus and other occipitotemporal regions are considered the "core" face processing network (Gobbini & Haxby, 2007; Haxby et al., 2002). Stronger selectivity for faces and functional connectivity within these core face processing regions have been shown to be associated with better face recognition (Elbich & Scherf, 2017; Furl et al., 2011; Turk-Browne et al., 2010; Q. Zhu et al., 2018). Other downstream non-visual regions, such as the intraparietal sulcus, auditory cortex, anterior temporal lobe (ATL) and the amygdala have also been identified as the "extended" face processing system, which has been associated with non-visual aspects of face memory like person knowledge and emotion processing (Gobbini & Haxby, 2007; Haxby et al., 2002).

Adaptation studies in fMRI have revealed how different parts of the core and extended face processing systems respond to sub-processes involved in face recognition. For instance, early work has shown that the FFA is not sensitive to changes in size of image but is selective to viewpoint

(i.e. different angles of the same face) (Andrews & Ewbank, 2004; Pourtois et al., 2005a; Rotshtein et al., 2005; Xu & Biederman, 2010). In contrast, downstream regions like the pSTS and inferior frontal cortex are not viewpoint selective and can integrate different views of the same face into a single identity (Andrews & Ewbank, 2004; Natu et al., 2010; Pourtois et al., 2005a, 2005b). Regions in the anterior temporal cortex have also been shown to hold view-invariant representations of identities (Kriegeskorte et al., 2007; Natu et al., 2010; Nestor et al., 2011; Sugiura et al., 2006), even in an acquired prosopagnosic who had lesions of the FFA and the occipital face area (OFA) (Yang et al., 2016).

Another critical aspect of person recognition is the ability to recall personal semantic information about individuals. It has been proposed that the bilateral anterior temporal lobe is critical for the storage of personal semantics. The ATL has been shown to be activated when retrieving semantic information about a face (Tsukiura et al., 2006, 2008, 2010). Other work has suggested that the ATL not only stores amodal personal semantic information, but also flexibly coordinates retrieval and integrates the information with face recognition (Deng et al., 2016; Volfart et al., 2020; Y. Wang et al., 2017). Some work has shown that the ATL sends information backwards and influences the processing of modality specific information in a top-down manner (Perrodin et al., 2015).

It has been proposed that the recognition of individuals is subserved by multiple distinct processes: first, each individual has an abstract "personal identity node," which in turn activates identity specific semantic information nodes, face recognition units and name codes (Bruce & Young, 1986; Burton et al., 1999). These models have been extended to account for parallel processing of expression separately from the process of identification (Gobbini & Haxby, 2007). Despite early work suggesting that these are separate processes that happen in parallel (Bruce &

Young, 1986; Gobbini & Haxby, 2007), there is emerging literature that suggests the systems may interact.

Behavioral work has demonstrated that faces that share similar conceptual identity characteristics are rated as more perceptually similar (Oh et al., 2021), and learning conceptual information about a person benefits recognition performance, even compared to seeing multiple views of the same face (Yovel & Schwartz, 2016). Schwartz and Yovel (2019) also demonstrate that encoding a face conceptually (i.e. encoding by asking about personality traits associated with the face) improves recognition over encoding a face perceptually (i.e. encoding by asking about the shape of the face). Neuroimaging work has shown that while there are three relatively independent sub-networks within face-selective regions corresponding to face identification, retrieval of semantic information and analysis of facial expression, there is still considerable overlap between the networks (Zhen et al., 2013).

While there has been shown to be a specific *f* factor that underlies some variance in memory for faces, other domains and cognitive processes also explain a significant amount of variance in performance (Gauthier, 2018; McCaffery et al., 2018; Verhallen et al., 2017). Face perception (i.e. the proficiency for distinguishing perceptually similar faces) has been shown to be separate from the memory for faces (Hacker & Biederman, 2020; Landi et al., 2021). Additionally, recent work has highlighted how face recognition in the "real-world" may be impacted by decision making processes (Devue et al., 2019) and that domain-general holistic processing is associated with individual differences in face memory (DeGutis et al., 2013; Richler et al., 2012; Wang et al., 2012).

Consistent with these findings, neuroimaging work has also been crucial for understanding the processes involved in face memory. Ramot and colleagues (2019) have shown that memory

for faces can be predicted by correlated fluctuations in fMRI activity at rest between regions in the core 'face processing network' and regions involved in vision, memory and social processing, rather than just within the face processing network. Other work has demonstrated that the connectivity between anterior temporal cortex and the core face processing network has been associated with stronger face memory performance (Levakov et al., 2023).

Individual Differences in Working Memory Capacity (WMC)

Many of the complex actions and behaviors that make up our lives require the maintenance, manipulation and updating of information over time. There has been a longstanding theoretical debate about which specific processes underlie working memory. One theoretical framework characterizes working memory as executive attention, which includes the inhibition of irrelevant information, shifting focus as task demands change, and updating maintained content (Engle, 2002). Other work has suggested that working memory can be defined by interactions between Primary Memory (the limited, short-term maintenance of information for active use) and Secondary Memory (more long-term, stable and unlimited storage). In this framework, working memory performance consists of the ability to hold content in Primary Memory paired with the effective encoding and retrieval of task-relevant information from Secondary Memory (Unsworth & Engle, 2007a). A third hypothesis suggests that working memory is primarily characterized by the ability to build, maintain, and update arbitrary bindings (Oberauer et al., 2007; Wilhelm et al., 2013).

Neuroimaging work has successfully identified regions involved in working memory that are modulated by task load (i.e. task difficulty) by systematically increasing the number of to-be-remembered stimuli (Li et al., 2022; Manoach et al., 1997; Mayer et al., 2007; McNab &

Klingberg, 2008; Todd & Marois, 2005). These load effects are primarily found in the dorsolateral prefrontal cortex, where activity increases monotonically as load demands increase (Linden et al., 2003; Manoach et al., 1997; Rypma & D'Esposito, 1999) and is sustained over the period of maintenance, even when the to-be-learned stimulus is no longer visually present (D'Esposito & Postle, 2012; Li et al., 2022; Riley & Constantinidis, 2016; Sreenivasan et al., 2014; Sreenivasan & D'Esposito, 2019). Additional evidence for load effects has also been shown across the posterior parietal cortex, where the intraparietal sulcus (IPS) shows increased activity with increasing memory loads (Manoach et al., 1997; Todd & Marois, 2004, 2005).

While it is widely accepted that there are capacity limits in working memory (Baddeley & Hitch, 1974; Engle, 2002; Unsworth & Engle, 2007a), there has been some debate about how many items a person can hold in their working memory. Miller (1956) initially proposed that we could hold a "magic number" of seven items, plus or minus two items, though more recent work has suggested that we are only able to hold two to five chunks of information, rather than individual items, in our working memory (Cowan, 2000, 2010; Daneman & Carpenter, 2004; J. Ryan, 1969). Other work has suggested that working memory capacity (WMC) is not limited by the number of items to be held, but rather the quality or precision of the memory that is the limiting factor, where representations in working memory could be more precise when there are fewer items to be held (Bays et al., 2011; Ma et al., 2014).

Regardless of how we conceptualize WMC and its limits, it is clear there are stable individual differences in the number of items that any given person can hold (Engle et al., 1999; Stevens et al., 2012). These individual differences have been linked to individual differences in higher order cognitive function. WMC has been shown to be related to intelligence (Chuderski et al., 2012; Johnson et al., 2013; Unsworth et al., 2014, 2015), reasoning ability (Kyllonen &

Christal, 1990) and procedural and declarative learning (Kyllonen & Christal, 1990). In addition to its relation to cognitive function, deficits in WMC have been correlated to dysfunction in psychiatric conditions such as schizophrenia (Perlstein et al., 2001), anxiety (Lapointe et al., 2013; Moriya & Sugiura, 2012), depression (Berman et al., 2011), bipolar disorder (Thompson et al., 2006), substance abuse (Grenard et al., 2008), and neurodevelopmental disorders, such as autism and attention-deficit/hyperactivity disorder (Gathercole & Alloway, 2006). One pillar of the National Institute of Health's Research Domain Criteria (RDoC) project is to formally characterize this relationship and relate neurocognitive markers of working memory to psychopathology across broad diagnostic categories (Bilder et al., 2013).

Neuroimaging studies have identified several candidate brain measures that may be implicated in individual differences in working memory performance and capacity. Univariate load related activity in the prefrontal cortex (Assem et al., 2020; Burgess et al., 2011; Kondo et al., 2004; M. Osaka et al., 2003; Rypma et al., 2002; Rypma & D'Esposito, 1999) and posterior parietal cortex (Todd & Marois, 2005) have been correlated with working memory performance and capacity, as has the maintenance of task-relevant information, indexed by multivariate pattern analysis (MVPA), in visual, parietal, and frontal cortex (Bettencourt & Xu, 2016; Hallenbeck et al., 2021). Other lines of work have shown that the functional connectivity of the fronto-parietal control network with the default mode (Avery et al., 2019; Keller et al., 2015; Murphy et al., 2020; J. Zhu et al., 2021) and saliency networks (Fang et al., 2016) in addition to general measures of the functional integration of network structure at rest (Alavash et al., 2015; Cohen & D'Esposito, 2016; Stevens et al., 2012), is related to working memory performance.

Overview of Dissertation

In this dissertation, I describe a series of studies I conducted to investigate the integration processes across the three domains described above. In Chapter 2, I report results from a paired associate learning task that manipulated the semantic relatedness of paired associates and whether the words were learned through testing or restudying. I supplement traditional measures of behavioral performance with a novel behavioral representational similarity approach to investigate bidirectional interactions of episodic experiences and semantic knowledge and test hypotheses about the mechanism of the testing effect from neurobiologically inspired frameworks such as the Non-Monotonic Plasticity Hypothesis (Ritvo et al., 2019). In Chapter 3, I present an experiment focusing on memory for faces and personal identity information, where a large battery of behavioral tasks was used to index latent cognitive processes. I use these cognitive factors to performance on the Cambridge Face Memory Task (CFMT) (Duchaine & Nakayama, 2006a), a canonical measure of face memory, and the Personal Identity Memory (PIM) task, which we developed to be a more ecologically valid measure of personal identity that includes face recognition in novel contexts and the recall of associated personal semantic information. In this chapter, I include data-driven clustering analyses that identify distinct cognitive profiles that accomplish the tasks using different strategies. In Chapter 4, I present a study that uses a wide range of neuroimaging measures, including structural MRI, resting state fMRI and multiple measures derived from task fMRI, to predict individual differences in performance on a working memory task, WMC and individual differences in psychiatric function.

Chapter 2: Behavioral representational similarity analysis reveals how episodic learning is influenced by and reshapes semantic memory

Introduction

Despite early theories that proposed a psychological and neurobiological separation between semantic and episodic memory systems (Sherry & Schacter, 1987; Squire, 2004), there is an increasing body of work that suggests the two systems are more intertwined than previously believed (Irish & Vatansever, 2020; Renoult et al., 2019). Neuroimaging experiments have demonstrated shared neural activation (Burianova & Grady, 2007) and functional connectivity (Burianova et al., 2010; Rajah & McIntosh, 2005) during episodic and semantic memory processes, and pre-existing semantic knowledge can act as a scaffold to facilitate the acquisition of new episodic memories (Audrain & McAndrews, 2022; Baldassano et al., 2018; Liu et al., 2019). Moreover, semantic relatedness has been shown either facilitate (Antony et al., 2022; Bulevich et al., 2016; Liu & Ranganath, 2021; Payne et al., 2012; van Kesteren et al., 2020; Wing et al., 2022) or impair (Antony & Bennion, 2022; Craig et al., 2013) episodic memory performance, depending on factors such as recall delay, degree of relatedness within the to-be-learned pairs, and the semantic relatedness of the broader stimulus set (Antony et al., 2022). Episodic experiences can also influence semantic knowledge by integrating new information as learning occurs, or by emphasizing task or context-relevant semantic features in pre-existing semantic space (Connell & Lynott, 2014; Solomon & Thompson-Schill, 2017; Yee & Thompson-Schill, 2016). However, further specification of the mechanisms of these putative bidirectional episodic/semantic interactions is needed.

One common assessment of episodic memory involves presenting pairs of items and later probing retention of the associations. Although one-shot learning of paired associates is possible, many paradigms have participants with re-engage with the material through retrieval practice or restudying, and there is a well-established benefit of the former, known as the testing effect (Carpenter et al., 2009; Carpenter & Kelly, 2012; Delaney et al., 2010; Karpicke & Roediger, 2008; Kornell & Vaughn, 2016; Nungester & Duchastel, 1982; Rowland, 2014). There is debate as to whether the "desirable difficulty" (Bjork & Bjork, 2011; Kornell et al., 2011) or effortfulness (Pyc & Rawson, 2010) of searching for and retrieving a target association is what strengthens memory or whether testing is advantageous because the episodic experience of retrieval practice is more contextually similar to the final test (Morris et al., 1977).

While researchers have increasingly acknowledged the interdependence of episodic and semantic memory, there are relatively few studies of the testing effect that directly manipulate the semantic information within to-be-learned pairs of items (Rowland, 2014) or integrate its role into mechanistic accounts. Carpenter (2009) proposed that retrieving information from memory necessitates elaborative processes that induce spreading activation to semantically related information (Anderson, 1983; Collins & Loftus, 1975), which can provide additional retrieval cues (Pyc & Rawson, 2010). Consistent with this framework, one recent study showed that when to-be-learned images do not contain meaningful semantic information, there is no benefit for retrieval practice compared to restudying the images (Ferreira & Wimber, 2021). A separate account suggests that that testing supports memory by facilitating semanticisation (i.e. a shift towards more generic semantic representations as opposed to detail-rich episodic representations) (Lifanov et al., 2021) and relational processing, which promotes attention to semantic information (Rawson & Zamary, 2019).

The degree to which to-be-learned items have a pre-existing semantic relationship may influence how they are associated in memory. The episodic binding of two items need not be symmetrical, in the sense that the ability of item A to predict item B does not necessarily equate with the ability of item B to predict item A. For instance, when pairs of words are learned in one direction (cue word A→target word B), the act of testing an unrelated pair in the forward direction (A→?) also improves associative memory in the reverse direction (B→?), yet when related pairs are tested in the forward direction (A→?), it does not improve recall of the reverse direction (B→?) (Popov et al., 2019; Vaughn & Rawson, 2014). Recent neuroimaging work has also shown asymmetrical integration of associative pairs (Bein et al., 2020). For example, when novel faces are paired with famous faces, the neural representation of the novel face becomes more similar to the representation of the paired famous face, which itself shows minimal representational change. In contrast, when a novel face is paired with another novel face, the neural representations of the two faces become more similar but change equally.

One neurobiologically-inspired computational modeling account of associative learning known as the non-monotonic plasticity hypothesis (NMPH) attempts to explain the testing effect and account for the role of semantic information through the relative co-activation of to-be-learned items and the associated representational change. This framework proposes that changes in memory strength are driven by the relative activation of items, such that memory for items that are strongly co-activated is strengthened, while items that are moderately co-activated are weakened or differentiated (Ritvo et al., 2019; Sinclair & Barense, 2019). When paired items are restudied and brought to mind together, they are strongly co-activated, and thus strengthened (Detre et al., 2013). When paired items undergo retrieval practice, there is also strong co-activation, but because retrieval is often imprecise, it will also tend to moderately co-activate semantically related

concepts (Anderson, 1983; Collins & Loftus, 1975). According to the NMPH, this moderate activation suppresses memory for the related items and differentiates the target to reduce interference and strengthen memory more than restudying (Antony et al., 2017; Hulbert & Norman, 2015; Rafidi et al., 2018; Ritvo et al., 2019; Sinclair & Barense, 2019; Ye et al., 2020)

In the present preregistered study, we sought to investigate the influence of semantic relatedness on the testing effect and understand how episodic paired associate learning might sculpt pre-existing semantic space. We had participants learn semantically related and unrelated pairs of words via testing or restudying and assessed their memory the next day. Although we were interested in how cued recall accuracy would vary depending on semantic relatedness and learning condition, our primary focus was on whether and how the semantic representations of the words changed over learning. For this, we developed a behavioral representational similarity analysis approach, which we applied to data from a similarity-based word arrangement task that participants performed before and after learning. This allowed us to investigate the bidirectional interaction of episodic learning and semantic knowledge by indexing changes in the associative structure and semantic representation of individual words.

Given the existing computational modeling work and literature on the role of semantic information, we expected to see an overall memory benefit for semantically related pairs. We thus predicted that these already-advantaged pairs would have less to gain from testing than unrelated pairs. We anticipated that tested pairs would undergo more representational change, and that the amount of representational change would be correlated with behavioral performance. Finally, we expected to see asymmetric change in the semantic structure of related pairs, where representations of targets would get drawn towards those of the cues, and symmetric changes for unrelated pairs of words.

Methods

The experimental design and data analysis plan were preregistered prior to data collection on November 19th, 2020 on the Open Science Framework at https://osf.io/5q6th/.

Participants

Participants were recruited via Prolific (https://www.prolific.co/) and through the UCLA SONA Undergraduate Participant Pool. A power analysis (see Supplementary Methods for details) suggested we would need a sample size of at least 73, so we aimed to collect useable data from 80 participants. A total of 262 participants (145 from SONA, 117 from Prolific) completed the first session of the experiment. Of those, 183 returned for the second session within 28 hours of completing the first (88 from SONA, 95 from Prolific). After excluding participants who did not complete both sessions or who otherwise did not meet our strict inclusion criteria (described in the Supplementary Methods), we were left with 29 from SONA and 51 from Prolific. Participants from Prolific received monetary compensation and participants from SONA received course credit. The two samples were not significantly different on any key measures, so the samples were combined for a final N=80 (29 male; age range=18-39, mean age=24.33, SD=5.49). Participants from SONA had all completed at least high school level education; years of education was not collected from participants from Prolific. All participants provided informed consent prior to participating. This research was approved by the IRB of the University of California, Los Angeles.

Additionally, we noted in our pre-registration that we would exclude participants who reported rehearsing word pairs between sessions. Ultimately, we included the 8 participants who reported rehearsing word pairs between sessions, as we did not explicitly instruct participants not to rehearse and our survey question was not specific enough to determine the extent to which they

rehearsed (i.e. it did not distinguish whether they spent hours rehearsing all word pairs, or just happened to spontaneously recall one or two of them).

Material

Stimulus materials included 60 cue-target word pairs. Thirty of these pairs were semantically related and were drawn from the FSU Free Association Norms (Nelson et al., 2004). We restricted words to nouns with no homographs, a concreteness norm greater than 3.5, and deemed by Nelson et al. as appropriate for use in an experiment because they had of an acceptable number of normed associates. In order to reduce the possibility that a participant might simply guess the target word given the cue word, pairs were restricted to have a forward strength of association less than 0.5, meaning that fewer than half of people who saw a given cue word would generate the target word in a free association task. Finally, any pairs of words that together made a compound word or were similar to any English idiom were excluded.

For each of the related pairs, we compiled three measures of pair similarity: (1) forward association strength, (2) cosine similarity from latent semantic analysis (LSA) derived from a corpus of 100k English words (http://www.lingexp.uni-tuebingen.de/z2/LSAspaces/) and the *LSAfun* R package (Günther et al., 2015; Nelson et al., 2004), and (3) word2vec similarity, based off of a model trained on a subset of the Google News dataset, which contains 300-dimension vectors for 3 million words and phrases (https://code.google.com/archive/p/word2vec/). An additional 30 low relatedness pairs were selected to form the remaining 30 unrelated pairs. Target words of these pairs were shuffled until all 30 pairs had word2vec, LSA cosine similarities and (if the pair was normed), cue-to-target association strengths that were lower than the entire list of related word pairs to ensure no overlapping measures.

Procedure Overview

Participation in this experiment took place over two days, with the sessions occurring no more than 28 hours apart (see Figure 2.1 for a schematic of the procedure). On Day 1, participants first performed a multidimensional similarity rating task using a drag-and-drop interface (Figure 2.2A; similar to an approach from work in neuroimaging (Kriegeskorte & Mur, 2012), which used picture stimuli instead of words). Following this similarity-based word arrangement task (hereafter referred to as the SWAT), participants completed a learning task, where they were given two opportunities to initially learn a set of 60 words pairs (30 related; 30 unrelated). We note that in our pre-registration of this experiment we had stated that participants would only have one initial learning opportunity before the test/restudy manipulation; however, pilot data suggested that one learning opportunity was not enough to yield sufficient accuracy on Day 2. Then, participants were given a third opportunity to engage with each pair via either testing or restudying. Last, participants completed a short questionnaire about how distracted they were during the task. Participants received a link to the second part of the experiment the following day; if they did not complete the Day 2 session within 28 hours (i.e. before they had a second night of sleep), they were excluded from all analyses. The Day 2 session (Figure 2.1) began with testing of all word pairs ("final test"), and then participants performed another set of similarity judgements using the SWAT protocol. Testing was performed prior to the SWAT protocol on Day 2 to prevent the possibility that words encountered during the SWAT trials would trigger additional retrieval practice or other rehearsal, which could have influenced final test performance in unpredictable ways.

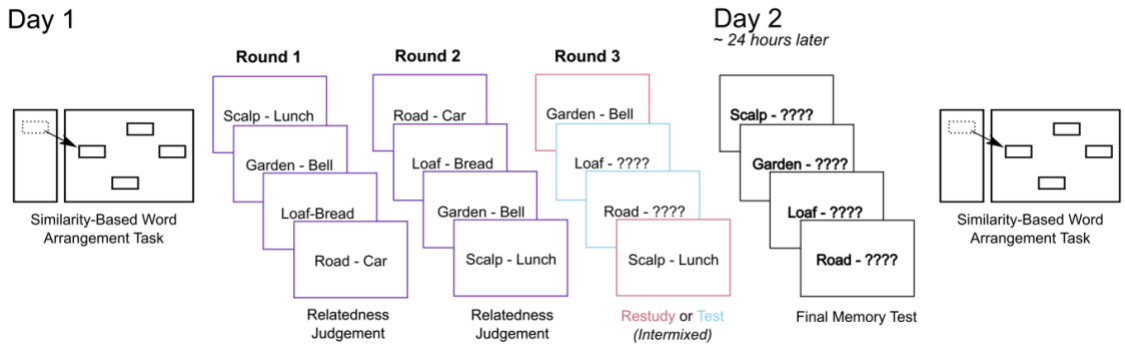


Figure 2.1: Procedural overview. The experiment took place over two days. Prior to learning, participants performed the Similarity-based Word Arrangement Task (SWAT), where they rated the similarity of subsets of words across four trials (60 words per trial). Next, participants had three opportunities to learn 60 pairs of words. During the first two opportunities (Rounds 1 and 2), participants made judgements about the relatedness of the words within the pair. For the third opportunity (Round 3), pairs were either restudied (illustrated here with maroon border) or tested (blue border). On Day 2, participants completed a final cued recall test for all learned pairs, followed by another four trials of the SWAT.

Word Pair Learning

Participants performed three rounds of word pair learning. During each of the first two rounds, all 60 pairs were presented on the screen in randomized order, with the text written in capital letters. Each pair appeared for 4 s with a 2 s ISI. For each pair, the cue word was presented on the left and the target word on the right. During the first round, participants were asked to make a judgement about how related the cue and the target word pairs were on a scale of 1-4, with 1 meaning “not related” and 4 meaning “very related”. The second round was structured the same as the first, but participants were asked to judge how likely it would be for those two words to appear on the same page of a book or magazine on a scale of 1-4, with 1 meaning “not at all likely” and 4 meaning “very likely”. These judgements allowed for incidental encoding and encouraged relational processing of the words in each pair. Relatedness judgements are described in Supplementary Figure 2.3.

In the final learning round, 30 of the pairs underwent retrieval practice (testing) and the other 30 were restudied. Participants were instructed that if they saw the cue and target words

together (just as they had in the prior two rounds) their task was simply to type the target word into the answer box; if they saw the cue word accompanied by four question marks ("????") their task was to attempt to recall the target word and type it into the answer box. If they could not remember the target word, participants were encouraged to take a guess, or they could leave the box blank. Asking participants to type the paired words in the restudy condition, rather than having them make an additional relatedness judgement as in the first two learning rounds, allowed us to match the behavioral response with that of the testing condition (i.e. typing a word). This also served to reduce the differences between behavioral responses in the restudy condition and the final test, where all pairs would be probed by asking the participant to type a word. The learning condition manipulation was randomly interleaved; although this interleaved design necessitates task switching within the learning opportunity block, there was no statistically significant difference in final recall accuracy between trials where the participant switched between testing and restudying and those where learning condition was consistent across consecutive trials (see Supplementary Results for more detail).

Participants were not given a time limit on recalling the second word in the pair. No feedback was provided, as feedback can provide an additional restudying opportunity that can enhance final test performance for tested items (Kang et al., 2007) and inflate testing effects (Rowland, 2014).

The assignment of the word pairs to either the test or restudy condition was counterbalanced by creating two matched sets of pairs with 15 related and 15 unrelated pairs. Words were always presented in the forward order (i.e., cue was always presented before the target). The sets were matched on concreteness, frequency, length of cue and target, word2vec and LSA cosine similarity measures. Each set of words was randomly assigned to either the test or

restudy condition independently for each participant. Memorability of the pairs of words was measured post-hoc by computing the average recall accuracy of the pair across participants; there was a range of accuracy across pairs, ranging from 91% (GENDER- FEMALE) to 5% (CHILDREN – BIRD) of participants recalling any given pair (Supplementary Figure 2.4). Despite the range of memorability across all pairs, there was no statistically significant difference in mean memorability across the two sets of words pairs (see Supplementary Results for more details).

Final Test

In the final test, performed on Day 2, participants were presented with cue words from pairs they had learned on the previous day (with the cue word on the left and "????" on the right, just as in the testing condition on Day 1) and were asked to type in the corresponding target word. There was no time limit on recall, and participants were encouraged to guess if they couldn't remember the pairs or otherwise leave the box blank. Responses were scored as correct if they were spelled correctly or if a spell-checking algorithm (https://textblob.readthedocs.io/en/dev/) identified the correct target word as the most likely word.

Similarity-based Word Arrangement Task (SWAT)

The SWAT (Figure 2.2A) was performed at the beginning of Day 1, prior to learning word pairs, and again at the end of Day 2, after the final test. Each session of the task was comprised of 4 trials. On each trial, participants received 60 words in a "word bank" on the left side of the screen. Participants clicked on a word to bring it over to a main arrangement area ("the canvas") and then dragged each word to the location of their choosing. Participants were instructed to take as long as

they needed to arrange the words such that more similar words were closer together and more dissimilar words were further apart. Trials lasted a median duration of 7.14 minutes.

Individual words were pseudo-randomly assigned to trials based on the to-be-learned pairs. The list of cues and targets were each split in half, to create 4 lists of 30 words. Each list was paired with each other list, except for the list that would form the to-be-learned pairs. This procedure created 4 trials of 60 words each, ensuring that each word would be arranged twice and that the two words in each to-be-learned pair were never both encountered on the same trial. This was an important constraint, as the mere act of thinking about the semantic relationship of the words in the to-be-learned pairs (or learned pairs in the case of the post-learning assessment) during a SWAT trial could bias participants' word placement decisions and corrupt our ability to sensitively measure the behavioral consequences of our experimental manipulations. The order of the 4 trials was randomized for each participant, as was the order of the words in the word bank on each trial.

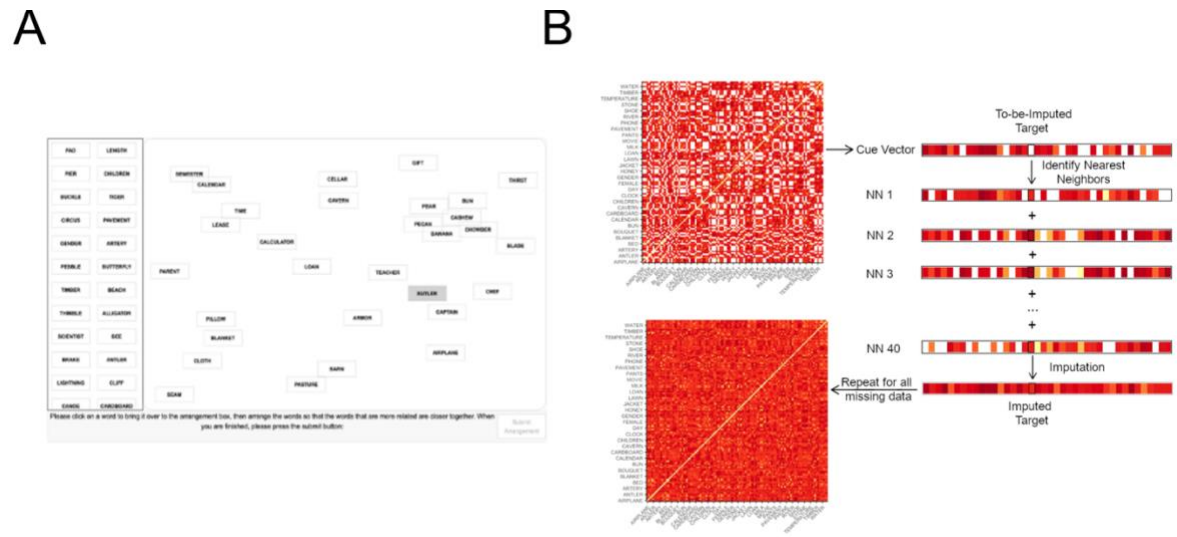


Figure 2.2: Similarity-Based Word Arrangement Task (SWAT) and behavioral representational similarity analysis imputation approach. A. On each SWAT trial, a set of 60 words initially appeared in a random order in a box on the left side of the screen. Words would move to the main canvas when clicked, and once there could be dragged to a chosen location. Participants were instructed to place words that were more similar closer together but not given any rules for how to judge similarity. Participants could move words around until they were satisfied with their final arrangement. The assessment included four SWAT trials, and each word occurred on two of these trials. **B.** Crucially, words from to-be-learned pairs never co-occurred on any SWAT trial to avoid potential contamination of their perceived relatedness, so the similarity of these pairs was imputed (see Methods). Euclidean distance was calculated for each pair of words as a proxy for dissimilarity and later converted to similarity for ease of interpretation.

Derivation of Semantic Similarity Metrics

After participants completed the SWAT arrangements, semantic dissimilarity was calculated for each pair of words by taking the Euclidean distance between the locations of each pair of words on the canvas (measured from the center of each word). Trials were combined using an evidence weighted average of scaled-to-match distance matrices (Kriegeskorte & Mur, 2012). However, because words within to-be-learned pairs were never included on the same trials, we could not directly measure the distance between these words. Thus, by design, our procedure produced an incomplete representational dissimilarity matrix. In order to reconstruct one of our primary measures of interest (i.e. the semantic distance between words in to-be-learned pairs, both before learning and after learning), SWAT trials were combined using an evidence-weighted average and the semantic dissimilarity of unmeasured data pairs was imputed using K-nearest

neighbors imputation using the KNNImputer function (Troyanskaya et al., 2001) from Python's *sci-kit learn* package (Pedregosa et al., 2011) with 40 neighbors (as was determined as an optimal number of nearest neighbors for imputation in simulations) and the "distance" weighting function (see Figure 2.2B for a visualization of this process). This imputation procedure was performed separately on each participant's pre-learning SWAT data and post-learning SWAT data. Since the imputation of not-directly-measured semantic distance ratings is a key innovation of our experimental paradigm, we conducted a number of analyses to confirm the validity of the imputation, and these are described in the Supplementary Methods and Supplementary Figures 2.2 and 2.3. Finally, semantic dissimilarity measures were converted to similarity measures for ease of interpretation by taking *1 – dissimilarity.*

Statistical Analyses

All statistical analyses were conducted in R (version 4.1.2; R Core Team, 2021) and visualized using the *ggplot2* R package (Wickham, 2016). Data and code are available on OSF at https://osf.io/5q6th/.

Preregistered Analyses

To investigate how semantic relatedness influences the testing effect, accuracy for tested and restudied pairs was calculated separately for semantically related and semantically unrelated pairs for each participant in the final session. A 2x2 (relatedness x learning condition) repeated measures ANOVA (RM-ANOVA) using the *rstatix* package (Kassambara, Alboukadel, 2021) was performed to detect differences between conditions on the final test. Furthermore, a single measure of the testing effect on behavioral performance was calculated for each semantic relatedness

condition (related pairs, unrelated pairs) and all pairs (regardless of condition) by taking the difference between the probability of a tested item being correctly recalled and the probability of a restudied item being correctly recalled. Next, tested pairs were split based on whether they were correctly recalled on Day 1. The accuracy on Day 2 was assessed in another 3x2 RM-ANOVA (Day 1 condition (correctly recalled, incorrectly recalled, restudied) x relatedness (related, unrelated). Although we initially preregistered that we would include all trials in the remainder of our analyses, we ultimately opted to exclude pairs that were tested and incorrectly recalled at Day 1 (mean number of pairs excluded=11.23, SD=4.57) because we were primarily interested in the effects of successful testing compared to restudying. Effect sizes for RM-ANOVAs are reported using generalized eta-squared (η_G^2), which measures the effect size with variation from other effects and includes variance due to individual differences (Lakens, 2013).

Change in semantic similarity was calculated for each word pair by taking the difference between similarity on Day 1 and Day 2. With this change measure, a negative value indicates that words within a pair became less similar over time (initial similarity > final similarity), while a positive value indicates that words within a pair became more similar over time (final similarity > initial similarity). As a manipulation check, the Day 2 semantic similarities of learned pairs (i.e. pairs of words that were either tested or restudied in the main part of our experiment), split by those that were correctly recalled at Day 2 and those that were not, were compared to random, unlearned pairs (i.e. all possible pairings of words from our stimulus set that were not restudied or tested during the learning portion of the experiment). This test provides a noise ceiling (as any changes in unlearned pairs can be thought of as noise) and ensures that learned pairs indeed show more representational change than unlearned pairs. Next, values for the learned pairs were entered into a linear mixed-effects model (LMM) with fixed effect predictors of semantic relatedness,

learning condition, and final recall success, and a random intercept of subject identity. We note that although these analyses were initially preregistered to use RM-ANOVAs and paired t-tests, we report our results in an LMM framework to be consistent with our exploratory analyses (see below) and account for variance from potential random effects and report the results from the RM-ANOVA in the Supplementary Results. A testing effect measure for similarity was calculated in a comparable way as we did for the memory recall performance data, by taking the difference between the raw similarity on the final day and the change in similarity across days for tested items and restudied items. These measures of the testing effect from the similarity data were correlated with the testing effect measure for performance in the learning task across all pairs. Results from this analysis are reported in the Supplementary Results.

Additionally, we evaluated asymmetrical representational change of each individual word by extracting the vector of similarity comparing each word to its top 20 nearest neighbors before and after learning. Analyses were restricted to the 20 nearest neighbors to reduce the influence of distant words in semantic space, which would be relatively uninformative for the definition of a given word. For example, it is much more useful to consider the definition of BLANKET in relation to words like PILLOW or CLOTH (the top two closest neighbors in our set, as defined by word2vec), where one can consider the specific connection or compare features, than its relationship to MATH or CHEF (the two least similar words in our set), where they share few features or associates. Nearest neighbors were identified by calculating the cosine similarity between the full semantic feature vectors extracted from word2vec and selecting the top 20 largest similarity values, excluding pairs of words where the distance is imputed. The similarity values for these pairs as measured by the SWAT were used as the vectorized representation for each word. The representation of the cue in the initial pair was then correlated with the target in the final pair

$r(Cue_{Day\,1}, Target_{Day\,2})$ and the cue in the final pair to the target in the initial pair $r(Cue_{Day\,2}, Target_{Day\,1})$. Taking the difference between the Fisher z-transformed correlation values $r\left(Cue_{Day\,1}, Target_{Day\,2}\right) - r(Cue_{Day\,2}, Target_{Day\,1})$ provided a single measure to index the amount of asymmetric change of each of the individual words, where a positive value would indicate that the target word becomes more similar to the cue word, while a negative value would indicate that the cue was drawn more towards the target, and a zero value would indicate that there was equal change for each word in the pair. Asymmetry values were Fisher z-transformed and entered into a linear mixed-effects model (LMM) with learning condition (tested vs restudied), position of word in pair, and relatedness of pair as fixed effect predictors and subject identity as a random intercept. Semantic relatedness and learning condition were iteratively tested as potential random slopes using likelihood ratio tests (using *varCompTest* from the *varTestnlme* R package Baey, C & Kuhn, E, 2019) and the variance of random effects in the final model was estimated using restricted maximum likelihood (REML). Follow up pairwise comparisons were used to investigate significant effects with Holm-Bonferroni corrections for multiple comparisons. Additionally, we tested whether the Fisher z-transformed asymmetry values were significantly different from zero using a series of two-tailed one-sample t-tests with Holm-Bonferroni corrections for multiple comparisons.

We note that we initially pre-registered that we would complete this analysis using all values from the row vector (rather than just the top 20 nearest neighbors). This analysis was initially attempted and resulted in no significant results. However, this analysis assumes that the measured representation of each word in our set is independent from that of the other words in our set; in a neuroimaging-based representational similarity analysis (which our analysis was inspired by), this is indeed the case. However, in our paradigm, the semantic representation of each

individual word is derived from its relationship to every other word in the set, and all of these words also underwent learning. As such, when comparing the representation of a given word across learning to all other words in the set, we are unable to isolate the change of that specific word from the changes in all the other words in the set, thereby inducing additional noise and making it more difficult to see any meaningful change for any individual word.

Finally, we note that we deviated from our pre-registration for both our analyses of the change in similarity and asymmetry values of learned pairs by separating pairs into those that were subsequently recalled at Day 2 and those that were forgotten at Day 2 (and include this distinction as a fixed effect predictor in our models) to allow us to test how our findings related to behavioral performance. Additionally, although we include observations about trials that were tested but incorrectly recalled at Day 1 in our basic behavioral analyses, we opted to exclude those trials from our analyses of representational space as we were primarily interested in the differential effects of our learning conditions (which theoretically only occur when testing is successful; Storm et al., 2014), and we ultimately did not have a sufficient number of trials that were tested and incorrectly retrieved at Day 1 to sufficiently power any analysis of representational change for that trial type.

Exploratory Analyses

As a complement to the preregistered analyses described above, several exploratory analyses were also performed. First, we performed an additional two-tailed paired t-test on the recall accuracy of tested pairs at the initial Day 1 test to determine whether semantically related pairs were recalled better than semantically unrelated pairs. Restudied pairs were excluded from this analysis as the accuracy of these pairs reflected the ability to correctly type the fully visible target word, rather than memory recall performance.

In addition to conducting a 2x2 RM-ANOVA on the similarity measures, we conducted a series of two-tailed one sample t-tests with Holm-Bonferroni adjustments for multiple comparisons to test whether the change in similarity in each condition was different from zero.

To further probe the effects of learning on semantic representations and representational change, we performed a series of LMMs, using the *lmer* function from the *lmerTest* R package (Kuznetsova et al., 2017) to estimate fixed and random effects. For each model, we included predictors of relatedness (related vs unrelated), learning condition (tested vs restudied), and recall success at Day 2 (recalled vs forgotten). Additional predictors were included for some models as necessary. Subject identity was entered as a random intercept for each model (which allows for variance in the intercept over participants), and semantic relatedness, learning condition, and position in pair (when relevant to model) were tested as potential random slopes sequentially using likelihood ratio tests (using *varCompTest* from the *varTestnlme* R package; Baey, C & Kuhn, E, 2019) for each model separately. Although the potential variance in the slopes was not the primary target of these analyses, the inclusion of random slopes allowed us to better explain variance in the model overall. All models were run with a maximum of 200,000 iterations for convergence. Once the final model was determined, significant main effects and interactions were probed using pairwise comparisons (using the *emmeans* R package; Lenth, 2022) with Holm-Bonferroni corrections for multiple comparisons. All models were estimated using REML, and two-tailed t-tests for fixed effects were estimated using Kenward-Roger's method. Effect sizes were estimated using partial eta-squared (η_p^2), as measured from the *effectsize* R package (Ben-Shachar et al., 2020). Unless otherwise noted, this procedure was used for all LMMs. Tables listing all coefficients, standard errors, degrees of freedom and t-values for each model are reported in Supplementary Tables 2.1 – 2.7.

In addition to our preregistered analyses of representational asymmetry described above, which operate on pairwise similarity values of cues and targets before and after learning, we also sought to analyze how each word within a given pair underwent representational change. To test this, we first computed each word's similarity with its top 20 nearest neighbors, and thus derived a 20-value representational vector for each word before and after learning. We used the Fisher z-transformed Pearson correlation between these vectors as a measure of change for each individual word. In addition to the fixed effect predictors of relatedness, learning condition, and recall at Day 2, this model included a fixed effect predictor of the word's position in the to-be-learned pair (cue vs target). This effect of word position was also tested as a potential random effect using likelihood ratio tests, as was done in previous models.

We additionally explored changes in the semantic distance of potentially interfering lure pairs (i.e. words in our set that were semantically related to the cue words of our to-be-learned pairs) to further explore the sculpting of semantic space due to learning. To do so, we calculated the semantic similarity (indexed by the LSA cosine similarity) for all potential 118 pair combinations for a given cue word in our to-be-learned set of words (excluding the associated to-be-learned target word and a word's similarity to itself). Given that these pairs were identified post-hoc after creation of the to-be-learned pairs, there was a wide range of similarity values. We then divided these lures into four classes of lures: weak/non-lures (LSA cosine similarity less than 0.2), moderate lures (LSA cosine similarity between 0.2 and 0.4), strong lures (LSA cosine similarity between 0.4 and 0.6) and very strong lures (LSA cosine similarity above 0.6); Supplementary Figure 2.5. For example, for the to-be-learned pair BLANKET – BED, SEMESTER would act as a weak/non-lure, TEMPERATURE would act as a moderate lure, SEAM might act as a strong lure and PILLOW would act as a very strong lure. Pairs that were not

correctly recalled at Day 2 were excluded from this analysis, as incorrect responses were often other words from our corpus (which would be considered lures in this analysis) and this retrieval may have influenced the similarity judgements in SWAT protocol (which was performed after the final test). Additionally, as in our other analyses, we excluded tested pairs that were incorrectly recalled at Day 1. This selection was repeated for the cues of all to-be-learned pairs separately for each individual, resulting in a range of 1652 – 5900 (mean=3106, SD=981) semantic lure pairs per participant. We used pairwise change in similarity across learning for the semantic lures as the dependent variable for an LMM regression with fixed effects of condition of the associated to-be-learned pair (tested vs restudied), relatedness of the associated to-be-learned pair (related vs unrelated), and strength of the lure pair (weak/non-lure, moderate lure, strong lure, and very strong lure). This model used a BOBYQA optimizer to ensure model convergence. As in our previous analyses, subject identity was included as a random effect in all models and relatedness and learning condition were independently and sequentially tested as potential random effects, as were potential two-way and three-way interactions. Significant main effects and interactions were probed by computing the contrast of the difference of each lure class and the non-lure pairs and comparing across learning condition (for example, the contrast [moderate lure – weak/non-lure for tested pairs] – [moderate lure – weak/non-lure for restudied pairs]) with Holm-Bonferroni corrections for multiple comparisons. Additionally, pairwise comparisons of all lure classes across learning condition were computed with Holm-Bonferroni corrections for multiple comparisons and are reported in the Supplementary Results. A summary of all analyses of representational change (both pre-registered and exploratory) is provided in Figure 2.3.

To supplement our analyses relating representational change and semantic structure to final recall success, we ran a generalized LMM with a logit link function (i.e. a mixed effects logistic

regression) using the *glmer* function from the *lme4* package (Bates et al., 2015). This model was fit using maximum likelihood estimation and a BOBYQA optimizer with a maximum of 200,000 iterations. We included fixed effects of learning condition (tested vs restudied), relatedness (related vs unrelated), Fisher z-transformed correlation of the cue and target across learning, difference of Fisher z-transformed correlation to normative semantic space across learning for both cues and targets, and Fisher z-transformed asymmetry value to predict the probability of final recall success (recalled vs forgotten). As in our previous LMMs, the effect of subject identity was included as a random effect, and random effects of relatedness and learning condition were independently tested as potential random effects using likelihood ratio tests. Significant main effects and interactions were probed using pairwise comparisons with Holm-Bonferroni corrections for multiple comparisons.

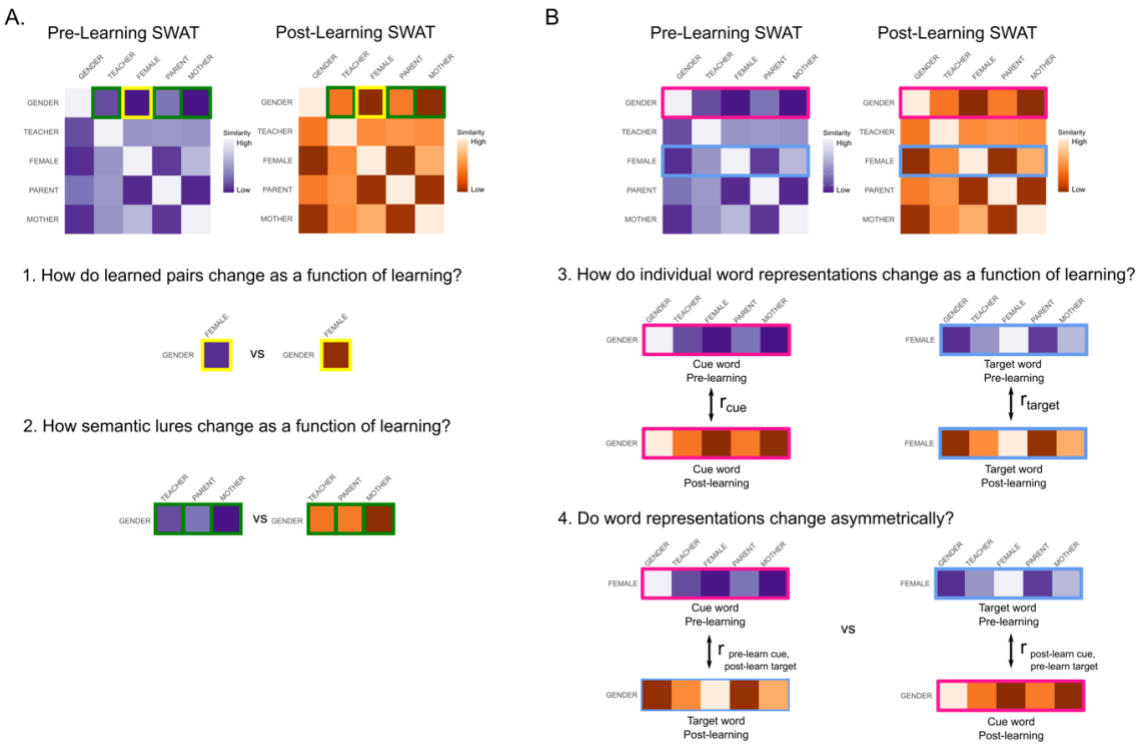


Figure 2.3: Analyses of Representational Change. A. Schematic of representational similarity matrices (RSMs) derived from the Similarity-Based Word Arrangement Task (SWAT) procedure before learning (purple RSM) and after learning (orange RSM); note that our real matrices would be 60x60 words rather than the 5x5 words used in this toy example. Using the pair GENDER-FEMALE for illustrative purposes, we illustrate four of our key analyses. A. Analyses of pairwise representational changes across learning. In Analysis 1, cells outlined in yellow highlight the pairwise distance of the cue word GENDER to its target FEMALE, and we compare how this distance changes across learning. In Analysis 2, we examine the change in pairwise distance across learning between cue words (e.g., GENDER) and semantically related non-target words (lures; green outlines). B. Analyses of individual word representations across learning. Pink outline reflects cue word GENDER, blue outline reflects target word FEMALE. We define the representation of an individual word as its row vector from the RSM (i.e. by its pairwise relationships to all other words in our set). In Analysis 3, we test how the representation of each word changes across learning by taking the Pearson correlation of the row vectors from the pre- and post-learning RSMs. In Analysis 4, we test whether the word representations in the to-be-learned pair change asymmetrically. In this analysis, we correlate the representation of the cue word before learning with that of the target word after learning and the representation of the cue word after learning with that of the target word before learning. The difference between these two values is calculated as a measure of asymmetry, where a positive value reflects the target being drawn towards the cue, a negative value reflects the cue being drawn towards the target and a value of zero reflects the cue and target being drawn towards each other symmetrically (or no representational change).

Data and Code Availability

The behavioral data generated in this study have been deposited in the Open Science Framework database under accession code https://osf.io/5q6th/. All code necessary to reproduce the analyses in this manuscript are provided at the same link.

Results

Recall Accuracy

We began by probing whether recall accuracy for the targets of each cue-target pair systematically varied based on the semantic relatedness (related vs. unrelated) and the learning condition (testing vs restudying, following two initial exposures); Figure 2.4. On Day 1, we could only assess recall accuracy for tested pairs, since performance for restudied pairs merely reflected participants' ability to type the visible target word. As expected, semantically related word pairs were recalled better than semantically unrelated pairs ($t_{(79)}$=9.979, p<0.001, d=1.12); Figure 2.4A. After a 24-hour delay (Day 2), a RM-ANOVA revealed significant main effects of semantic relatedness ($F_{(1,79)}$=362.227, p<0.001, η_G^2=0.462) and learning condition ($F_{(1,79)}$=25.240, p<0.001, η_G^2=0.045) but no interaction ($F_{(1,79)}$=0.076, p=0.78, η_G^2=9.94x10^{-5}); Figure 2.4B. Comparison of marginal means at the final test showed that related pairs (M=0.629, SD=0.183) had a higher probability of recall than unrelated pairs (M=0.281, SD=0.202), and tested pairs (M=0.496, SD=0.258) were more likely to be recalled than restudied pairs (M=0.415, SD=0.256).

Given that participants were provided with no feedback, it is possible that tested pairs that were not successfully retrieved on Day 1 would not benefit from testing, potentially obscuring an interaction between semantic relatedness and learning condition on Day 2. To investigate this, tested pairs were split into those that were correctly recalled at initial learning and those that were not, revealing a significant relatedness by learning condition interaction for Day 2 recall performance ($F_{(2,154)}$=23.531, p<0.001, η_G^2=0.054); Figure 2.4C and 2.4D. Follow up paired t-tests revealed significant testing effects (i.e. the contrast of pairs that were tested and correctly recalled at Day 1 versus pairs that were restudied) for both related pairs ($t_{(79)}$=9.575, p<0.001, d=1.070) and unrelated pairs ($t_{(79)}$=12.361, p<0.001, d=1.382), with a larger effect of learning condition for

unrelated pairs. Pairs that were tested but recalled incorrectly at Day 1 showed significantly lower accuracy on Day 2 than both restudied pairs (related: $t_{(77)}=13.602$, $p<0.001$, $d=1.540$; unrelated: $t_{(79)}=9.692$, $p<0.001$, $d=1.084$) and tested pairs that were recalled correctly at Day 1 (related: $t_{(77)}=22.293$, $p<0.001$, $d=2.524$; unrelated: $t_{(79)}=18.870$, $p<0.001$, $d=2.110$). These results indicate that semantic relatedness reduces the magnitude of the testing effect by improving recall of related restudied pairs. The benefit of semantic relatedness overcomes the relatively ineffective learning method of restudying, leaving less to gain by testing.

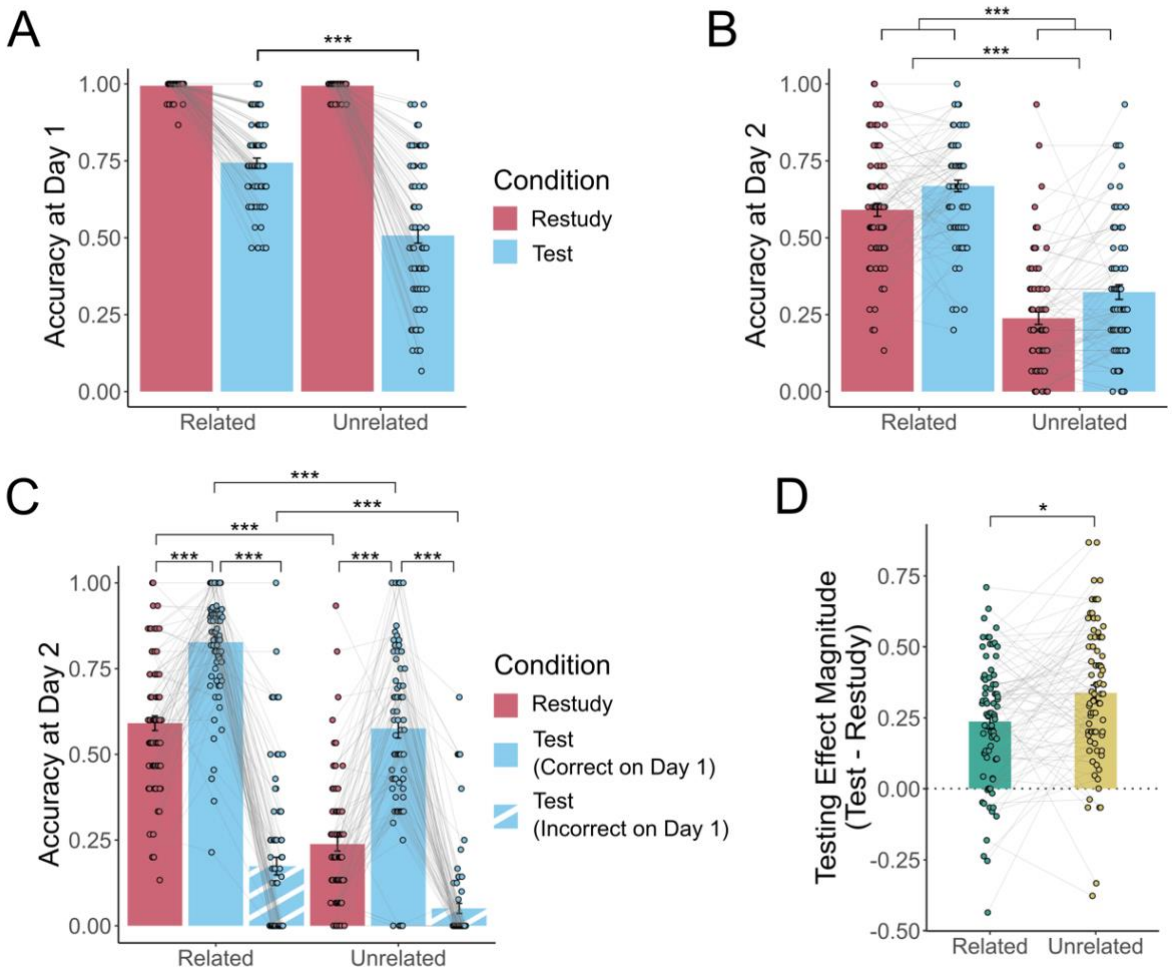


Figure 2.4: Retrieval accuracy across days as a function of semantic relatedness and learning manipulation A. On Day 1, a two-tailed t-test revealed that pairs that underwent testing during Round 3 showed a significant effect of relatedness (blue bars; $t_{(79)} = 9.979$, $p<0.001$, $d=1.12$). Restudied pairs from Day 1 are shown (maroon bars) but were not analyzed since the Day 1 restudy score merely reflects the ability to re-type the target word that was displayed on the screen, rather than memory strength. B. On Day 2 (our critical measure of learning outcomes), a RM-ANOVA revealed that there was better cued recall of target words from related pairs than unrelated pairs ($F_{(1,79)} = 362.23$, $p<0.001$, $\eta_G^2=0.462$), and better recall for tested pairs (blue bars) than restudied pairs (maroon bars) ($F_{(1,79)} = 25.24$, $p<0.001$, $\eta_G^2=0.045$), but no significant interaction. C. Splitting tested word pairs based on whether or not they were successfully recalled on Day 1 reveals a significant semantic relatedness by learning manipulation interaction ($F_{(2,154)} = 23.53$, $p<0.001$, $\eta_G^2=0.045$), showing a larger testing effect (i.e. an advantage for tested pairs that were correctly recalled at Day 1 over restudied pairs) for unrelated pairs than for related pairs. Solid blue bars reflect tested pairs that were recalled correctly at Day 1, striped bars reflect tested pairs that were not recalled correctly at Day 1. (**D**). Across all panels, open circles reflect means of individual participants (N = 80), with connecting lines showing within-subject differences across conditions. Error bars reflect standard error of the mean. Symbols reflect statistically significant differences across conditions using Holm-Bonferroni corrections for multiple comparisons (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

Change in Pairwise Representational Similarity

While differences in recall accuracy based on semantic relatedness and learning condition show that these factors are consequential for memory, they cannot show how this happens. To gain mechanistic insight, we turned to the changes in pairwise representational similarity (measured by the difference between within pair similarity at the final and initial Similarity-Based Word Arrangement Task (SWAT) assessments; Figure 2.2), which provides a more direct measurement of how the semantic representations of our word set change across learning; Figure 2.3.

First, we ran a linear mixed-effects model (LMM) testing whether there were differences in the change in similarity for pairs that were correctly recalled at Day 2 relative to those incorrectly recalled at Day 2 and to a control condition of random word pairings that were never experienced during learning. Pairs that were correctly recalled at Day 2 changed more than those that were incorrectly recalled ($t_{(158)}$=2.566, p=0.0225, d=0.20) and random pairs ($t_{(158)}$=3.112, p=0.0066, d=0.25), but the change in pairs that were incorrectly recalled at Day 2 was not significantly different from that of random pairs ($t_{(158)}$=0.547, p=0.585, d=0.04); Figure 2.5A.

We next ran a series of one-sample t-tests (with Holm-Bonferroni corrections for multiple comparisons) to determine whether change in similarity in our conditions of interest was significantly different from zero; Figure 2.5B. For this analysis (and all hereafter), we opted to exclude tested pairs that were incorrectly recalled at Day 1 because they did not incur the benefit of testing. Significant changes in similarity were observed for related pairs that were correctly recalled at Day 2, regardless of learning condition (tested: $t_{(79)}$=3.788, p=0.002, d=0.423; restudied: $t_{(79)}$=4.258, p<0.001, d=0.476), and unrelated pairs that were tested and correctly recalled at Day 2 ($t_{(74)}$=3.085, p=0.017, d=0.356). All other comparisons were not significantly different from zero (p-values>0.1). When change in similarity across conditions was analyzed in an LMM with fixed

effects of relatedness, learning condition, and final recall success, there were no main effects or interactions between relatedness and learning condition; there was, however, a significant main effect of final recall success ($t_{(528)}$=1.965, p=0.050, η_p^2=0.0073), where pairs that were correctly recalled at Day 2 showed significantly more change in similarity than those that were not.

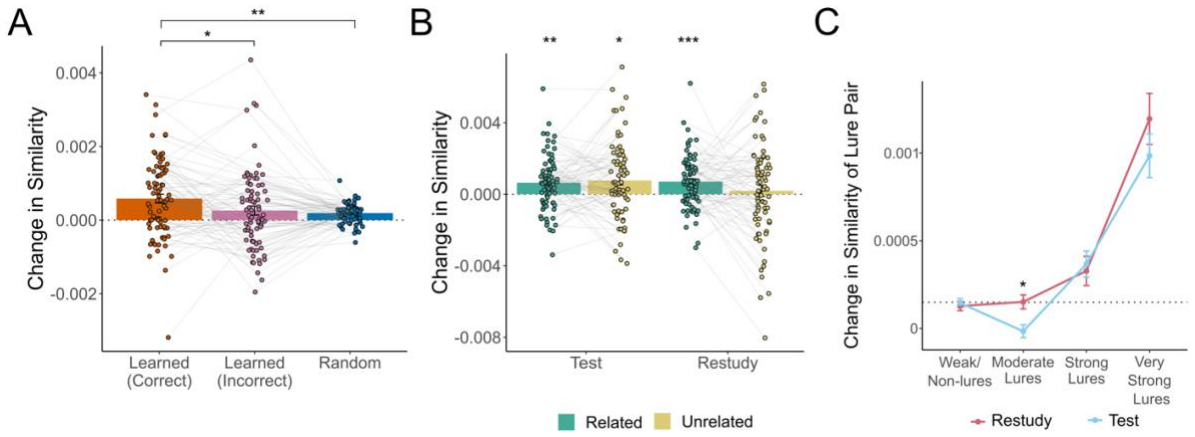


Figure 2.5: Learning-induced changes in representational similarity for paired words. A. Linear mixed effects modeling revealed that pairs of words that were experienced as to-be-learned pairs (regardless of whether tested or restudied) and that were correctly recalled at final test (regardless of learning manipulation; orange bar) became more similar after learning than pairs that were not successfully recalled at final test (pink bar; $t_{(158)}$ = 2.566, p = 0.023, d=0.20) and arbitrary pairings of words that were never experienced as to-be-learned pairs (blue bar; $t_{(158)}$ = 3.112, p = 0.007, d=0.25). Y-axis indicates change in similarity (post-learning assessment minus pre-learning assessment). B. Two-tailed one sample t-tests reveal significant learning-induced similarity change for pairs of words that were tested (regardless of semantic relatedness; related: $t_{(79)}$ = 3.788, p = 0.002, d=0.423; unrelated: $t_{(74)}$ = 3.085, p = 0.017, d=0.356) and for semantically related restudied pairs ($t_{(79)}$ = 4.258, p = 0.0004, d=0.476). Green indicates semantically related pairs, yellow indicates semantically unrelated pairs. For both panels A and B, open circles reflect means of individual participants (N = 80), with connecting lines showing within-subject differences across conditions. C. Words that may interfere with successful recall of to-be-learned pairs (i.e. lures) were defined by their LSA cosine similarity to a given to-be-learned cue. Change in similarity of these potential lure pairs was calculated across learning and entered into a linear mixed effects model based on data from 248,408 pairs of words across 80 participants. This model revealed a significant learning condition by lure strength interaction ($t_{(248400)}$ = 2.840, p = 0.0045, η_p^2=3.25x10^{-6}), where very strong lures are drawn towards a given cue word regardless of learning condition (restudy: z=6.689, p<0.001, d=0.15; test: z=6.029, p<0.001, d=0.12), while strong lures showed no change relative to weak/non-lures (restudy: z=1.707, p=0.32, d=0.022; test: z=1.964, p=0.20, d=0.022). In contrast, moderate lures are pushed away from cue words (relative to baseline weak/non-lures) more when the associated to-be-learned pairs are tested than restudied (z=2.840, p=0.014, d=0.03). Maroon lines reflect lures and symbols associated with restudied pairs, blue lines and symbols reflect lures associated with tested pairs. Dotted line reflects average change in similarity for baseline non-lure pairs. Across all panels, error bars reflect standard error of the mean. Symbols reflect statistically significant differences across conditions (panels A and C) or versus zero (panel B) using Holm-Bonferroni corrections for multiple comparisons (* p < 0.05, ** p < 0.01, *** p < 0.001).

Although correctly recalled pairs showed the most overall representational change, it is also possible that there might be changes within the local semantic neighborhoods of the learned

cue words that reflect the repulsion of potential competitor words (i.e. potential lures) to reduce interference; Figure 2.3A. To test for this, we first characterized the strength of potential lures for each cue word using the LSA cosine similarity between the cue word and all other words in our 120-word set. For example, for the pair GENDER – FEMALE, the word MOTHER might interfere with recall, while CAVERN likely would not. We then calculated the change in similarity across learning for cues in successfully recalled to-be-learned pairs and their potential lures (e.g., GENDER – MOTHER). We used this change in similarity as the outcome variable of an LMM with fixed effects of relatedness, learning condition, and lure strength and random effects of relatedness and learning condition. This model showed a significant learning condition by lure strength interaction ($t_{(248400)}$=2.840, p=0.00451, η_p^2=3.25x10^{-6}); Figure 2.5C. Follow up t-tests revealed that very strong lures are drawn together more than weak/non-lures when they were associated with both tested (z=6.029, p<0.001, d=0.12) and restudied pairs (z=6.689, p<0.001, d=0.15). In contrast, moderate lures associated with tested pairs were pulled together less than tested weak/non-lures (z=4.182, p<0.001, d=0.03). Because there was generalized semantic change even for weak/non-lures (see Supplementary Results), we further probed this interaction by contrasting each lure bin with the weak/non-lures across learning condition. This analysis revealed that moderate lures associated with tested pairs are drawn together less than those associated with restudied pairs (z=2.840, p=0.014, d=0.03). All other baseline-corrected comparisons were not significant (p-values>0.05). Additional pairwise comparisons between other lure bins and other significant effects are reported in Supplementary Table 2.2.

These results show that successful recall not only pulls to-be-learned word pairs closer together in representational space, but also sculpts the overall representational space by drawing highly similar words closer to the cue word to potentially serve as additional retrieval cues for the

to-be-learned target. Testing additionally repels moderate lures that are unlikely to serve as retrieval cues and could potentially interfere with successful recall.

Change in Overall Representational Similarity Structure

A complementary approach to our analyses of the relationship of words within a to-be-learned pair is to investigate how the semantic relationship of each word changes with respect to all other words in the set. To explore this, we extracted from the full representational similarity matrix the row vector reflecting a word's similarity to its 20 nearest semantic neighbors and compared this across learning; Figure 2.3B. For example, the representation of GENDER can be defined by its similarity to its nearest semantic neighbors, including CHILDREN, MOTHER, TEACHER, and PARENT. By comparing the similarity of GENDER to each of these words across learning, we can quantify how much the representation of GENDER changes. When Fisher z-transformed correlation values were entered into an LMM with fixed effects of relatedness, learning condition, final recall success, and word position (cue vs target), and random effects of word position and learning condition, there was a significant relatedness by word position by final recall success interaction ($t_{(967)}$=2.607, p=0.009, η_p^2=0.007), Figure 2.6A, in addition to significant relatedness by final recall success interaction ($t_{(967)}$=1.986, p=0.047, η_p^2=0.0041), and position by final recall success interaction ($t_{(964)}$=2.581, p=0.009, η_p^2=0.0068). Additionally, there was a main effect of final recall success ($t_{(963)}$=2.660, p=0.007, η_p^2=0.0073). Follow up t-tests revealed that for related pairs that were successfully recalled at Day 2, target words underwent more learning-induced representational change than cue words ($t_{(372)}$=3.546, p=0.002, d=0.18).

Comparing the correlation of representations across learning can identify asymmetry of change for paired words but does not provide information about *how* the structure of the pair

changes. For instance, our previous analyses showed that GENDER changes relatively more than its target FEMALE, but it cannot tell us whether GENDER becomes more similar to FEMALE, or whether the changes are unrelated to its to-be-learned target; Figure 2.3B. To investigate this, we calculated a single asymmetry measure by subtracting the correlation of the cue after learning and target before learning from the correlation of the cue before learning and target after learning to determine whether how the representations change relative to each other. Here, a positive value would suggest the representation of the target is drawn towards that of the cue, a negative value would suggest the cue is drawn towards the target, and a value of zero would suggest that the relative representational change of the cue and target is symmetric. An LMM on our asymmetry measure with fixed effects of relatedness, learning condition, and final recall success and a random effect of learning condition showed a significant main effect of relatedness ($t_{(452)}$=2.414, p=0.016, η_p^2=0.01), where the asymmetry value was significantly smaller for related pairs than unrelated pairs; Figure 2.6B. We additionally found that unrelated pairs did not show any significant asymmetry relative to zero ($t_{(78)}$=0.814, p=0.418, d=0.18). In contrast, related pairs showed a numerically negative asymmetry value; however, despite a moderate effect size, this effect was only a non-significant trend after corrections for multiple comparisons ($t_{(79)}$=2.157, p=0.068, d=0.49).

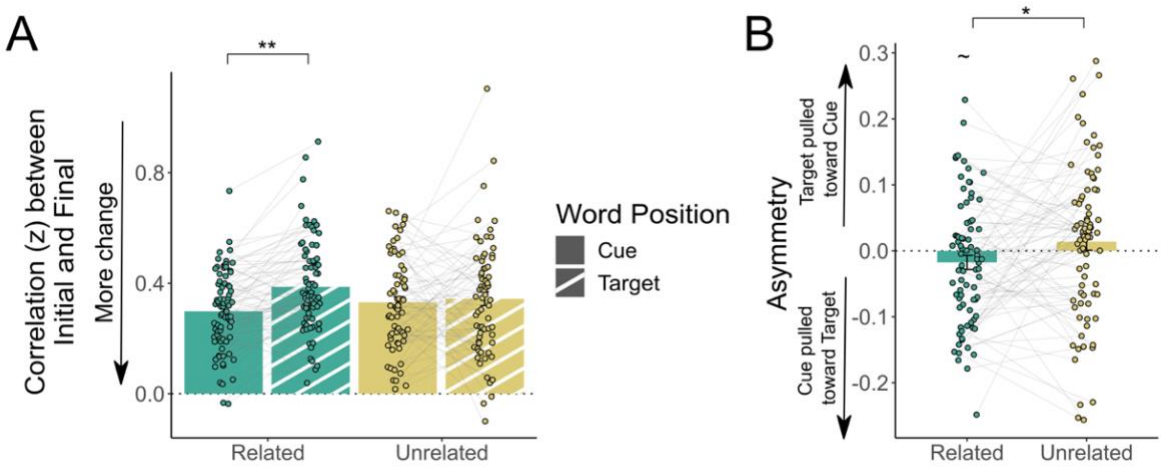


Figure 2.6: Symmetry of representational change within pairs across learning. A. Linear mixed effects modeling revealed that target words in semantically related pairs showed less representational change (i.e. a higher correlation between their initial representation and final representation) across learning than did cue words when correctly recalled at Day 2 ($t_{(372)}$ = 3.546, p =0.002, d=0.18). This effect was not observed when comparing cues and targets in unrelated pairs. Green indicates semantically related pairs, yellow indicates unrelated pairs; solid bars indicate cue words, striped bars indicate target words. B. Related pairs of words show significantly less asymmetrical change in representation than unrelated pairs ($t_{(452)}$ = 2.414, p = 0.016, h_p^2=0.01). All displayed correlation values are Fisher r-to-z transformed. Open circles reflect means of individual participants (N=80), with connecting lines showing within-subject differences across conditions. Error bars reflect standard error of the mean. Symbols reflect statistically significant differences across conditions using Holm-Bonferroni corrections for multiple comparisons (~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Relating Accuracy to Representational Change

To further probe the behavioral relevance of representational change for learning outcomes, we conducted an item analysis (with each word pair considered an 'item'). The average accuracy at final test across participants for each word pair (regardless of learning condition or semantic relatedness) was significantly correlated with its similarity after learning (Figure 2.7A; $r_{(58)}$=0.46, p<0.001) and average change in similarity (Figure 2.7B; $r_{(58)}$=0.39, p=0.002), suggesting that word pairs that are considered more similar after learning and that show greater learning-induced representational change are more likely to be remembered.

Although comparing the average pairwise change in similarity to average accuracy across participants provides a valuable link between the re-sculpting of semantic space and behavioral performance, it overlooks the fact that semantic relatedness and learning condition may have

differential effects on the recall success of a word pair. To investigate the relative contribution of these processes to behavioral performance, we conducted a mixed effects logistic regression predicting the Day 2 recall outcome of each individual word pair. Echoing our previous analyses, this model showed a significant relatedness by learning condition interaction ($z=2.424$, $p=0.014$, $\eta_p^2=0.0016$), in addition to significant main effects of relatedness ($z=10.572$, $p<0.001$, $\eta_p^2=0.028$) and learning condition ($z=6.324$, $p<0.001$, $\eta_p^2=0.010$). Follow up tests revealed that there was a larger benefit of testing over restudying pairs on the probability of successful recall at Day 2 for unrelated pairs ($z=12.431$, $p<0.001$, $d=0.20$) than related pairs ($z=9.706$, $p<0.001$, $d=0.16$).

Additionally, this model revealed a significant main effect of the change in similarity of the cue representation ($z=2.453$, $p=0.0142$, $\eta_p^2=0.0015$), suggesting that more change in the representation of the cue across learning is associated with a higher probability of recall at Day 2; Figure 2.7C. Finally, this model showed a significant relatedness by condition by change in target representation across learning interaction ($z=2.075$, $p=0.0380$, $\eta_p^2=0.0011$). Investigation of the slopes revealed that for tested unrelated pairs, there was a significant negative relationship between the probability of final recall success and the change of the representation of the target across learning ($z=2.691$, $p=0.0071$, $d=0.16$), suggesting that more change in the target across learning (i.e. lower correlation values) is associated with higher probability of subsequent recall. This slope was significantly more negative than the slope from unrelated restudied pairs ($z=2.321$, $p=0.020$, $d=0.11$); Figure 2.7D. There was no relationship between change in target representation and probability of subsequent recall across learning for related pairs (p-values>0.1). Together, these results show while the magnitude of representational change that a pair undergoes is associated with its probability of subsequent recall, there may be multiple processes underlying the change

that depend on both the characteristics of the word pair itself and the learning conditions, and that these processes do not all impact the probability of successful recall.

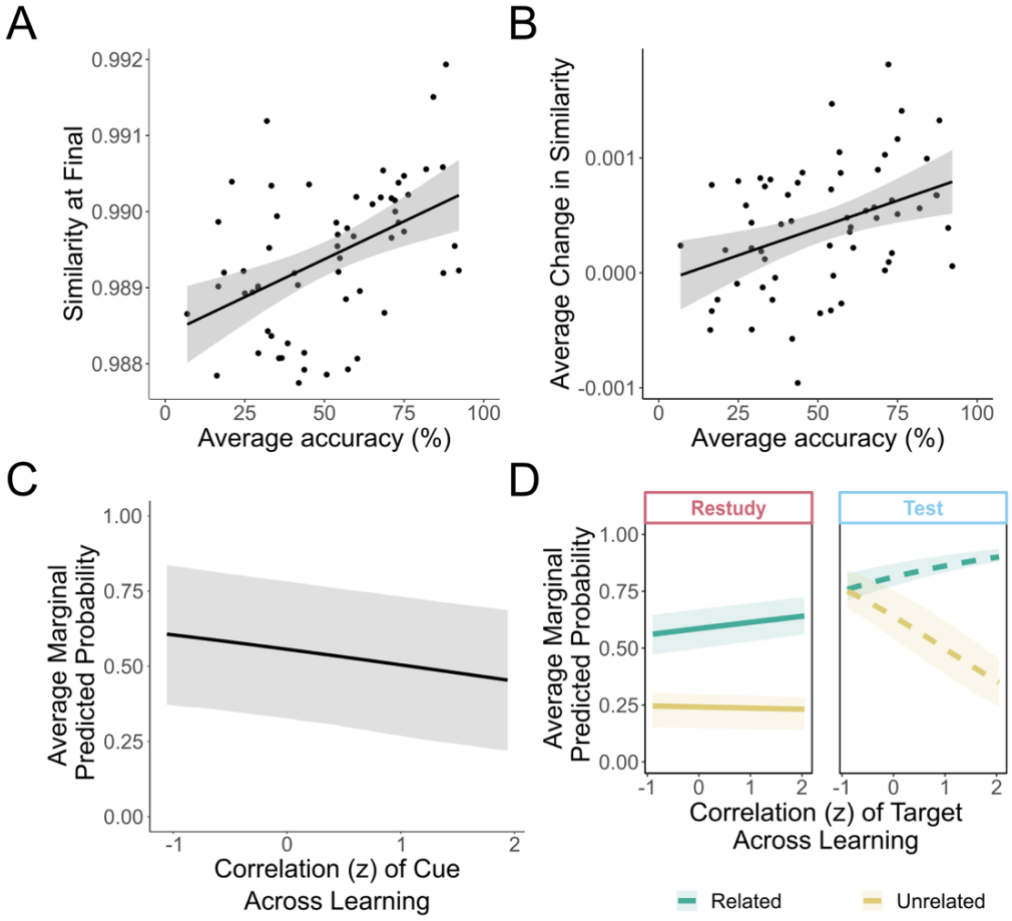


Figure 2.7: Relating representational change to recall accuracy on the final test. A/B. When averaging across 80 participants, word pairs that showed greater within-pair similarity after learning (A; $r_{(58)} = 0.46$, $p<0.001$), and those that showed more representational change (difference in within-pair similarity between final and initial assessment) (B; $r_{(58)} = 0.39$, $p = 0.002$) were recalled with greater accuracy at the final test. N = 60 word pairs. Closed circles reflect average similarity (A) or change in similarity (B) and average final test accuracy across 80 participants. Shaded area reflects 95% CI. C. More change in cue representation after learning (i.e. lower correlation between initial and final) is associated with a higher probability of recall at Day 2 ($z = 2.453$, $p = 0.0142$, $\eta_p^2=0.0015$). D. The impact of the change in target representation across learning on subsequent recall depended on the semantic relatedness of the pair and whether the pair was tested or restudied at Day 1. For unrelated pairs that were tested at Day 1, more change in the representation of the target across learning (i.e. lower values on x-axis) was associated with higher probability of subsequent recall ($z = 2.691$, $p = 0.0071$, $d=0.16$). No other relationships between target word representational change and behavior were significant. Green indicates semantically related pairs, yellow indicates semantically unrelated pairs. Dashed lines indicate tested pairs of words, non-dashed lines indicate restudied pairs of words. For both C and D, N = 3,902 pairs of words from 80 participants. Values on x-axis reflect Fisher's z-transformed correlation of cue representations across learning (C) and measured values of Fisher's z-transformed correlation of targets across learning (D); values on y-axis reflect average marginal predicted probability (probability of subsequent recall across all participants across change of the predictor of interest, holding other predictors constant). Shaded areas reflect 25th-75th percentile of the average marginal predicted probability.

Discussion

Three primary questions were addressed in the current work. First, we sought to determine how semantic relatedness between paired words influences the testing effect. Second, we created a novel extension of a multi-arrangement similarity paradigm (Kriegeskorte & Mur, 2012) to investigate how paired associate learning, supported by either testing or restudying, can shape the semantic representations of individual words. Finally, we assessed whether learning-induced changes in semantic representation were associated with behavioral performance.

To evaluate our first question, we systematically manipulated semantic relatedness between the cue and target with a to-be-learned pair of words and compared accuracy between tested and restudied pairs after approximately 24 hours. We found that although relatedness increases overall performance, it decreases the magnitude of the testing effect by substantially improving performance for restudied pairs, such that the relative additional benefit conferred by testing is less than for unrelated pairs. Crucially, we only observed this interaction between semantic relatedness and learning condition when tested items were split between those successfully and unsuccessfully recalled at the initial testing. This is consistent with previous work showing that, in the absence of feedback, the mnemonic benefits of testing only occur if the target item is successfully recalled during the initial test (Halamish & Bjork, 2011; Kornell et al., 2011; Storm et al., 2014).

Accuracy alone, however, can only provide limited insight into exactly how semantic relatedness differentially improves memory for tested and restudied pairs of words. To address this gap, we developed a novel extension of a multi-arrangement paradigm to simultaneously measure the semantic similarity of sixty words at a time and impute the semantic similarity of words in to-be-learned pairs without them ever being directly measured against one another. In

this analysis, we showed that successful learning, especially of related pairs, draws paired words closer together in semantic space more than unsuccessful learning attempts and pairs that did not undergo learning.

Showing that pairs are drawn together, however, does not show how they become more similar. It is possible that both items within a pair change symmetrically to become more similar to each other; alternatively, one item may remain relatively stable while the other changes. Extant literature investigating these potential hypotheses (Caplan et al., 2014; Kahana, 2002; Madan et al., 2010; Popov et al., 2019; Vaughn & Rawson, 2014) tends to compare outcome measures like accuracy and reaction times when probing pairs in the forward (i.e. A→B) vs backward (i.e. B→A) directions. These measures, while useful for answering some questions, are less effective for exploring associative asymmetry of changes in semantic space, as they cannot compare the overall representations of concepts.

To this end, we compared the semantic representations of individual words across learning. We found that for related pairs, learning induced greater representational change in the semantic structure of cue words than target words, while unrelated cue and target words both underwent an equivalent degree of change. We then adapted an approach from neuroimaging literature for investigating asymmetrical representational change (Bein et al., 2020; Schapiro et al., 2012). If the correlation between pre-learning cues with post-learning is less than the correlation between post-learning cues with pre-learning targets, this implies that learning draws cues towards targets in semantic space. This was indeed the pattern we observed for related word pairs (although as an isolated effect, the negative asymmetry value narrowly failed to survive corrections for multiple comparisons; however, the change in asymmetry relative to unrelated pairs was significant).

The idea that testing creates a directionally-specific (i.e. asymmetric) associative relationship, where the cue-to-target relationship is strengthened without influencing the backward associative target-to-cue link, is consistent with prior theoretical accounts. According to the dual memory theory (Rickard & Pan, 2018) this process occurs by creating an episodic "cue memory" where the cue and target are encoded in the context of a retrieval task, whereas restudying creates a bidirectional association. The transfer appropriate processing account (Morris et al., 1977) posits that the benefit of testing stems from greater episodic contextual similarity between retrieval practice and the final test, relative to restudying.

Consistent with this framework, our results show asymmetric change in cue and target representations across learning; however, this asymmetric change depends on the pre-existing semantic relatedness, rather than learning condition, suggesting that the asymmetrical change within a pair may be driven by the semantic information within the to-be-learned pairs, rather than the creation of an episodic "cue memory" during testing. Other work has suggested that prior knowledge plays a crucial role in the symmetry of concept representations after learning (Bein et al., 2020; Caplan et al., 2014; Popov et al., 2019). For instance, when pairs of famous and novel faces are learned, multivariate neural representations of novel target faces are drawn towards those of their paired cue faces only when there is pre-existing knowledge about the cue face (Bein et al., 2020). While this asymmetric representation is in the opposite direction to the one we observed in our data, it is important to note that in that study there was no pre-existing relationship between the paired faces and no prior knowledge surrounding the novel faces. In contrast, the word stimuli used in our study had a rich network of semantic associations prior to learning, with pre-existing semantic relationships between half of the pairs. It is possible that the assimilation of a target item representation into that of its paired cue item only occurs when existing semantic information

about the cue can scaffold the integration of the novel information into the existing knowledge. When there is pre-existing knowledge about both items in a pair, as was the case in our study, the cue representation instead changes asymmetrically to become more predictive of the upcoming target (Estes & Jones, 2009; Schapiro et al., 2012).

Although words in our corpus that are strongly associated with a given cue word are drawn towards that cue word regardless of the learning condition or relatedness of the associated to-be-learned pair, we only show learning-induced asymmetric sculping of the overall semantic space for semantically related pairs. Accounts of the testing effect such as the elaborative encoding account (Carpenter & Delosh, 2006) or the semantic mediator hypothesis (Carpenter, 2009, 2011; Pyc & Rawson, 2010) propose that mental elaboration during the search for correct answer during testing facilitates later recall by dynamically creating additional retrieval routes via the activation of concepts connecting cues and targets (Carpenter, 2009, 2011) or by increasing relational processing relative to restudying (Rawson et al., 2015; Rawson & Zamary, 2019). Other work has shown that pre-existing semantic relationships between words facilitate integration of the pair, potentially amplifying these effects (Bein et al., 2015). It is possible that even though semantic associates of the cues in unrelated pairs create new links within the to-be-learned pair, the paired words are less likely to co-activate shared concepts enough to change the overall semantic space (Ritvo et al., 2019).

We additionally conducted a set of analyses comparing participants' idiosyncratic semantic representations (derived from the SWAT) to normative semantic representations (derived from word2vec) to test whether these elaborative connections made during learning are truly novel or reflect the sculpting of existing features; see Supplementary Results and Supplementary Figure 2.6. We show that after learning, words in tested pairs are drawn closer to their normative

representations, suggesting that even though learning drives novel connections, testing shapes features that already exist, rather than adding entirely new features to a representation.

The non-monotonic plasticity hypothesis (NMPH) may also help explain the effects we observed of testing on representational change. This account posits that while testing and restudying both strongly co-activate representations of the paired items, testing additionally requires a search process for the to-be-learned target that induces moderate co-activation of other similar items (as has been previously shown to occur during the retrieval of highly similar episodic memories; Kuhl et al., 2011; Wimber et al., 2015), weakening these connections and reducing interference (Ritvo et al., 2019). This is precisely what we found in our analyses of lure representations – testing exerted the biggest impact on moderate strength lures, which were significantly repelled away from cue words, relative to weak/non-lure pairings. This effect is not only consistent with the predictions of NMPH, but also with an emerging body of work showing that competition adaptively distorts and repels overlapping episodic representations so they become less similar (Chanales et al., 2021; Drascher & Kuhl, 2022; Rafidi et al., 2018).

Our last goal was to evaluate the linkage between learning-induced changes in semantic representations and final recall success. To do so, we examined how the mean retrieval success of each word pair (averaged across participants) relates to its mean learning-induced change in representational similarity, and how individual differences in multiple factors affecting representational change relate to the subsequent recall of a given pair. Using the first approach, we showed that word pairs that undergo a greater amount of pairwise representational change (regardless of learning condition) are more likely to be remembered at the final test. Our individual differences approach showed that pairs are more likely to be recalled after a delay when the representation of the cue changes more across learning, while learning-induced change in the

representational space of the target is only associated with final recall success in unrelated pairs that underwent testing. These findings highlight how changes in the representation of the cue (to make it more predictive of the target) are crucial regardless of learning condition, but it may only be necessary to sculpt the representation of the target to create elaborative links between words in a pair if they do not already exist.

One potential limitation of our work comes from our use of pairwise similarity metrics derived indirectly via imputation. If our imputation method was unreliable, it might cast doubt on our behavioral representational change results. We believe that this is not the case and have performed extensive validation analyses of our imputations (see Supplementary Results). We believe that our ability to impute the subjective semantic relatedness of pairs without ever having participants directly judge them is a key innovation of our work over existing approaches such as semantic priming and free association that can only show the relative magnitude of the effects of semantic relatedness through measures like accuracy and reaction time. Moreover, we expect our imputation approach will allow researchers to infer pairwise relationships without running the risk of biasing participants by presenting to-be-learned pairs before learning, nor evoking demand characteristics by having participants explicitly judge the similarity of already-learned pairs, which may occur in traditional multi-arrangement paradigms.

Another potential limitation could come from our admittedly restricted assay of semantic space. Due to experimental time constraints, we were unable to include additional words beyond those in the to-be-learned pairs in our SWAT protocol that would enrich our measurement of semantic space and serve as a null hypothesis test, as they should undergo little or no representational change. We also ensured that the distributions of semantic association across conditions did not overlap so that we could treat relatedness as a dichotomous variable and actively

avoided very strongly related pairs of words so that participants could not easily guess the target word in the absence of successful learning. These design constraints may have resulted in a truncated range of semantic relatedness across all pairs. Recent work has shown that the effect of semantic relatedness may depend on the range of strength of association across the entire stimulus set (Antony et al., 2022), so future work may opt to choose a broader range to determine if this impacts the results.

Despite the general stability of semantic knowledge over the course of one's lifetime, our results demonstrate that even a brief session of episodic learning can subtly yet systematically re-sculpt semantic space. Our novel behavioral representational similarity approach identifies multiple processes supporting episodic memory, where new connections are established between a cue and target, shared semantic information asymmetrically changes cues to become more predictive of their paired target and testing minimizes associations with potentially interfering semantic lures. Together, these changes impart a lingering residue on semantic memory that facilitates later episodic recall. These results are consistent with recent neuropsychological, behavioral, and neuroimaging evidence that the episodic and semantic memory systems may interact through gradients of activation of shared cognitive processes (Burianova et al., 2010; Burianova & Grady, 2007; Rajah & McIntosh, 2005). In this framework, episodes are comprised of both general conceptual reinstatement and episode-specific sensory processing, while recall of semantic memory often includes episodic information about when and where the information was acquired (Renoult et al., 2019). Future studies will be needed to better characterize whether these subtle learning-induced semantic distortions are short-lived or whether they can endure for weeks or months.

Supplementary Methods

Determining our sample size

Meta-analysis of testing effect studies reveals that the advantage of tested words pairs over restudied word pairs on final memory performance is typically a medium size effect, with a mean effect size of g = 0.50 across 159 effect sizes from 61 studies (Rowland, 2014). When restricting the meta-analysis to studies that only included semantically unrelated pairs (g = 0.67) or studies that only included semantically related pairs (g = 0.66), it appears that the size of the testing effect is not heavily impacted by semantic relatedness. However, there is limited literature on the topic, so individual empirical studies might be more informative than a larger meta-analysis(Rowland, 2014). One particularly relevant study (Bulevich et al., 2016) directly compared semantically related triads to semantically unrelated triads in a repeated measures testing effect design with multiple cycles of a study/test manipulation. They found that relatedness had an effect size of d = 2.94 after one test and an effect size of d = 2.58 after two tests, both of which constitute large effect sizes. They also reported a significant learning condition (test vs restudy) x relatedness interaction effect size of d = 0.55, suggesting a medium effect size.

Although the literature suggests medium to large effect sizes for the impact of our learning condition and semantic relatedness manipulations on final test performance, we anticipated that the effect of these manipulations on our representational similarity measures would be considerably smaller (i.e. learning-induced perturbations of semantic space should be subtle, given the overall stability of participants' pre-experimental semantic knowledge). Since there was nothing in the literature to anchor our effect size calculations for our novel representational similarity approach, we assumed a Cohen's d of 0.2 (a small effect size) when calculating our target sample size. Using this effect size, we reasoned that we would need a sample size of at least

73 useable participants to reach a power of 0.80 for the learning condition x relatedness interaction (calculations performed using PANGEA (v0.2); https://jakewestfall.shinyapps.io/pangea/). As such, we aimed to collect useable data from 80 participants.

Inclusion/Exclusion criteria:

Recruitment was restricted to adults (age 18-40) from the United States, Canada, and Mexico, whose first language is English, who have no current or ongoing mental health or neurological condition and who had successfully completed a minimum of 10 prior studies on Prolific.

Additionally, participants who failed to meet the following pre-registered criteria were excluded:

- Fail to complete Day 2 within 28 hours of completing Day 1
- Fail to respond correctly to attention checks
- Have a median RT of less than 500 ms in the first learning session while making relatedness judgements
- Fail to show a difference between relatedness judgements on related and unrelated words
- Report that they believe their data should be excluded or report some sort of technical issue
- Report distraction of greater than 5 out of 7 on our distraction scale for either of the days, or report three or more distractions occurring on either day
- Do not perform a meaningful arrangement (see below for details) on one or more of the trials of the word arrangement task
- Do not have a sufficient performance on the memory tasks. For example, the participants will be excluded if any of the following are true:

    - Those who fail to correctly retype three or more of the restudy trials on the initial learning day
    - Those who fail to recall more than 25% of the tested associates during the initial learning session, or more than 25% across all pairs during the second session
- Explicitly report writing pairs down between sessions

Additionally, initial and final arrangements following imputation were visually inspected to determine compliance with the task. Participants who judged similarity based on lexical characteristics (i.e. by the first letter of each word) or who did not use any meaningful arrangement and randomly placed words on the page, were excluded. Meaningfulness was judged by three independent raters, and participants were excluded if at least two of the three raters agreed that either arrangement was considered not meaningful.

Supplementary Results

Validation of Imputation

Given that one of our main dependent variables in our analyses is the change in dissimilarity/similarity before and after learning, it is imperative to validate that our imputed values are a meaningful approximation of what they would be if they were experimentally measured. First, we sought to determine the appropriate number of nearest neighbors in our imputation algorithm. To do so, we artificially 'lesioned' a random sample of observed pairwise dissimilarity values obtained from an independent sample of pilot participants, and we varied the number of neighbors used from 1 to 50 to impute the missing ('lesioned') values. For each number of neighbors, the Pearson correlation of the imputed dissimilarity value and the true measured dissimilarity value was calculated, as well as the mean error between the imputed and measured

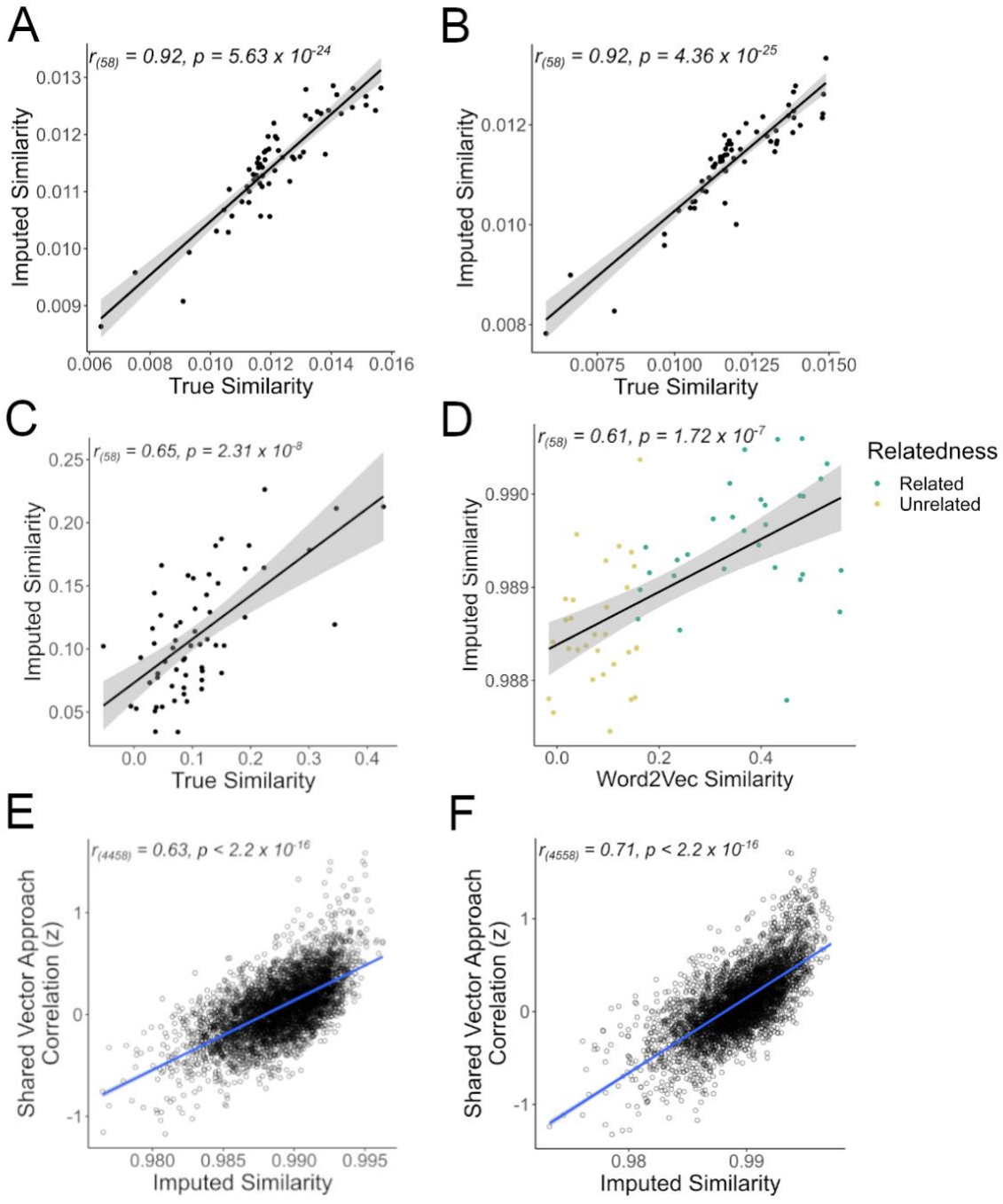
dissimilarity values. A final value of 40 neighbors was chosen, as it maximized the correlation between true and imputed value and minimized error.

Once an optimal number of neighbors was identified from our pilot sample, we turned to our measured data and lesioned a new set of 60 dissimilarity measures from our raw, unimputed data. Although we are theoretically only interested in imputing 60 pre-defined pairs of interest, our actual imputation task requires the imputation of large "blocks" of missing data in the similarity matrix. Lesioning additional values from our unimputed behavioral dataset (rather than 60 uncorrelated values from a fully measured matrix) ensured that the validation imputation process would have access to the approximately the same amount of data as in the final imputation process and equated the total amount of missing data across the validation and true procedure. The lesioned data were imputed, and the new imputed values were correlated with the true values. We also obtained the normative similarity of each to-be-learned pair from word2vec, and then correlated these values with the imputed dissimilarity scores. There were significant correlations between the lesioned and measured values before learning ($r_{(58)} = 0.92$, $p = 5.63 \times 10^{-24}$) and after learning ($r_{(58)} = 0.92$, $p = 4.36 \times 10^{-25}$) (Supplementary Figure 2.1A-B).

We additionally repeated this procedure on semantic space as measured by word2vec. Because the semantic space from word2vec did not include the blocks of missing data that our measured similarity did, we first removed the similarity values that would be missing from the measured data to equate amount of missing data in the imputation. We then artificially lesioned an additional 60 pairs (as was done in the previous validation analyses), imputed the lesioned word2vec values and compared the imputed values to the true values. Just as in the measured behavioral data, there was a significant correlation between the imputed and true values ($r_{(58)} = 0.65$, $p = 2.31 \times 10^{-8}$) (Supplementary Figure 2.1C). Additionally, we compared imputed

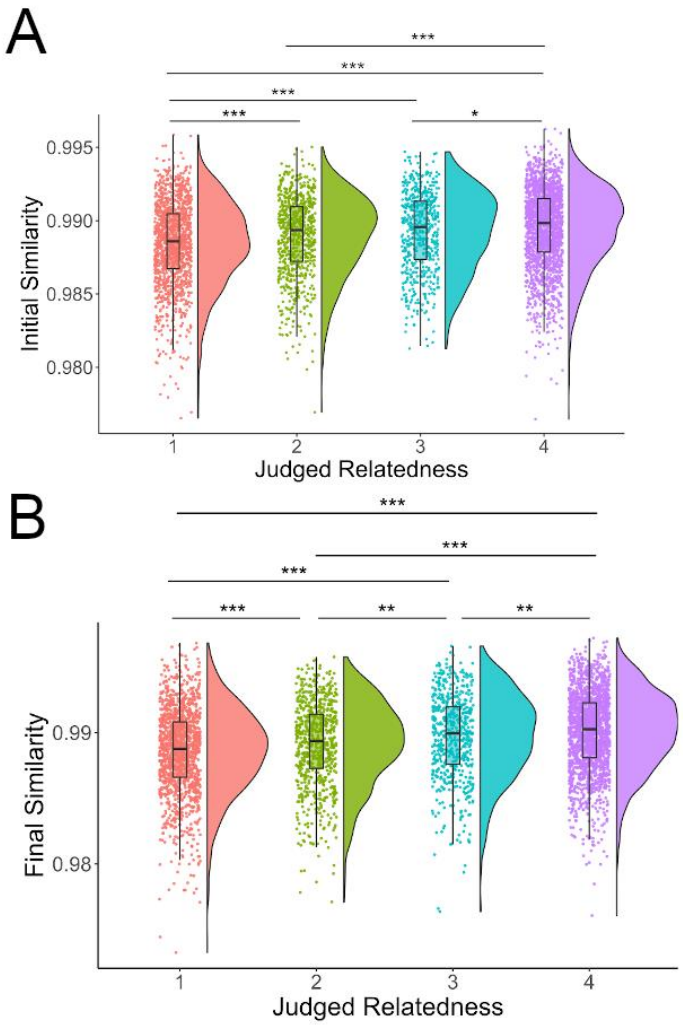
dissimilarity values to similarity values derived from word2vec corpus. This showed a significant correlation ($r_{(58)} = 0.61$, $p = 1.72 \times 10^{-7}$) (Supplementary Figure 2.1D).

An alternative way to approach measuring similarity for a pair of words A and B would be to identify the set of 90 words that both word A and word B were measured against. It would then be possible to operationalize word A and word B by a 90-dimension vector containing those shared values. Concretely, even if the pair GENDER – FEMALE was never directly measured, we could operationalize both GENDER and FEMALE by their distance from the words PARENT, MOTHER, and TEACHER (among the other words that GENDER and FEMALE were both compared against). We could then measure similarity by computing the correlation between the two 90-dimension vectors. There was a significant correlation between the Fisher z-transformed correlation values from the shared vector approach similarity and the imputed similarity of our to-be-learned pairs for both the initial ($r_{(4558)} = 0.63$, $p < 2.2 \times 10^{-16}$) (Supplementary Figure 2.1E) and final arrangements ($r_{(4558)} = 0.71$, $p < 2.2 \times 10^{-16}$) (Supplementary Figure 2.1F). This shared vector approach has the benefit of avoiding imputation entirely; future work may explore whether there are optimal conditions for choosing one approach over the other, or whether the approaches perform similarly.



Supplementary Figure 2.1: Comparing imputed similarity to measured values. A-C Imputed and measured similarity from a "lesioned" semantic space showed significant correlations before learning (A: $r_{(58)}$ = 0.92, p = 5.63 x 10^{-24}), after learning (B: $r_{(58)}$ = 0.92, p = 4.36 x 10^{-25}), and when the procedure was performed on normative semantic space as measured by word2vec (C: $r_{(58)}$ = 0.65, p = 2.31 x 10^{-8}). D. Imputed similarity was significantly correlated ($r_{(58)}$ = 0.61, p = 1.72 x 10^{-7}) with similarity ratings derived from word2vec. Each dot reflects a single similarity score for a pair of words. E-F. Imputed similarity showed significant correlations to a shared vector similarity approach both before learning (E: $r_{(4458)}$ = 0.63, p < 2.2 x 10^{-16}) and after learning (F: $r_{(4458)}$ = 0.71, p < 2.2 x 10^{-16}). For A-D, N = 60 word pairs. For E and F, N = 4,560 word pairs across 80 subjects. Green represents semantically related pairs, yellow semantically unrelated pairs. Shaded area reflects 95% CI.

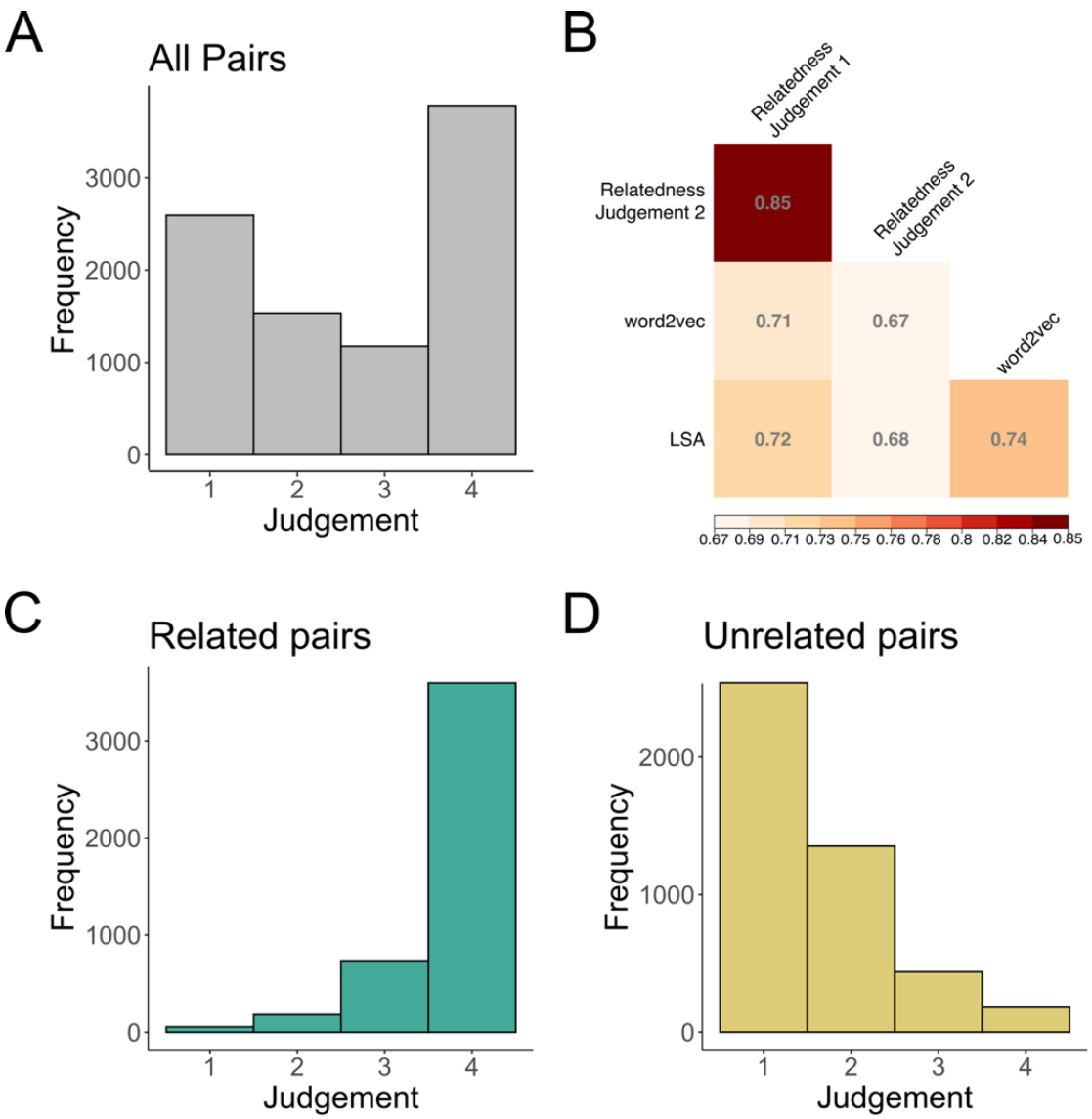
Finally, a linear mixed model was used to compare imputed dissimilarity across participants' explicit behavioral judgements of similarity from the learning phase. Linear mixed modeling showed that imputed dissimilarity was significantly different across relatedness judgements for both initial arrangement ($F_{(3,4538)} = 42.43$, $p = 4.92 \times 10^{-27}$, $\eta_p^2 = 0.03$) and final arrangement ($F_{(3,4539)} = 68.87$, $p = 1.52 \times 10^{-43}$, $\eta_p^2 = 0.04$) (Supplementary Figure 2.2).



Supplementary Figure 2.2: Comparing imputed similarity to relatedness judgements. Linear mixed modeling showed that imputed similarity was significantly different across relatedness judgements for both the initial arrangement (A: $F_{(3,4538)} = 42.43$, $p = 4.92 \times 10^{-27}$, $\eta_p^2 = 0.03$) and final arrangement (B; $F_{(3,4539)} = 68.87$, $p = 1.52 \times 10^{-43}$, $\eta_p^2 = 0.04$). N = 4,550 imputed similarity scores from 80 independent subjects. Dots reflect an individual imputed similarity score. Boxplots reflect median, 25th and 75th percentiles; whiskers show 95% CI. Symbols reflect statistically significant differences across conditions using Holm-Bonferroni corrections for multiple comparisons (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Behavioral Judgements of Relatedness

On average, participants rated semantically related pairs (M = 3.737, SD = 0.578) as more similar to each other than semantically unrelated pairs (M = 1.559, SD = 0.780) ($F_{(1,79)}$ = 3328.743, p = 2.4 x 10^{-66}, η_G^2 = 0.941). Similarity, participants rated semantically related pairs (M = 3.711, SD = 0.609) more likely to be found on the same page of a book or magazine than semantically unrelated pairs (M = 1.676, SD = 0.864) ($F_{(1,77)}$ = 1704.76, p = 2.75 x 10^{-54}, η_G^2 = 0.896). Judgements across questions were significantly correlated with each other ($r_{(4394)}$ = 0.854, p = 0.00012), as well as with word2vec similarity (relatedness judgement: $r_{(4550)}$ = 0.710, p < 2.2 x 10^{-16}; page judgement: $r_{(4533)}$ = 0.675; p < 2.2 x 10^{-16}) and LSA (relatedness judgement: $r_{(4550)}$ = 0.718; p < 2.2 x 10^{-16}; page judgement: $r_{(4533)}$ = 0.675; p < 2.2 x 10^{-16}). Results are visualized in Supplementary Figure 2.3.

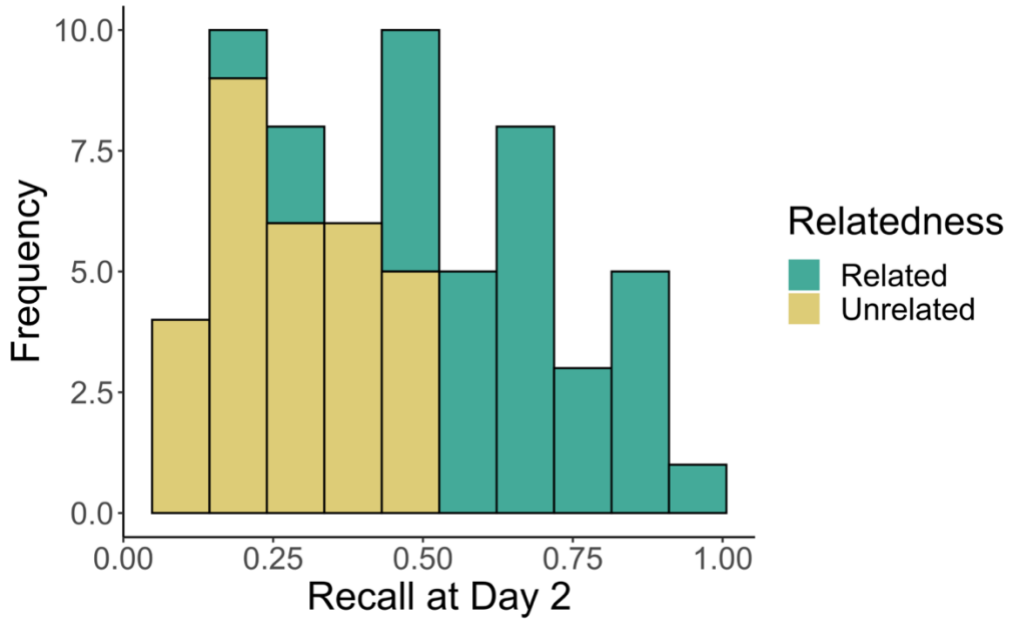


Supplementary Figure 2.3: Behavioral Judgements of Relatedness During Learning. A. Histogram of all relatedness judgements during learning. B. Correlations of participants' relatedness judgements to normative measures of semantic relatedness. C/D. Histogram of relatedness judgements for semantically related and semantically unrelated pairs.

Memorability of Word Pairs

The memorability of each word pair was operationalized by measuring the average accuracy for each word pair across subjects. There was a wide range of memorability across word pairs (Supplementary Figure 2.4; M = 0.455, SD = 0.233, range 0.05 – 0.913). When average accuracy was entered into a between subjects ANOVA, there was no main effect of word set

($F_{(1,116)}$ = 0.623, p = 0.432, η_G^2= 0.005) and no interaction of word set and learning condition ($F_{(1,116)}$ = 0.226, p = 0.635, η_G^2= 0.002).



Supplementary Figure 2.4: Distribution of memorability across word pairs. Memorability was operationalized by computing the average recall accuracy for each word pair across subjects and learning condition. Memorability ranged from 0.05 - 0.913. Despite the wide range of memorability, there was no difference in memorability across word pair sets used to randomize learning condition ($F_{(1,116)}$ = 0.623, p = 0.432, η_G^2= 0.005) or interaction with learning condition ($F_{(1,116)}$ = 0.226, p = 0.635, η_G^2 = 0.002). N = 60 pairs of words.

Investigating Impact of Task Switching

One potential confound to our behavioral results could stem from task-switching during the third learning opportunity (where our learning condition manipulation was introduced). Our test condition (where participants had to actively retrieve and type a cued target word from memory) could be considered quite different from our restudy condition (where participants were asked to restudy the pair by re-typing the presented target word). Because test and restudy trials were intermixed, there was necessarily some degree of task switching each time participants encountered a trial of a different condition than the prior trial. Given the added cognitive load associated with a task switch, it is possible that switch trials would have different learning

outcomes than non-switch trials. To test for the potential impact of task switching during learning, we characterized each trial as either a switch trial (where the learning condition was different from the previous trial; i.e. $trial_{t-1}$ was a test trial and $trial_t$ was a restudy trial) or a non-switch trial (where the learning condition was consistent with the prior trial; i.e. both $trial_{t-1}$ and $trial_t$ were restudy trials) and included this trial type as a factor with learning condition and semantic relatedness in a RM-ANOVA on the behavioral accuracy at Day 2. Regardless of whether or not we bifurcated the test trials into those correctly and incorrectly recalled at Day 1, there was no main effect of trial type (switch trial vs non-switch trial; non-bifurcated: $F_{(1, 79)} = 0.061$, $p = 0.806$, $\eta_G^2 = 3.94 \times 10^{-5;}$ bifurcated: $F_{(1,44)} = 0.828$, $p = 0.368$, $\eta_G^2 = 0.001$), and trial type did not interact with either of our other factors (all p-values > 0.05).

Change in Similarity – To-Be-Learned Pairs

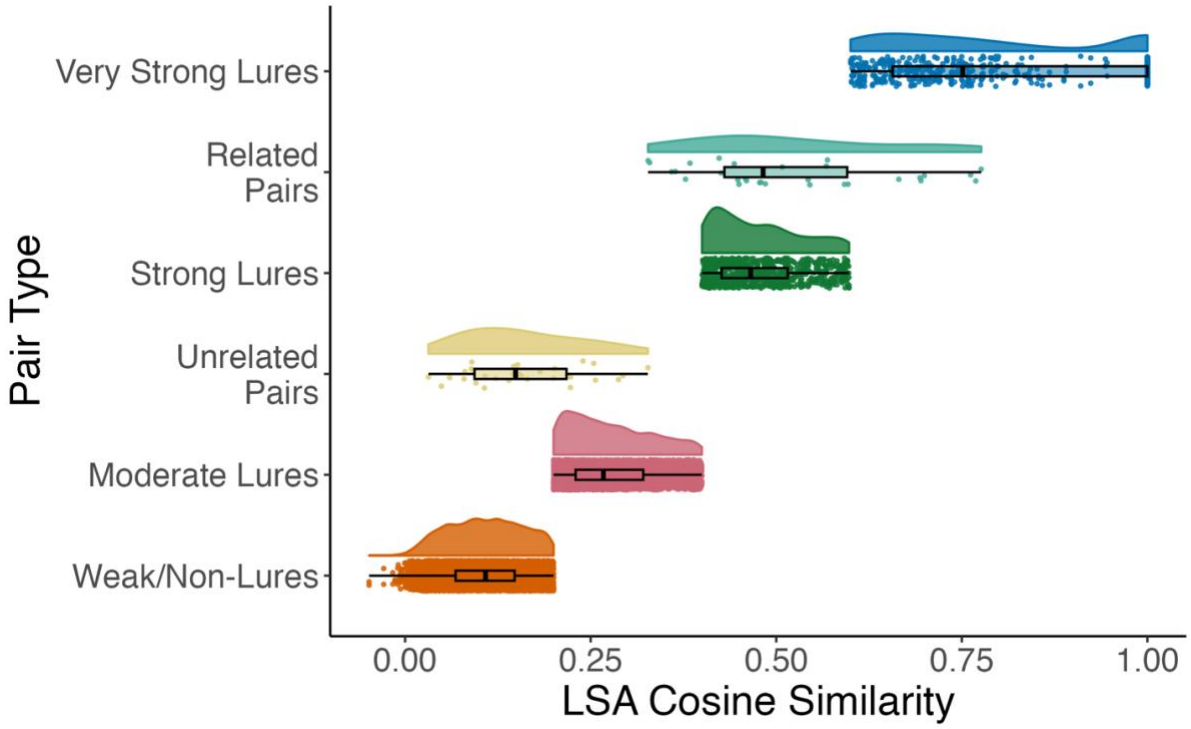
In our main analyses, we showed that the pairwise change in similarity in our conditions of interest were significantly different from zero and report a linear mixed-effects modeling (LMM) approach to show that there were no differences in the amount of pairwise change in similarity across relatedness and learning condition. However, we pre-registered that we would test hypotheses surrounding the differences in pairwise change of similarity using an RM-ANOVA approach, so results of those analyses are included here for completeness. Using an RM-ANOVA on the change in similarity across learning showed no significant main effect for relatedness ($F_{(1,49)} = 1.453$, $p = 0.234$, $\eta_G^2 = 0.004$), no main effect for learning condition ($F_{(1,49)} = 0.415$, $p = 0.523$, $\eta_G^2 = 0.001$), and no relatedness by learning condition interaction ($F_{(1,49)} = 0.394$, $p = 0.533$, $\eta_G^2 = 0.001$). These results are consistent with those from an LMM framework reported in the main text.

Change in Similarity – Lures

In addition to the analyses of the change in similarity of very strong, strong, and moderate lures relative to a weak/non-lure baseline, we additionally tested the other pairwise comparisons of each lure condition across learning condition (i.e. the lure strength by learning condition interaction). In addition to the results reported in the main text, we showed that for restudied pairs, the representations of very strong lures are drawn towards cue words more so than strong lures ($z=5.068$, $p=4.02\times10^{-7}$, Cohen's $d=0.13$) and moderate lures ($z=6.624$, $p=3.50\times10^{-11}$, Cohen's $d=0.15$) are drawn towards cue words. For tested pairs, very strong lures show a similar pattern where they are drawn towards cue words more so than strong lures ($z=4.326$, $p=1.52\times10^{-5}$, Cohen's $d=0.10$) and moderate lures ($z=7.260$, $p=3.87\times10^{-13}$, Cohen's $d=0.15$) are drawn towards cue words; additionally, tested strong lures are also drawn towards cue words more so than tested moderate lures ($z=4.090$, $p=0.0003$, Cohen's $d=0.05$).

Additionally, for each lure strength bin and learning condition combination, we computed a one-sample t-test (with Holm-Bonferroni corrections for multiple comparisons) to determine whether there was change in semantic similarity across learning. For lures associated with restudied pairs, weak/non-lures ($t_{(73369)}=4.815$, $p=5.88\times10^{-6}$, Cohen's $d=0.018$), moderate lures ($t_{(33934)}=3.786$, $p=3.060\times10^{-4}$, Cohen's $d=0.02$), strong lures ($t_{(7861)}=3.929$, $p=2.583\times10^{-4}$, Cohen's $d=0.04$), and very strong lures ($t_{(2242)}=8.234$, $p=5.304\times10^{-16}$, Cohen's $d=0.17$) were drawn closer together after learning.

For lures associated with tested pairs, weak/non-lures ($t_{(799776)}=5.726$, $p=1.03\times10^{-8}$, Cohen's $d=0.02$), strong lures ($t_{(9791)}=4.904$, $p=5.88\times10^{-6}$, Cohen's $d=0.05$), and very strong lures ($t_{(2849)}=7.918$, $p=2.408\times10^{-14}$, Cohen's $d=0.15$) were significantly drawn together after learning; the change of moderate lures associated with tested pairs was not significantly different from zero ($t_{(38679)}=-0.438$, $p=0.661$, Cohen's $d=0.002$).



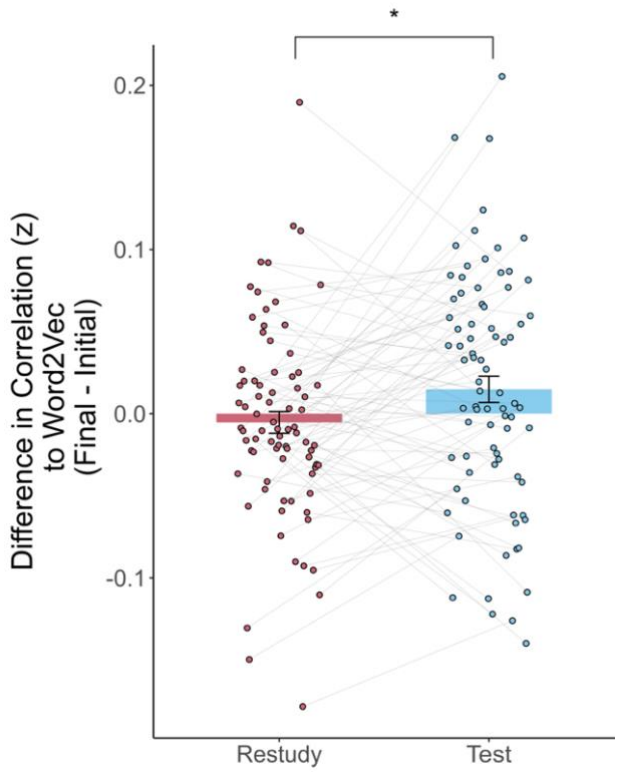
Supplementary Figure 2.5: Distribution of LSA cosine similarity of to-be-learned pairs and potential lure pairs. LSA cosine similarity was computed for all related (n = 30 pairs; M = 0.52, SD = 0.13) and unrelated (n = 30 pairs; M = 0.15, SD = 0.8) to be learned pairs of words. Lure pairs (i.e. all 118 potential word pair combinations for the 60 to-be-learned cues) were each defined as a non-lure (LSA < 0.2; n = 9,056 pairs, M = 0.11, SD = 0.05), weak lure (LSA between 0.2 and 0.4; n = 4011 pairs, M = 0.28, SD = 0.06), moderate lure (LSA between 0.4 and 0.6; n = 844 pairs, M = 0.48, SD = 0.06) and strong lure (LSA above 0.6; n = 429, M = 0.79, SD = 0.06). Dots reflect the LSA cosine similarity value for a given pair of words in our corpus. Boxplots reflect median, 25th and 75th percentiles; whiskers show 95% CI.

Representational change relative to normative space

Another way to test for differential representational change of cue and target words is to compare each word's representation following learning to a normative semantic template. The observed changes can provide insight into how the representational changes relate to a more general, normative structure. That is, we can test whether the changes result from sculpting existing semantic information or by creating idiosyncratic elaborative connections that would not be captured by a normative representation. To do so, each similarity vector (defined by the 20 closest

nearest neighbors, as was reported in analyses in the main text) was correlated with its respective similarity vector derived from word2vec. This value was Fisher z-transformed and the difference in correlation across learning was entered into an LMM. As with the previous analyses in the main text, the semantic relatedness of the pair, learning condition, recall success at final test and word's position in the to-be-learned pair were included as a fixed effect predictors and tested as potential random slopes. Ultimately, no random slopes were included, and only subject identity was included as a random intercept.

This model showed a significant main effect of learning condition ($t_{(1137)} = 2.211$, $p = 0.027$, Cohen's $d = 0.07$), where tested pairs showed more change than restudied pairs; Supplementary Figure 2.6. No other predictors were significant (all t's < 1, all p-values > 0.5). Taken together with our results from the main text—where we show that strongly related lures are pulled closer to the to-be-learned cue and that testing pushes moderately related lures further away relative to the baseline general changes to semantic space—these results suggest that testing may shape existing features of the paired concepts, rather than adding new features to the representation, and that this process occurs for both cue and target words.



Supplementary Figure 2.6: Relating idiosyncratic semantic structure to normative semantic space. Words in tested pairs, regardless of word position in pair or semantic relatedness of pair, showed more change relative to analogous normative semantic representations derived from word2vec ($t_{(1137)}$ = 2.211, p = 0.027). All displayed correlation values are Fisher *r*-to-*z* transformed. Maroon bar reflects restudied pairs, blue bar reflects tested pairs. Open circles reflect means of individual subjects, with connecting lines showing within-subject differences across conditions. Error bars reflect standard error of the mean. Symbols reflect statistically significant differences across conditions using Holm-Bonferroni corrections for multiple comparisons (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Correlation of Behavior and Change in Similarity

In the main text, we report that the pairwise similarity at the final SWAT protocol and the amount of representational change across learning is correlated with the final recall accuracy of the pair. Despite the relationship between pairwise change in similarity and behavioral accuracy, there was no correlation between the magnitude of an individual's behavioral testing effect and their average change in similarity for tested and restudied pairs when averaging across words ($r_{(78)}$=-0.16, p=0.15).

Fixed Effects

	Coefficient	*t-value*	*S.E.*	*p-value*
Pair Type - Learned, Correct at Day 2	0.0004	3.1121	0.0001	0.0021
Pair Type - Learned, Incorrect at Day 2	0.00007	0.5465	0.0001	0.5852
Num.Obs.	240			
AIC	-2637.4			
BIC	-2620			
ICC	0.2			
RMSE	0.0007			

Random Effects

	Variable	*Variance*	*Std. Dev*
PTID	(Intercept)	1.78E-07	0.00042243
Residual		6.37E-07	0.00079788

Supplementary Table 2.1: Fixed and random effects for LMMs investigating pairwise change depending on whether pair was recalled correctly at Day 2 and whether it was designated as a to-be-learned pair vs semantic lure

Fixed Effects

	Coefficient	*t-value*	*S.E.*	*p-value*
Memory Success at Day 2	0.0003*	1.965	0.0002	0.0499
Learning Condition	0.00006	0.3933	0.0002	0.6942
Relatedness	-0.0002	-1.3161	0.0002	0.1887
Lure Strength - Weak Lures				
Lure Strength - Moderate Lures				
Lure Strength - Strong Lures				
Relatedness x Lure Strength - Weak Lures				
Relatedness x Lure Strength - Moderate Lures				
Relatedness x Lure Strength - Strong Lures				
Learning Condition x Lure Strength - Weak Lures				
Learning Condition x Lure Strength - Moderate Lures				
Learning Condition x Lure Strength - Strong Lures				
Learning Condition x Relatedness				
Num.Obs.	605			
AIC	-5685.6			
BIC	-5659.2			
ICC	0.1			
RMSE	0.002			

Random Effects

	Variable	Variance	Std. Dev
PTID	(Intercept)	4.89E-07	0.00069945
Residual		3.96E-06	0.00198879

Supplementary Table 2.2 Fixed and random effects for LMMs investigating pairwise change of to-be-learned pairs

Fixed Effects

	Coefficient	*t-value*	*S.E.*	*p-value*
Learning Condition	0.0001	1.0232	0.0001	0.3062
Relatedness	0.0002*	2.1957	0.0001	0.0281
Lure Strength - Weak Lures	0.00004	0.6908	0.00005	0.4897
Lure Strength - Moderate Lures	0.0002*	2.3437	0.00009	0.0191
Lure Strength - Strong Lures	0.0010***	6.1001	0.0002	<0.0001
Relatedness x Lure Strength - Weak Lures	-0.00009	-1.2492	0.00007	0.2116
Relatedness x Lure Strength - Moderate Lures	-0.0001	-0.8615	0.0001	0.389
Relatedness x Lure Strength - Strong Lures	0.00009	0.4255	0.0002	0.6705
Learning Condition x Lure Strength - Weak Lures	-0.0002**	-2.8405	0.00007	0.0045
Learning Condition x Lure Strength - Moderate Lures	0.000006	0.0485	0.0001	0.9613
Learning Condition x Lure Strength - Strong Lures	-0.0002	-1.0528	0.0002	0.2924
Learning Condition x Relatedness	-0.0002*	-2.3051	0.00007	0.0212
Num.Obs.	248508			
AIC	-1747564.2			
BIC	-1747355.7			
ICC	0.01			
RMSE	0.007			

Random Effects

	Variable 1	Variable 2	Variance	Std. Dev
PTID	(Intercept)		4.78E-07	0.00069123
	Relatedness		7.26E-07	0.00085178
	Learning Condition		6.17E-07	0.00078558
	(Intercept)	Relatedness	-2.86E-07	-0.4856958
	(Intercept)	Learning Condition	-3.58E-07	-0.6593156
	Relatedness	Learning Condition	2.16E-07	0.32304773
Residual			5.15E-05	0.00717846

Supplementary Table 2.3: Fixed and random effects for LMMs investigating pairwise change of semantic lures.

Fixed Effects

	Coefficient	*t-value*	*S.E.*	*p-value*
Memory Success at Day 2	-0.0059	-0.3988	0.0148	0.6902
Learning Condition	0.0002	0.0118	0.0177	0.9906
Relatedness	0.0356	2.4146	0.0147	0.0161
Num.Obs.	605			
AIC	-257.7			
BIC	-222.5			
ICC	0.1			
RMSE	0.17			

Random Effects

	Variable 1	Variable 2	Variance	Std. Dev.
PTID	(Intercept)		0.00021896	0.01479713
	Learning Condition		0.00754632	0.08686956
	(Intercept)	Learning Condition	0.00070642	0.54956229
Residual			0.03256754	0.18046478

Supplementary Table 2.4: Fixed and random effects for LMM investigating asymmetry

	Coefficient	*t-value*	*S.E.*	*p-value*
Memory Success at Day 2	-0.078	-2.6604	0.0293	0.0079
Learning Condition	-0.0234	-0.7893	0.0296	0.4301
Word Position	-0.0071	-0.2384	0.0299	0.8116
Relatedness	-0.0309	-1.1301	0.0274	0.2586
Memory Success at Day 2 x Learning Condition	0.0494	1.258	0.0393	0.2086
Memory Success at Day 2 x Position	0.0995	2.5814	0.0385	0.01
Memory Success at Day 2 x Relatedness	0.077	1.9863	0.0387	0.0472
Learning Condition x Word Position	0.0249	0.7673	0.0324	0.4431
Learning Condition x Relatedness	0.0296	0.9085	0.0325	0.3638
Word Position x Relatedness	0.0438	1.3558	0.0323	0.1754
Memory Success at Day 2 x Learning Condition x Word Position	-0.0329	-0.7266	0.0452	0.4676
Memory Success at Day 2 x Learning Condition x Relatedness	-0.0583	-1.2841	0.0454	0.1994
Memory Success at Day 2 x Word Position x Relatedness	-0.1179	-2.6071	0.0452	0.0092
Num.Obs.	1210			
AIC	-164			
BIC	-57			
ICC	0.3			
RMSE	0.18			

Random Effects

	Variable 1	Variable 2	Variance	Std. Dev
PTID	(Intercept)		0.01523587	0.12343366
	Word Position		0.01122176	0.10593279
	Learning Condition		0.00499679	0.07068795
	(Intercept)	Word Position	-0.005304	-0.4056352
	(Intercept)	Learning Condition	-0.0013259	-0.1519599
	Word Position	Learning Condition	0.00370919	0.49534014
Residual			0.03841899	0.19600763

Supplementary Table 2.5: Fixed and random effects for LMMs investigating overall change in semantic space across learning

Fixed Effects

	Coefficient	*t-value*	*S.E.*	*p-value*
Memory Success at Day 2	-0.0018	-0.1975	0.0093	0.8434
Learning Condition	0.0207	2.2159	0.0093	0.0269
Word Position	0.0009	0.1002	0.0093	0.9202
Relatedness	-0.0058	-0.6275	0.0093	0.5304
Num.Obs.	1210			
AIC	-881.5			
BIC	-845.8			
ICC	0.05			
RMSE	0.16			

Random Effects

	Variable	Variance	Std. Dev
PTID	(Intercept)	0.00134274	0.03664345
Residual		0.02616052	0.16174213

Supplementary Table 2.6: Fixed and random effects for LMMs investigating difference in representational change relative to normative semantic space

Fixed Effects

	Coefficient	*t-value*	*S.E.*	*p-value*
Learning Condition	1.1809	6.3238	0.1867	<0.0001
Relatedness	-1.7256	-10.5724	0.1632	<0.0001
Change of Cue	-0.4286	-2.4531	0.1747	0.0142
Change of Target	0.1215	0.7137	0.1702	0.4754
Change in Cue Correlation to Word2vec	-0.1566	-0.7601	0.206	0.4472
Change in Target Correlation to Word2vec	-0.1818	-0.852	0.2134	0.3942
Asymmetry Value	-0.0111	-0.0588	0.1884	0.9532
Relatedness x Learning Condition	0.6167	2.4537	0.2513	0.0141
Relatedness x Change of Cue	0.1676	0.6562	0.2555	0.5117
Relatedness x Change of Target	-0.1509	-0.609	0.2478	0.5426
Relatedness x Change in Cue Correlation to Word2vec	0.4085	1.3245	0.3084	0.1853
Relatedness x Change in Target Correlation to Word2vec	0.155	0.5005	0.3097	0.6167
Relatedness x Asymmetry Value	0.027	0.0946	0.2852	0.9246
Learning Condition x Change of Cue	0.0391	0.132	0.2964	0.895
Learning Condition x Change of Target	0.2558	0.8572	0.2984	0.3913
Learning Condition x Change in Cue Correlation to Word2vec	-0.2701	-0.7493	0.3604	0.4537
Learning Condition x Change in Target Correlation to Word2vec	0.4831	1.2923	0.3738	0.1962
Learning Condition x Asymmetry Value	-0.2636	-0.7977	0.3304	0.425
Relatedness x Learning Condition x Change of Cue	0.3182	0.7712	0.4127	0.4406
Relatedness x Learning Condition x Change of Target	-0.8631	-2.0745	0.416	0.038
Relatedness x Learning Condition x Change in Cue Correlation to Word2vec	0.4297	0.823	0.5222	0.4105
Relatedness x Learning Condition x Change in Target Correlation to Word2vec	-0.1218	-0.2325	0.5238	0.8162
Relatedness x Learning Condition x Asymmetry Value	0.2818	0.5972	0.4718	0.5504
Num.Obs.	3902			
AIC	4430.3			
BIC	4618.4			
ICC	0.1			
RMSE	0.42			

Random Effects

	Variable 1	Variable 2	Variance	Std. Dev
PTID	(Intercept)		3.90E-07	0.00062453
	Learning Condition		4.11E-07	0.00064075
	Relatedness		2.94E-07	0.00054253
	(Intercept)	Learning Condition	-1.70E-07	-0.4237948
	(Intercept)	Relatedness	-9.88E-08	-0.2915987
	Learning Condition	Relatedness	-3.26E-08	-0.0938087
Residual			3.66E-05	0.00604818

Supplementary Table 2.7: Fixed and random effects for LMM investigating behavioral relevance of representational change

Chapter 3: Understanding individual differences in memory for faces: cognitive factor decomposition depends on the nature of memory assessment.

Introduction

Humans rely on our ability to differentiate and recognize faces every day to navigate our social worlds. Despite its ubiquity, the processes underlying it are deceivingly complex. An extensive body of literature has identified a core and extended network of brain regions associated with face processing, including regions across the ventral visual stream, amygdala, ventral anterior temporal lobe and posterior superior temporal sulcus (Collins & Olson, 2014; Grill-Spector & Malach, 2004; J. V. Haxby et al., 1996, 2002; Jonas et al., 2015; Kanwisher & Yovel, 2006; Pitcher et al., 2007). Activation and face-selectivity in these regions has been associated with face memory ability in healthy (Elbich & Scherf, 2017; Furl et al., 2011) and congenital prosopagnosia populations (Duchaine & Nakayama, 2006; Guo et al., 2018). These experimental results are congruent with lesion and stimulation studies (Damasio et al., 1982; Jonas et al., 2015) that implicate these regions in face processing.

Other theoretical work, however, has suggested that memory for faces is more expansive than just the visual recognition of faces: each individual is associated with an abstract "personal identity node," which in turn activates identity specific semantic information nodes, face recognition units and name codes (Bruce & Young, 1986; Burton et al., 1999). Recent reviews emphasize that while learning about individuals may begin with visual familiarization, it also includes the acquisition of personal knowledge, including personal biographical semantic information, episodic information, personality traits and personal beliefs (Anzellotti & Young, 2020; Kovács, 2020).

The association of faces with personal semantic information changes how they are represented in the brain. Faces learned with biographical information are represented more similarly to each other than to faces learned without (Verosky et al., 2013), and learning about individuals through even brief social interactions changes how faces are represented in face-selective cortex (Sliwinska et al., 2022). Additionally, behavioral work has also shown that novel faces learned with similar conceptual information are represented as more perceptually similar (Oh et al., 2021), highlighting how conceptual knowledge impacts even the basic perception of faces.

Although personal semantic information is not necessarily probed during a visual recognition task, there is still evidence that having information associated with a face improves performance on these tasks. Recognition is improved when a face is given a name, even relative to seeing multiple views of a face (Yovel & Schwartz, 2016), and encoding a face conceptually (i.e. encoding by asking about personality traits associated with the face) improves recognition over encoding a face perceptually (i.e. encoding by asking about the shape of the face) (Schwartz & Yovel, 2019). Together, these findings suggest that considering a face as a socially meaningful concept by integrating identity information across domains, as opposed to just an image-based percept, improves recognition.

Despite the growing literature suggesting that face memory may be impacted by more than just the visual characteristics of the face itself, there have been relatively few studies explicitly investigating how processes outside of the core face-processing network impact memory for faces. One recent study demonstrated that hyperconnectivity between face-selective regions and anterior temporal cortex, which has been proposed to be a hub for semantic personal identity knowledge, is associated with stronger face memory performance (Levakov et al., 2023). Other neuroimaging work has highlighted more wide scale connectivity between face-selective regions and regions

implicated in social, memory and auditory regions is associated with better face memory performance measured by the CFMT (Ramot et al., 2019).

In the present study, we sought to understand how cognitive systems outside of the traditional face processing system support memory for faces. To that end, we applied exploratory factor analysis (EFA) (Fabrigar et al., 1999) to a broad battery of behavioral tasks indexing face perception, working memory, associative memory, socio-emotional cognition, and general fluid intelligence. We use the resulting factors to explain performance on the Cambridge Face Memory Task (CFMT) (Duchaine & Nakayama, 2006a), a canonical measure of face memory ability where images of faces are learned in isolation without any socially relevant information, and a task that we developed – the Personal Identity Memory (PIM) task. This new task seeks to index face memory and personal identity knowledge in a more ecologically valid fashion – identities are presented through dynamic videos and are associated with semantic information, and participants are tested on identity recognition in novel contexts and on the recall of the personal semantic information. Given the literature associating performance on the CFMT with connectivity between the face-processing network and other cognitive domains, we expected to see factors encompassing face perception, memory and social cognition predicting performance on the CFMT, with similar combinations predicting performance on the recognition phase of PIM task. Although we expected that recognition performance on the PIM task may be predicted by similar factors as the CFMT, we hypothesized that the more complex and ecologically valid nature of the PIM task may allow for the possibility of multiple strategies to accomplish the task, while the CFMT may be a more process-pure measure of face memory. To test this hypothesis, we used k-means clustering (Hartigan & Wong, 1979), a data-driven clustering approach, to identify potential profiles of cognitive performance in our sample, which we subsequently used to test whether

individuals used different cognitive processes to support face memory performance depending on their abilities.

Methods

Task Battery

In order to investigate the array of cognitive processes that may underlie face memory ability, we designed a battery of cognitive tasks to probe performance across 6 cognitive domains: face processing/memory, object processing, general long-term memory, working memory, social cognition and general/fluid intelligence. A brief description of the tasks included and the outcome measure associated with each task is listed below; additional detailed descriptions of the tasks are provided at: https://jankawis.github.io/battery_of_tasks_WIS_UCLA/intro.html.

Face Processing/Memory Tasks:

- **Cambridge Face Memory Test** (CFMT; original (Duchaine & Nakayama, 2006a), Australian (McKone et al., 2011) and Female versions (Arrington et al., 2022)): Participants study 6 faces (cropped to exclude features like hair and glasses) presented in black and white. After the initial learning phase, they are tested using a three alternative forced choice format across three phases. In the first phase, participants are presented with images used in the learning phrase. The second and third phases included novel views, with the third phase additionally including random Gaussian noise overlayed on the images.
- **Glasgow Face Matching Task** (GFMT; long version) (Burton et al., 2010): Participants are simultaneously presented with two faces and must decide whether the images are photos of the same person.

- **Face Memory and Perception** (FMP) task: A novel task designed to parametrically dissociate processes of face memory and face perception. On each trial, participants perform a match to sample task, where they are presented with a target face and two probe faces and must decide which of the probe faces are a match to the target. Face perception difficulty was modulated by the perceptual similarity of the probe and target. Face memory difficulty was modulated by the delay between presentation of the target and the probes. In the No Delay condition, the target and probe faces appear on the screen simultaneously for 3s. In the Unfilled Delay condition, targets appear for 3s, followed by a 4s delay with a blank screen, after which the probe faces appear. In the Filled Delay condition, targets appear for 3s, followed by a 4s delay during which there is an additional face presented that depicts an emotion. After the delay, the probe faces appear and participants must identify the match to the target, following which they must perform a two-alternative forced choice task identifying the emotion on the face presented during the delay.

Object Processing/Memory Tasks:

- **Cambridge Car Memory Test** (CCMT) (Dennett et al., 2011): A task designed to match the format of the CFMT but use cars as the target stimuli. Participants study images of 6 cars, which they are tested using a three alternative forced choice test across three phases. In the first phase, participants are presented with images used in the learning phrase. The second and third phases included novel views, with the third phase additionally overlaying random Gaussian noise on the images.
- **Vanderbilt Expertise Tests** (VET) (McGugin et al., 2012): A task designed to be similar in structure to the CCMT, but expand the domains of stimuli. Participants study 6 target

images, and then complete two rounds of testing with feedback using a 3 alternative forced choice format, with each testing round followed by an additional study round. Finally, participants complete a series of test trials without feedback. Our task battery included the birds, leaves, and planes tests.

- **Mnemonic Similarity Task** (MST) (Stark et al., 2019): Participants are presented with a continuous flow of images of common objects. Participants must judge whether each object is new (i.e. has not been presented before), old (i.e. the exact image has been presented before) or a lure (i.e. a similar image to an image that has been presented before).

General Associative Memory Tasks:

- **Relational and Item-Specific Encoding** (RISE) Test (Ragland et al., 2012): Participants are presented with pairs of common objects. After a delay, participants are presented with intact or shuffled pairs and must identify whether pairs are shuffled or intact. Item-specific blocks were not included.

Working Memory Tasks:

- **Parametric Go/No-Go** (PGNG) Task (Langenecker et al., 2007; Votruba & Langenecker, 2013): Participants are presented with a sequential series of letters and must press a key when a target letter is presented. This task parametrically modulates the difficulty in two ways: the number of targets (2 or 3) in a target set and the instructions, creating 4 "levels." In the baseline level, there is a two-letter target set (x and y), and the participant must press a key when they see either target. In Level 1, there is a three-letter target set (x, y, and z) and the participant must press a key when they see any of the targets. In Level 2, there is a

two-letter target set, but participants must inhibit their response if the target is repeated. In Level 3, there is a three-letter target set and the participant must inhibit their response if the target is repeated.

- **Digit Span**: Participants are presented with sequential strings of digits from 2-9 digits long. After the entire string of digits is presented, participants must recall the string. Participants are given two strings of each length and complete all string lengths. Modeled after the digit span test in the WAIS-III (Ryan & Lopez, 2001).
- **Automated Operation Span** (AOS) Test (Unsworth et al., 2005): Participants are asked to recall strings of letters with lengths of 3-7 letters. Between the presentation of each letter, participants are presented with a simple arithmetic equation and must determine whether it is correct or incorrect.
- **Spatial Working Memory Capacity** (SCAP) Task: Participants are presented with spatial arrays of 3, 5 or 7 dots. After a short delay, participants view a probe dot and must indicate whether the location of the probe dot was present in the target array. Modeled on (Glahn et al., 2002)
- **Fractal N-back task** (Ragland et al., 2015): Participants are presented with a sequential series of fractal images and must perform 0-back, 1-back and 2-back tasks. In the 0-back task, participants must press a key when they detect a specific fractal target. In the 1-back and 2-back conditions, participants must press a key when a fractal is a match for the fractal presented 1 or 2 trials previously (for the 1-back and 2-back tasks, respectively).

Socio-emotional Cognition Tasks:

- **Emotion Labeling Task** (Palermo et al., 2013): Participants are presented with a single face depicting an emotion and must choose from 6 options which emotion is displayed.
- **Emotion Matching Task** (Palermo et al., 2013): Participants are presented with three emotional faces, two of which are depicting the same emotion. Participants must select the face where the emotion does not match the other two.

General/Fluid Intelligence Tasks:

- **Raven's Advanced Progressive Matrices** (RAPM) (Raven et al., 1998): Participants are presented with 8 symbols that form a pattern and must identify which symbol (out of 8 choices) completes the pattern. Participants must complete as many trials as possible out of 18 trials within 15 minutes.
- **Letter Sets Task** (Ekstrom et al., 1976): Participants are presented with 5 four-letter strings. Four of these five strings follow a pattern, and the participant must identify which string does not follow the pattern. This task includes 15 trials; the participant must complete as many as possible in 7 minutes.
- **Number Series Task** (Fisher et al., 2014; Thurstone, 1938): An adaptive test where participants are shown several series of digits with 1-2 missing digits that follow a rule and must identify the missing digit(s) to complete the pattern. All participants are shown the first three moderately difficult items and depending on how many items are completed correctly in the first set, complete a set of 3 additional questions. The difficulty of the second set is adaptively set depending on performance on the first set.

Personal Identity Memory (PIM) Task

In order to characterize memory in a more holistic and ecologically valid manner, we developed a novel task: the Personal Identity Memory (PIM) Task (Figure 3.1). This task was comprised of 24 identities (12 male and 12 female), learned across 4 blocks of 6 identities (3M/3F). The task was designed with the cover story that the identities were secret agents and the participant was their spy master, so they must learn the information to best manage their agents. Each identity was paired with a code name, a country, a hobby, and a vice. Code names were the names of common animals to avoid pre-existing associations with common first names and ease of translation to Hebrew for planned follow-up studies. Code name and location were randomized for each participant, but hobby and vice were associated with the same identity across participants, as they were chosen to be at least loosely consistent with the video (for example, one identity is shown playing the piano and her hobby was "Plays in Orchestra").

Each identity was presented through a dynamic 10s video, which was created using video clips of international reality television shows, including Big Brother - UK, Big Brother – Canada, The Bachelor – Australia, and Survivor – Australia. Each video showed an identity across multiple contexts/locations and wearing different outfits to create a multi-dimensional, dynamic person representation of each identity. Videos were in color, did not include sound and had the identity's name displayed below the video. After each video completed, participants were shown an "agent ID card," which contained an image of the identity from the video with the identity's associated semantic information and given an additional 15s to learn the associated semantic information. Presentation order of identities within a block and overall order of blocks were randomized for each participant.

After all identities in a block were studied, recognition memory was queried using novel photos and matched lures in a Recognition Task. Participants were presented with either a novel image of the identities from the block (i.e. an image that was not taken from the introduction video) or a matched lure in a random order. After making a Yes/No judgment as to whether they had seen the identity before, participants were asked to make a confidence judgment from 1 (Very confident) to 3 (Not at all confident).

Finally, participants completed a semantic recall task, where they were required to recall the semantic information associated with a given identity. In this Recall Task, there were no lure images and participants were cued using a screenshot of the identity from the introduction video. First, participants performed free recall of the associated name and location (indexed by typing their responses), after which they performed a 6 alternative forced choice multiple choice test of all four pieces of semantic information. The options for the multiple choice phase came from the information associated with the other identities in the block. No time limit was placed on either the free recall or multiple choice phases of the Recall Task.

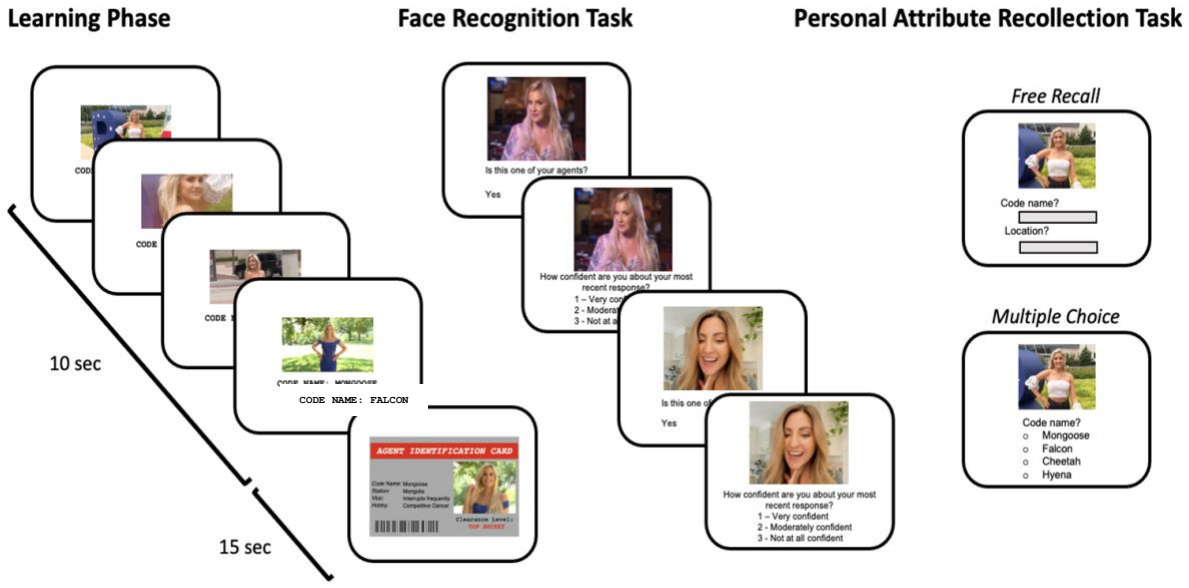


Figure 3.1. Personal Identity Memory (PIM) task design. Each 'agent' was first introduced along with their code name by a silent 10-sec video montage. Video clips were used instead of static images as they showed each agent from a variety of angles, with varied emotional expressions and environmental contexts. After the video concluded, participants were shown an 'agent identification card' with additional personal details. After each set of 6 agents were introduced, participants were given a face recognition task, where they were shown stills of their agent drawn from a new context and stills of matched lures. Participants responded whether the image shown was one of their agents or not and asked to rank the confidence of their judgement. Once the participants had gone through the 12 recognition trials (6 agents, 6 lures), they proceeded to the personal attribute recollection task. In this task, participants were shown a still image of one of their agents and freely recalled the associated code name and location of the agent. After the free recall task, participants were given a multiple-choice test of all four attributes to allow for retrieval of information that was not encoded strongly enough to be recalled without a cue. This process continued until participants encoded and were tested on all 24 of their agents (4 sets of 6 agents).

Performance Summary Measures:

The following task performance summary measures were derived for use in further analysis:

- CFMT, CCMT, GFMT, RISE, SCAP, RAPM and Letter Sets tasks: overall accuracy across all trials (although note that for the RAPM and Letter Sets tasks, the total number of trials was the 18 and 15, respectively, not the number of trials that a participant completed).
- VET: overall accuracy across all three subscales
- AOS: overall accuracy of string recall across all string lengths

- FMP: accuracy for No Delay and Filled Delay conditions, averaging over the levels of perceptual difficulty.
- Fractal N-back: d' from the 2-back condition
- MST: Recognition Score, which was calculated as the difference of hit rate (response of "old "to targets) and false alarm rate (response of "old" to foil images) and Lure Discrimination Index (LDI), calculated as the difference in probability of a "similar" response to a lure image and probability of a "similar" response to a foil image.
- PGNG: percent correct to target trials (PCTT; the percentage of correct target responses across Levels 1-3) and percent correct inhibitory trials (PCIT; the percentage of correctly inhibited trials across Levels 2 and 3)
- Digit Span: maximum capacity, operationalized as the maximum length of string recalled successfully before making two sequential errors
- Number Series: total number of questions correct, with the assumption that "easier" blocks in the hierarchy of Set 2 difficulties were all performed correctly. For example, if a participant is assigned to the second most difficult Set 2, it is assumed that the two easier sets would have been completed correctly.
- Emotion Labeling and Matching: The two tasks measuring emotion processing were the least reliable at capturing individual differences and collecting the necessary number of trials to reach a sufficient reliability would have been too time intensive to collect (Kadlec et al., 2023). As such, we scored each trial as correct or incorrect and took the average percentage correct across both tasks to increase the number of trials.
- PIM – Recognition Phase: overall accuracy measure, which simply scores every trial as correct or incorrect and takes the percentage of correct trials.

- PIM – Recall Phase: overall accuracy on all multiple choice questions and accuracy on free response questions (after responses had been checked for spelling errors – see below for spell-checking procedure).

Additional Survey Tasks:

In addition to the tasks listed above, participants also completed the 20 Question Prosopagnosia Index (PI20) (Shah et al., 2015) and the OSVIQ (Blazhenkova & Kozhevnikov, 2008).

Data collection

All data were recruited online through the online platform Prolific (www.prolific.co). We first recruited participants to participate in a battery of tasks that spanned three days. 298 participants began the first day of this task battery; 257 (131 females, 120 males, 6 not stated; mean age 29.8 +/- 7.7) completed the first day of experimental testing. Of those, 244 completed the full 3-day battery. Participants were paid at a rate of $9.50/hour, plus an additional $9 if they finished all three days of testing within one week.

Eight months after the completion of the original battery, we invited participants who completed the full 3-day battery back to complete additional versions of our tasks to increase our power and ability to reliably estimate individual differences in cognition. We used simulation studies (Kadlec et al., 2023) to determine how many trials would be necessary to achieve a test/re-test reliability of 0.8. We then either administered the same version of a task multiple times (CCMT, SCAP), administered extended task versions with novel stimuli (GFMT, Emotion Matching, Emotion Labeling) or administered additional versions of the task with novel stimuli (MST, PGNG, RISE, PIM, FMP, CFMT, VET). For these tasks, trials were concatenated across

both versions and used as though they were collected in one sitting. However, if a participant had one version of a task (for example, if they completed the PGNG in the original battery but did not complete it or had their data excluded in one of the datasets), the data from the single version would be used. Similarly, we included participants who completed 2/3 forms of the CFMT.

For tasks where one version of a task (MST, N-back, digit span) or a single extended version (GFMT) contained a sufficient number of trials, only a single version of the task was included in the battery. We also included two additional tasks to the battery – the Vanderbilt Expertise Task and a fractal N-back task.

In this replication dataset, each task was advertised individually; a total of N=183 returned and completed at least one task in the replication dataset; N=41 successfully completed all tasks in the replication dataset.

Inclusion/exclusion criteria

We first excluded data on a per-task basis, depending on a combination of accuracy, reaction time (RT) and individual trial responses. Data was excluded from a task if accuracy was greater than 3SD below the mean and one of the following criteria were met:

- Average RT was 3 standard deviations (SD) faster than the group mean
- Average sequence length of a single train of responses was 3SD greater than the group mean sequence length.

If a participant's accuracy was less than 0.5SD below the mean, they were included regardless of their RT and individual trial responses. The only exception to this was the MST, where if the standard deviation of the RT was less than 2 SD below the SD of the group, they were

excluded regardless of performance, as this pattern of responses suggested that the task was being performed by a script/bot rather than a human.

After excluding participants based on accuracy, RT, and individual trial responses, we additionally excluded participants from specific tasks based on the following criteria, which would indicate that they were not paying attention:

- CFMT: Four or more incorrect trials in Stage 1 (i.e., Stage 1 score less than 83%). For participants with three incorrect trials in Stage 1, data were excluded if performance on the other stages indicated a lack of attention rather than a valid measure of poor performance.
- FMP: Accuracy below chance (50%) in face-matching trials.
- PGNG: Accuracy less than 3 SD below the mean for the two target identification stages (baseline level). If accuracy was less than 2 SD below the mean, performance on the rest of the task was evaluated to determine whether low accuracy was because of a genuine low performance or lack of attention.
- N-Back: Accuracy less than 3 SD below the mean for the 1-back blocks. If accuracy was less than 2 SD below the mean, performance on the rest of the task was evaluated to determine whether low accuracy was because of a genuine low performance or lack of attention.
- VET tasks: Incorrect or missing responses on 2 out of 3 catch trials.

Once data was excluded at the task level, we excluded participants from the battery entirely if they were missing either of our outcome measures (the CFMT or the PIM task). Additionally, we excluded participants who were completely missing more than 4 tasks across both sessions.

After applying all inclusion/exclusion criteria, we were left with a final sample of N = 101 (51 female, 50 male, mean age: 31.04 +/- 8.43).

Data Preprocessing:

PIM Task

Items where fewer than 70% of all subjects completed a version of the PIM (i.e. not just the sample we describe below), or 75% of the subjects who scored above the 75th percentile overall, responded correctly were removed, as these items were likely uninformative. This threshold resulted in 3 target items being removed from the original version of the PIM and 2 target items being removed from the replication version of the PIM.

For the free response accuracy score, the cosine similarity score between the semantic embeddings of the correct and incorrect responses from word2vec were calculated. Any responses with a cosine similarity of above 0.55 for recall of the country and 0.6 for the code name were given half credit.

Multiple Imputation

After the application of our exclusion criteria, 2.3% of the final dataset was missing. Missing data was imputed using multiple imputation with the *mice* R package (Buuren & Groothuis-Oudshoorn, 2011). We used 5 imputations with the predictive mean matching method and averaged across the iterations to create a complete dataset. Descriptive statistics for each task prior to imputation, including the amount of missing data and the standard deviation of the imputed values, are provided in Table 3.1.

Data Analysis

Validation of PIM task

To investigate the construct validity of the PIM task, we took the Pearson correlation of the outcome measures from the PIM task (raw recognition accuracy score, overall multiple choice accuracy) and overall accuracy on the CFMT, GFMT, and PI20. Additionally, we computed the Pearson correlation between the PIM – Recognition Accuracy and PIM – Overall MC scores.

Suitability for Exploratory Factor Analysis (EFA)

Our next goal was to investigate the underlying latent structure of our task battery. To that end, we used exploratory factor analysis (EFA), which seeks to identify latent constructs that explain shared variance across tasks (Fabrigar et al., 1999). To determine whether our data were appropriate for EFA, we used the Bartlett's Test of Sphericity (Bartlett, 1950) and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1970; Kaiser & Rice, 1974) using the R *psych* package (Revelle, 2023).

Bartlett's Test of Sphericity tests whether the correlation matrix is significantly different from an identity matrix – that is, where there is any correlation between the variables. A significant value in this test indicates that there are correlations between the variables, suggesting that the data may be appropriate for a factor analysis.

Testing whether the matrix is an identity matrix (i.e. that all off-diagonal correlations are not different from zero) is not necessarily a good indicator as to whether the data can be explained by a factor structure, as the presence of significant correlations does not necessarily indicate the existence of latent constructs. The KMO test determines whether a latent factor structure may

explain the data by measuring the strength of the partial correlations of a variable, relative to the original correlations. The KMO value is defined as

$$KMO = \frac{\sum\sum_{j\neq k} r_{jk}^2}{\sum\sum_{j\neq k} r_{jk}^2 + \sum\sum_{j\neq k} p_{jk}^2}$$

Where $r_{jk}$is the correlation between two variables and p_{jk} is the partial correlation. When variables distinctly load onto latent factors, they will be strongly correlated but the partial correlations should be small, leading to a KMO value of close to 1.0. According to Kaiser (1970), a KMO value of >0.9 is "marvelous," between 0.8-0.9 is "meritorious," between 0.7-0.8 is "middling," between 0.6-0.7 is "mediocre," between 0.5-0.6 is "miserable" and below 0.5 (when the partial correlation matrix would be equivalent to the correlation matrix), is considered unacceptable. The KMO test also provides a single measure of sampling adequacy (MSA) for each individual variable, which can be interpreted using the same guidelines as the overall KMO statistic and calculated as:

$$MSA_j = \frac{\sum_{j\neq k} r_{jk}^2}{\sum_{j\neq k} r_{jk}^2 + \sum_{j\neq k} p_{jk}^2}$$

Number of factors to extract

We used several methods to determine the number of factors to extract. First, we used Horn's parallel analysis (Horn, 1965), which generates synthetic shuffled data (i.e. data where there is no underlying factor structure) with the same dimensions as the original dataset and iteratively extracts the eigenvalues for an increasing number of factors. This process was repeated 100 times and the 95th percentile of the saved eigenvalues is computed and compared to the eigenvalues from the true data. The crossover point where the eigenvalues of the permuted data are greater than those of the true data indicate the maximum number of factors to extract.

We additionally used Kaiser's rule (Kaiser, 1960), where we computed the eigenvalues of the correlation matrix and determined how many eigenvalues were greater than 1, which would indicate the maximum number of factors to extract. Although this threshold is arbitrary (Howard, 2016), it provides a useful upper bound for the number of factors to consider.

Once we established that our data were factorable and identified the lower and upper limits on potential number of factors to extract from Horn's parallel analysis and Kaiser's rule, we used Exploratory Factor Analysis (EFA) with maximum likelihood (ML) estimation (De Winter & Dodou, 2012) and an orthogonal varimax (Kaiser, 1958) rotation on our sample to test whether there was a latent cognitive structure underlying our measured task variables. Although the latent cognitive processes that are reflected by the extracted factors are unlikely to be truly independent, downstream regression analyses (described below) would be negatively impacted by correlation between factors and as such, necessitated an orthogonal rotation.

Evaluating Factor Structure

As has been done in prior literature (Schöttner et al., 2023), we further examined each potential factor structure using 3 criteria: (1) *model fit indices* and (2) *robustness*, (3) *interpretability*. We used multiple converging indicators to measure *model fit*. First, we investigated the total percent of variance in our measured data explained by each factor model, which should be maximized while still ensuring the other two criteria. According to established guidelines (Hu & Bentler, 1999), we additionally used the Tucker-Lewis Index (TLI) (Bentler & Bonett, 1980; Tucker & Lewis, 1973), root mean square of residuals (RMSR) (Bentler, 1995), root mean square error of approximation (RMSEA) (Steiger & Lind, 1980) as well as the 90% confidence interval of the RMSEA.

TLI is an incremental fit index, which compares the fit of a given model to a baseline model where all observed variables are uncorrelated and takes the complexity of the model into account. A TLI of greater than 0.95 is considered a good fit, while above 0.9 is considered an acceptable fit.

RMSR and RMSEA, on the other hand, are absolute fit measures, which compare a model specified fit to the sample data. RMSR is a residual-based fit index, which compares the observed covariance matrix to the covariance matrix that is specified by a hypothesized model. A smaller RMSR indicates a better fit, with an RMSR < 0.05 being considered a good fit and RMSR < 0.08 considered acceptable.

RMSEA indicates the fit of covariance structure indicated by the hypothesized model to that of the observed covariance matrix, per degree of freedom. As with RMSR, a smaller RMSEA indicates a better fit, with RMSEA < 0.06 with a lower value of the confidence interval being close to 0 and upper value being less than 0.08 being considered a good fit, and an RMSEA < 0.08 suggesting an acceptable fit.

To measure *robustness*, we performed the following subsampling procedure. First, we sampled 80% of our full sample without replacement and extracted and rotated factors following the procedure from our main factor analysis, saving the rotated factor loadings. We then used the Hungarian algorithm (Kuhn, 1955), implemented using the *RcppHungarian* package in R (Silverman, 2022) to re-order the factors in each solution to align with the first subsampled factor structure and ensure that we could compare equivalent factors across permutations. We repeated this process 1000 times and then took the average absolute correlation of the factor loadings for a given factor across all iterations. Here, a higher correlation value reflects a more stable clustering

solution across randomly permuted sub-samples of our data. This process was repeated 1000 times for structures with 2-5 factors.

Finally, to assess *interpretability*, we considered measured variables with a factor loading of greater than 0.4, which would indicate significance both practically and statistically (Norman & Streiner, 2014; Watkins, 2018). Because we used an orthogonal rotation, we can interpret the factor loadings as the correlation between the underlying factor and the measured variable, with the squared factor loading reflecting the percent of variance in the factor explained by a given variable. As such, a measured variable with a loading of greater than 0.4 would explain at least 16% of the variance in a factor. We sought to optimize interpretability by preferring factor structures where factors had at least 2 and preferably 3 variables with loadings of greater than 0.4 (Hair et al., 2009), where there was a simple structure such that variables load on only one factor (Thurstone, 1947) and where factors had an internal consistency (calculated as Cronbach's alpha for the salient factor loadings) of > 0.7 (Watkins, 2018).

Prediction of Outcome Measures

Once we established the best factor structure for our data, we extracted Bartlett factor scores for each participant for each factor. Factor scores extracted using this method are only impacted by the shared factors and minimize the influence of the unique variance of the set of variables, resulting in factor scores that are only correlated to their associated factor and not to the other factors. Moreover, this method produces unbiased estimates of the factor score parameters (DiStefano et al., 2009). We used these factor scores as predictors in a regression, along with covariates of age and gender, to predict average CFMT score, PIM recognition accuracy and PIM overall multiple choice accuracy. Regressions were standardized using the *QuantPsyc* R package

(Fletcher, 2022) to compare the relative strength of predictors across outcome variables. Adjusted R^2 was used to compare the percent of variance that was explained for each outcome measure.

Clustering Analysis

To complement our analyses of what predicts successful face memory performance, we performed k-means clustering on a sample of our data that was constrained to have a relatively narrow range on our outcome measures. Restricting the range allowed us to control for variance in performance and better investigate whether there were multiple potential strategies used to accomplish our tasks.

To that end, we created two partially overlapping subsets of participants by selecting participants who were within 0.75SD of the mean CFMT score (N = 56) and those who were within 0.75SD of the mean on the PIM recognition score (N=49). For each sample, we used the *NbClust* function from the *NbClust* R package (Charrad et al., 2014), to compare solutions from 2-4 clusters using the Calinski and Harabasz (CH) index (Calinski & Harabasz, 1974), Duda index (Duda & Hart, 1974), CIndex (Hubert & Levin, 1976), Gamma Index (F. B. Baker & Hubert, 1975) and Beale Index (Beale, 1969), which were the top 5 performing indexes in simulation studies comparing methods to best recover factor solutions (see Charrad et al., 2014 and Milligan & Cooper, 1985 for a review of the methods).

Additionally, we measured the stability of the clustering solutions using the *clusterboot* function from the *fpc* R package (Hennig, 2023), which resamples the clustering data 100 times through a bootstrapping procedure and performs clustering on each bootstrapped sample. Bootstrapped cluster solutions were compared to the original sample using the Jaccard similarity index, which measures the proportion of the common items in both sets relative to the union of all

potential items in both sets. Jaccard similarity values were averaged over bootstrapping iterations to determine stability of a cluster; a stable cluster should show an average Jaccard similarity of greater than 0.75, with values above 0.85 considered highly stable. A Jaccard value between 0.6-0.75 suggests that clusters may be present in data, but the assignment of individual data points to a cluster may be unreliable and values below 0.6 indicate an unreliable cluster (Hennig, 2007).

Once we identified the appropriate number of clusters in each subset of our data, we performed k-means clustering (Hartigan & Wong, 1979) using a Euclidean distance function and with 25 random starts on the extracted factor scores in each of our subsets of data. We used a two-tailed independent samples t-test with unequal variance across groups on the outcome measures for each group to validate that the groups resulting from the clustering were matched on their respective outcome measures – although it was our goal to have groups that were matched on their outcome measure, this analysis did not necessarily guarantee this outcome because the k-means clustering algorithm did not include outcome measures. Average factor scores for each group were computed and compared across subsets of data to investigate whether these analyses identified similar cognitive profiles in the two samples.

Given that the k-means clustering analysis is designed to maximize differences between clusters, it would be circular to attempt statistical analyses comparing the distributions of factor scores between groups. To statistically determine whether there was a trade-off between the Face Perception and Working Memory factors, as might be suggested from the clustering analyses, we returned to the full sample and added an interaction term between the Face Perception and Working Memory factors to each of the regressions on our outcome variables. We then identified the Johnson-Neyman regions of significance (Johnson & Fay, 1950; Johnson & Neyman, 1936) to probe the specific nature of each significant interaction.

Results

Validation of PIM Task

We first sought to compare the outcome measures of the Personal Identity Memory (PIM) task (Figure 3.1) to established measures of face memory to ensure adequate construct validity. To that end, we computed the Pearson correlation of between the PIM Recognition score and the PIM – Overall Multiple Choice score and performance on two previously established tasks measuring face memory – the Cambridge Face Memory Task (CFMT) and the Glasgow Face Matching Task (GFMT). Additionally, performance on the two PIM outcome measures were correlated with scaled scores from the 20-item Prosopagnosia Index (PI20), a self-report measure indexing difficulty with face memory in everyday life (Shah et al., 2015); Figure 3.2A. PIM – Recognition score was significantly correlated with performance on the CFMT ($r = 0.620$, $t_{(99)} = 7.897$, $p = 3.99 \times 10^{-12}$), performance on the GFMT ($r = 0.380$, $t_{(99)} = 4.072$, $p = 9.41 \times 10^{-5}$) and PI20 ($r = 0.27$, $t_{(99)} = 2.741$, $p = 7.28 \times 10^{-3}$). PIM – Overall MC Score was significantly correlated with the CFMT ($r = 0.36$, $t_{(99)} = 3.837$, $p = 2.20 \times 10^{-5}$) and the PI20 ($r = 0.22$, $t_{(99)} = 2.226$, $p = 0.0283$), but not the GFMT ($r = -0.047$, $t_{(99)} = -4.682$, $p = 0.641$). Additionally, PIM – Recognition and PIM – Overall MC scores were significantly correlated ($r = 0.56$, $t_{(99)} = 6.721$, $p = 1.16 \times 10^{-9}$).

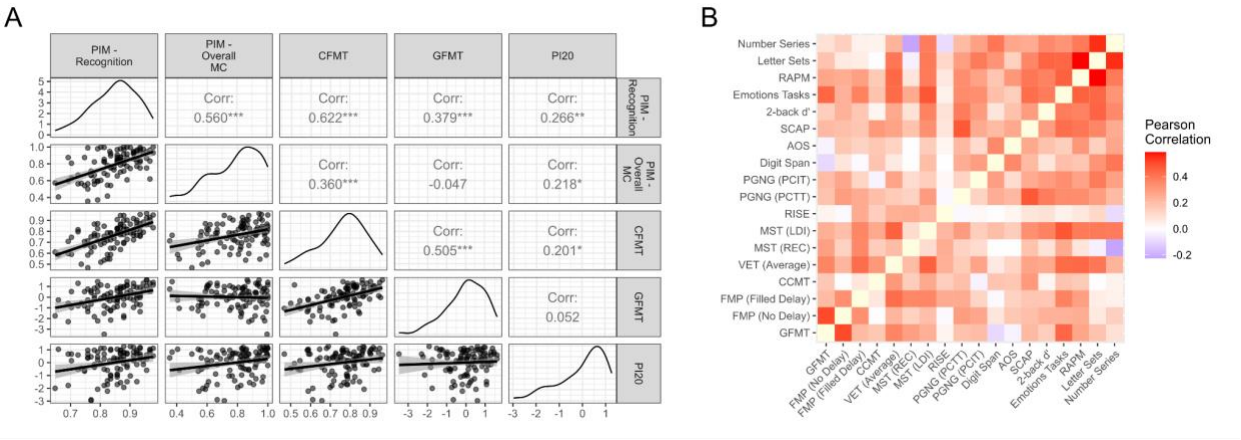


Figure 3.2: Pearson correlations of behavioral tasks. A. Correlations between PIM - Recognition Score and PIM – Overall MC Accuracy with established measures of face processing ability suggest that . Lower triangle depicts scatter plots of variables with linear trend line with each dot representing a subject (n = 101); upper triangle reflects Pearson's correlation value between two variables. Diagonal reflects distribution of each variable. Symbols reflect statistically significant Pearson correlations (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). B. Pearson correlation values between measures from behavioral task battery, after imputation. Moderate correlations across tasks suggest data are appropriate for use in Exploratory Factor Analysis. Abbreviations: PIM: Personal Identity Memory Task; CFMT: Cambridge Face Memory Test; GFMT: Glasgow Face Matching Test; PI20: 20 question Prosopagnosia Index; FMP: Face Memory and Perception Task; CCMT: Cambridge Car Memory Task; VET: Vanderbilt Expertise Task; MST (REC): Mnemonic Similarity Task (Recognition Score); MST (LDI): Mnemonic Similarity Task (Lure Discrimination Index); RISE: Relational and Item Specific Encoding; PGNG (PCTT): Parametric Go/No-Go Task (Percent Correct to Targets); PGNG (PCIT): Parametric Go/No-Go task (Percent Correct to Inhibitory Trials); AOS: Automated Operation Span; SCAP: Spatial Capacity; RAPM: Raven's Advanced Progressive Matrices

Suitability for Exploratory Factor Analysis (EFA)

Both Bartlett's test of Sphericity ($\chi^2_{(153)} = 531.255$, $p = 3.043 \times 10^{-43}$) and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (overall KMO = 0.76) indicated that our data were appropriate for EFA. Descriptive statistics prior to imputation (including percent of scores imputed and standard deviation of imputed scores across iterations, where appropriate) and MSA for individual measures are reported in Table 3.1. Additionally, Figure 3.2B visualizes the correlations between all variables used in the EFA.

Construct	Task Name	n	Percent Imputed	SD of Imputed Values	Mean	Standard Deviation	MSA
Outcomes	CFMT	101	0		0.765	0.112	
	PIM - Recognition Task	101	0		0.848	0.076	
	PIM - Overall MC	101	0		0.79	0.161	
Face Processing	GFMT	97	0.04	0.064	0.879	0.077	0.642
	FMP - No Delay	96	0.05	0.058	0.876	0.071	0.612
	FMP - Filled Delay	96	0.05	0.112	0.661	0.136	0.652
Object Processing	CCMT	101	0		0.702	0.132	0.574
	VET - Birds	99	0.02	0.132	0.699	0.165	
	VET - Planes	97	0.04	0.11	0.668	0.136	
	VET - Leaves	101	0		0.608	0.105	
	VET - Average	95			0.66	0.11	0.836
	MST - REC	98	0.03	0.079	0.803	0.132	0.699
	MST - LDI	98	0.03	0.196	0.378	0.173	0.867
General Associative Memory	RISE	101	0		0.779	0.086	0.589
Attention	PGNG - PCTT	100	0.01	0.142	0.882	0.087	0.797
	PGNG - PCIT	100	0.01	0.155	0.739	0.181	0.805
Working Memory	Digit Span	98	0.03	1.264	7.041	1.45	0.613
	AOS	101	0		0.517	0.223	0.733
	SCAP	101	0		0.869	0.086	0.838
	2-back d'	97	0.04	0.844	3.118	0.776	0.764
Social Cognition	Emotion Tasks	101	0		0.693	0.078	0.866
General/Fluid Intelligence	RAPM	100	0.01	0.144	0.403	0.185	0.849
	Letter Sets	100	0.01	0.121	0.607	0.191	0.753
	Number Series	93	0.079	3.24	11.892	2.672	0.84

Table 3.1: Descriptive statistics for battery of tasks. Tasks are grouped by hypothesized cognitive constructs. Missing values in the SD of Imputation column reflect no imputation was needed. VET – Average statistics calculated prior to imputation. MSA values were not included for individual VET tasks and outcome measures as they were not included in the EFA. Abbreviations: PIM: Personal Identity Memory Task; CFMT: Cambridge Face Memory Test; GFMT: Glasgow Face Matching Test; PI20: 20 question Prosopagnosia Index; FMP: Face Memory and Perception Task; CCMT: Cambridge Car Memory Task; VET: Vanderbilt Expertise Task; MST (REC): Mnemonic Similarity Task (Recognition Score); MST (LDI): Mnemonic Similarity Task (Lure Discrimination Index); RISE: Relational and Item Specific Encoding; PGNG (PCTT): Parametric Go/No-Go Task (Percent Correct to Targets); PGNG (PCIT): Parametric Go/No-Go task (Percent Correct to Inhibitory Trials); AOS: Automated Operation Span; SCAP: Spatial Capacity; RAPM: Raven's Advanced Progressive Matrices

Dimensionality of Latent Space

Kaiser's rule and Horn's parallel analysis (Figure 3.3A) suggested that 2-5 factors would be appropriate to investigate for our data. Fit measures for each model are reported in Table 3.2 and Figure 3.3B. Robustness results (Figure 3.3C) across suggested that both the 2-factor structure was most stable, followed by the 4-factor structure. The 3-factor and 5-factor solutions were least stable, as they both resulted in factors that showed average absolute correlations of less than 0.6.

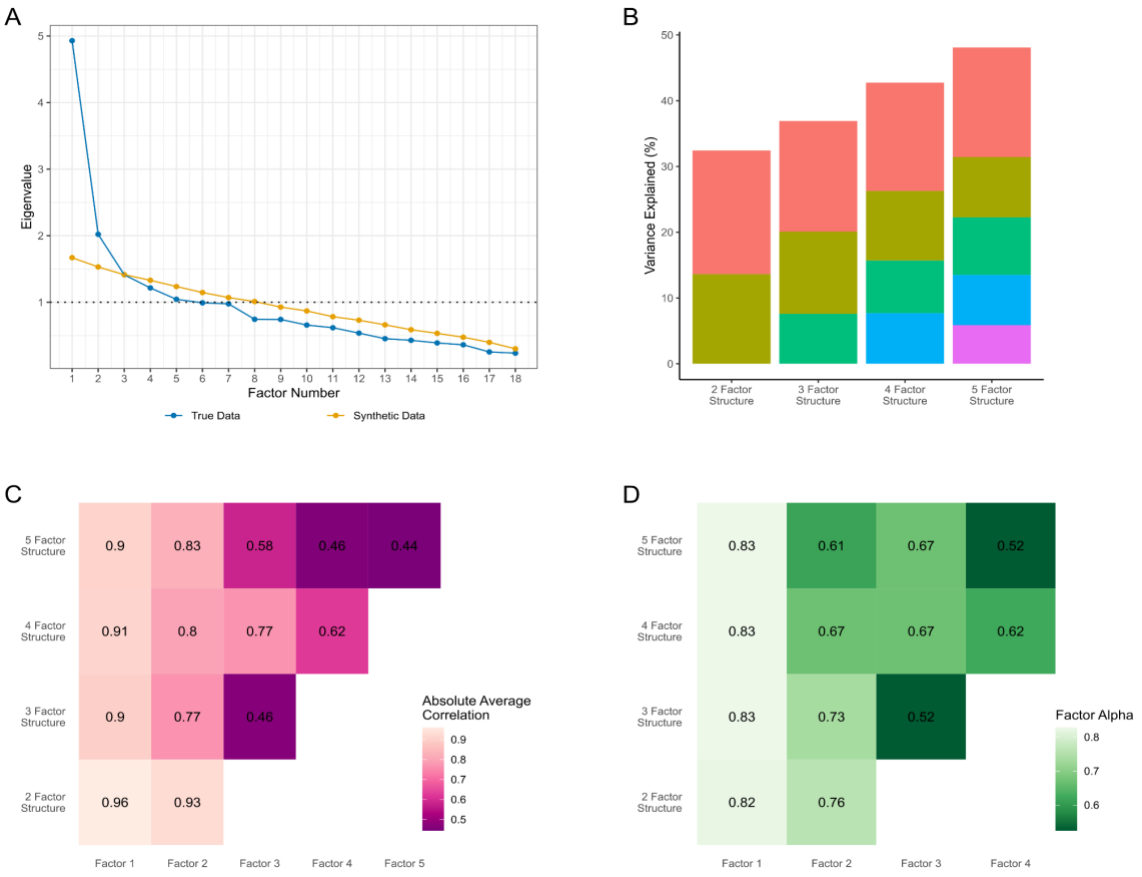


Figure 3.3: Determining dimensionality of latent space. A. Horn's parallel analysis results showing the eigenvalues of the factors extracted from the true data (blue) and synthetic data (orange). B. Model fit, as measured by the total variance explained by each potential factor structure. Colors within each bar reflect the variance explained by individual factors, although note that the factors do not necessarily reflect the same construct across structures (i.e. Factor 1 in the 2-factor structure does not necessarily reflect the exact same construct as Factor 1 in the 5-factor structure). C. Factor robustness, as measured by the average absolute correlation across 100 subsamples that each selects 80% of the sample. Rows reflect the factor structure (i.e. how many factors were extracted) and columns reflect the individual factor (matched across subsamples via the Hungarian algorithm). As with panel B, note that the ordering of factors across structures is not necessarily the same. D. Internal reliability of each factor, as measured by Cronbach's alpha across the salient variables for each factor. Note that Factor 5 in the 5 factor solution only had 1 salient variable, so alpha could not be calculated.

In terms of interpretability, no factor solution showed a simple solution where each variable only loaded on a single factor, and the 5-factor solution did not meet our criteria that all factors had at least 2 salient variables with loadings of greater than 0.4. Additionally, only the 2-factor solution showed sufficient internal reliability of all extracted factors; all others showed at least one factor with an internal reliability of less than 0.7, with the 3 and 5 factor showing internal reliability values close to 0.5 (Figure 3.3D).

Model	Variance Explained	Chi Squared (df)	TLI	RMSEA (90% CI)	SRMSR
2-factor model	32.40%	166.528 (118)	0.841	0.063 (0.039 0.086)	0.073
3-factor model	36.90%	135.713 (102)	0.871	0.056 (0.027 0.082)	0.065
4-factor model	42.70%	103.957 (87)	0.923	0.043 (0 0.073)	0.052
5-factor model	48.10%	74.413 (73)	0.992	0.01 (0 0.06)	0.045

Table 3.2: Model fit indices. Abbreviations: df – degrees of freedom; TLI – Tucker Lewis Index; RMSEA – Root Mean Square Error of Approximation; CI – Confidence Interval; SRMSR – Standardized Root Mean Square of the Residual

Taken together, our results suggested that the 2-factor and 4-factor structures were the most stable and best fitting models. The 4-factor model, while slightly less stable than the 2-factor model, explained more variance and showed better indices of model fit. As such, we only consider the 4-factor model (Figure 3.4) for the remaining analyses. Loadings of the models with 2, 3 and 5 factors are reported in Supplementary Figures 3.1-3.3.

Interpretation of Latent Factors

Before using our latent cognitive factors to predict performance on tasks measuring face memory, we sought to interpret them using the variables that saliently loaded on each factor.

Our first factor had salient positive loadings from Letter Sets, Raven's Advanced Progressive Matrices (RAPM), Number Series, 2-back d', the combined Emotion Labeling/Emotion Matching score, percent correct to inhibitory trials on the Parametric Go/No-

Go task, Lure Discrimination Index (LDI) from the Mnemonic Similarity task and overall accuracy on the Vanderbilt Expertise Tasks (VET). Taken together, we interpreted this factor as a General Intelligence factor (Figure 3.4A).

Our second factor had salient positive loadings from the Face Memory and Perception Task (FMP) – Filled Delay score, Recognition Score from the Mnemonic Similarity Task, overall accuracy on the VET and combined Emotion Labeling/Emotion Matching score. Given that the tasks that load on this factor do not all fall in a single domain, we interpreted this factor as an Episodic Long-term Memory factor. The only task that did not fit this interpretation was the combined Emotion Matching/Emotion Labeling score. (Figure 3.4B).

Our third factor had salient positive loadings on the Glasgow Face Matching Task (GFMT) and the FMP – no delay condition. Together, this suggests that the factor reflects Face Perception (Figure 3.4C).

Our final factor shows significant positive loadings on the Automated Operation Span score, Digit Span, percent correct to target on the Parametric Go/No-Go task and performance on the Spatial Capacity Working Memory task. Together, we interpreted this factor as a Working Memory factor (Figure 3.4D).

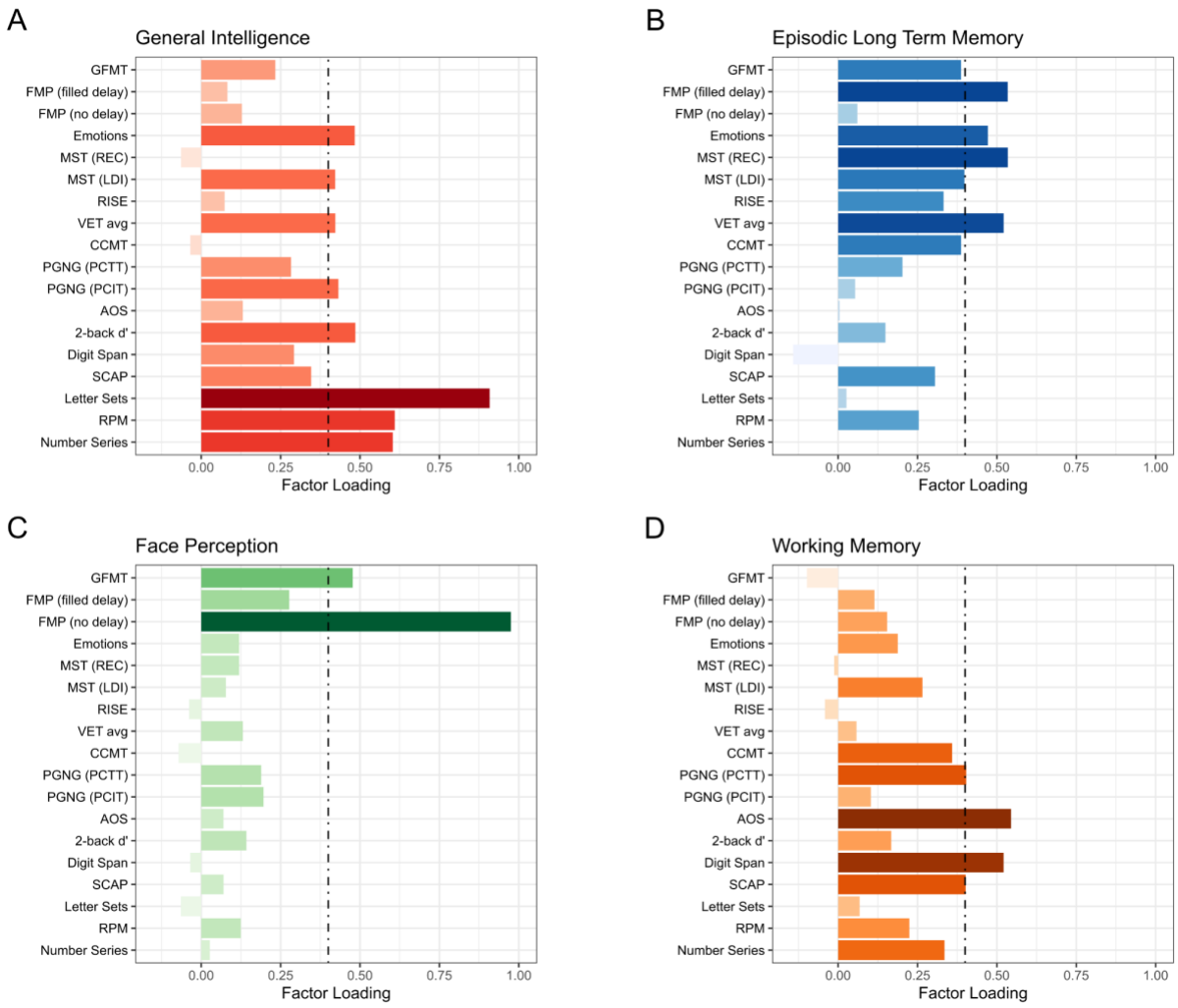


Figure 3.4: Latent factor structure. Factor loadings from the 4-factor solution using N = 101 subjects. Factors are ordered by their explained variance. Dotted line reflects a loading of 0.4, which was our threshold for saliency. Abbreviations: PIM: Personal Identity Memory Task; CFMT: Cambridge Face Memory Test; GFMT: Glasgow Face Matching Test; PI20: 20 question Prosopagnosia Index; FMP: Face Memory and Perception Task; CCMT: Cambridge Car Memory Task; VET: Vanderbilt Expertise Task; MST (REC): Mnemonic Similarity Task (Recognition Score); MST (LDI): Mnemonic Similarity Task (Lure Discrimination Index); RISE: Relational and Item Specific Encoding; PGNG (PCTT): Parametric Go/No-Go Task (Percent Correct to Targets); PGNG (PCIT): Parametric Go/No-Go task (Percent Correct to Inhibitory Trials); AOS: Automated Operation Span; SCAP: Spatial Capacity; RAPM: Raven's Advanced Progressive Matrices

Predicting Face Memory Performance

Once interpretable factors were established, we used multiple linear regression to determine how well they could explain variance in face memory performance (as measured by performance on the CFMT, PIM – Recognition Task and PIM – Overall MC Score), in addition to demographic measures of age and gender.

Our factors could significantly predict performance on the CFMT ($F_{(6, 94)} = 10.44$, $p = 7.535 \times 10^{-9}$, adjusted $R^2 = 0.362$), PIM – Recognition Task ($F_{(6, 94)} = 5.757$, $p = 3.89210^{-5}$, adjusted R2 = 0.222) and PIM – Overall MC Score ($F_{(6, 94)} = 2.972$, $p = 0.011^{9}$, adjusted $R^2 = 0.1058$).

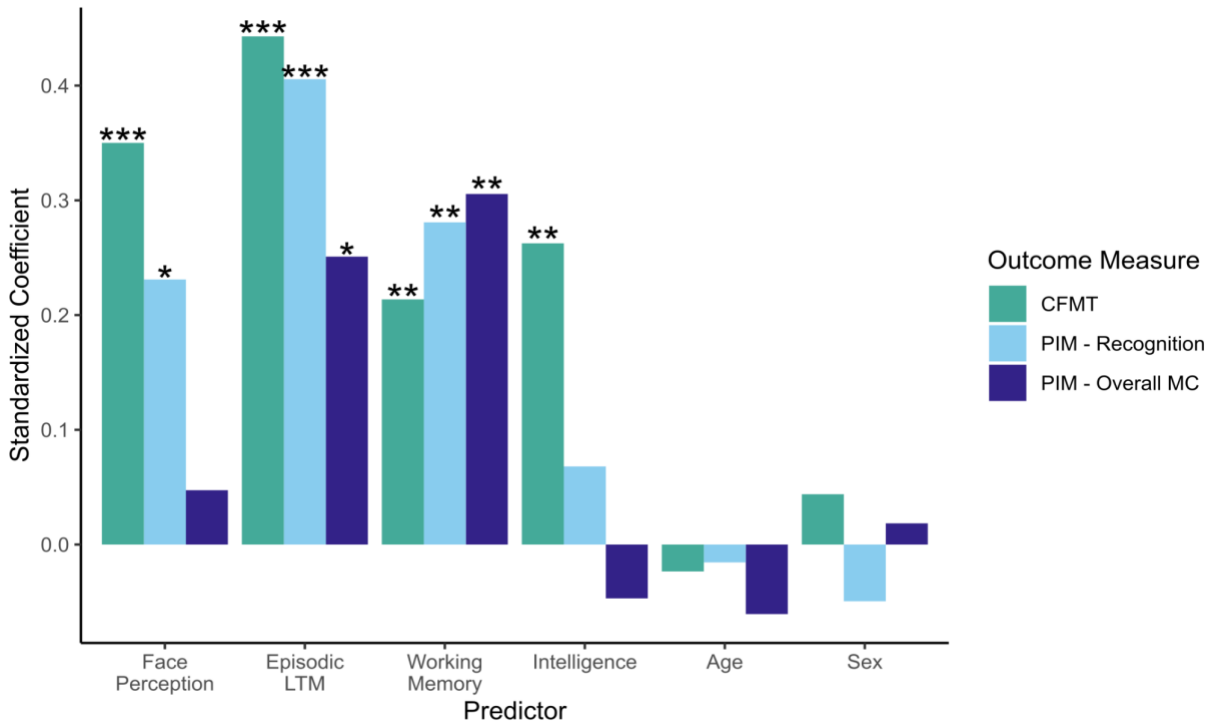


Figure 3.5: Explaining measures of face memory. Standardized coefficients from each regression using N = 101 participants using the four latent factors (in addition to covariates of age and sex) to predict performance on the Cambridge Face Memory Test (green), PIM - Recognition Score (light blue), and PIM - Overall Multiple Choice Score. Symbols reflect statistical significance of coefficients (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Abbreviations: LTM: Long-term memory

Despite being able to explain significant amounts of variance in each of our three outcome measures, there were different patterns of factor scores that significantly predicted each outcome measure (Tables 3.3-3.5; Figure 3.5). While all factor scores significantly predicted performance on the CFMT, only the Face Processing, Episodic Long-Term Memory and Working Memory factors significantly predicted performance on the PIM – Recognition Task. In contrast, only the Episodic Long-Term Memory and Working Memory factors predicted PIM – Overall MC Scores.

Effect	**Estimate**	**SE**	**95% CI**	**t statistic**	**p**
Intercept	0.107	0.355	(-0.598 0.812)	0.301	0.764
Face Perception	0.229	0.089	(0.053 0.405)	2.582	0.011
Episodic LTM	0.345	0.077	(0.193 0.497)	4.509	1.9x10-5
Working Memory	0.218	0.069	(0.081 0.356)	3.153	0.002
Intelligence	0.063	0.083	(-0.102 0.227)	0.759	0.45
Sex	-0.098	0.179	(-0.454 0.257)	-0.549	0.584
Age	-0.002	0.011	(-0.023 0.019)	-0.174	0.863

Table 3.3: Coefficients for regression model predicting PIM - Recognition Score. Abbreviations: LTM: Long-term memory

Effect	**Estimate**	**SE**	**95% CI**	**t statistic**	**p**
Intercept	0.042	0.321	(-0.596 0.681)	0.131	0.896
Face Perception	0.347	0.08	(0.188 0.507)	4.32	3.87x10-5
Episodic LTM	0.377	0.069	(0.239 0.515)	5.435	4.31x10-7
Working Memory	0.166	0.063	(0.042 0.291)	2.648	0.009
Intelligence	0.243	0.075	(0.094 0.392)	3.233	0.002
Sex	0.087	0.162	(-0.235 0.409)	0.538	0.592
Age	-0.003	0.01	(-0.022 0.016)	-0.289	0.774

Table 3.4: Coefficients for regression model predicting CFMT. Abbreviations: LTM: Long-term memory

Effect	Estimate	SE	95% CI	t statistic	p
Intercept	0.205	0.38	(-0.551 0.96)	0.538	0.592
Face Perception	0.047	0.095	(-0.142 0.236)	0.494	0.623
Episodic LTM	0.214	0.082	(0.051 0.376)	2.602	0.011
Working Memory	0.237	0.074	(0.09 0.385)	3.2	0.002
Intelligence	-0.043	0.089	(-0.22 0.133)	-0.488	0.627
Sex	0.037	0.192	(-0.345 0.418)	0.191	0.849
Age	-0.007	0.011	(-0.03 0.015)	-0.631	0.530

Table 3.5: Coefficients for regression model predicting PIM - Overall Multiple Choice Score. Abbreviations: LTM: Long-term memory

Clustering Analysis

So far, we have used multiple linear regression to show that different cognitive processes linearly predict performance on face memory tasks. However, an alternative approach is to select a sample of participants who are matched on performance and apply an unsupervised clustering algorithm, like k-means clustering, to determine whether there are multiple potential strategies that can support performance. To that end, we applied k-means clustering on the factor scores from two partially-overlapping samples of participants who were within 0.75SD of the mean on the CFMT and on the PIM – Recognition task.

For the sample constrained on the CFMT, indices for determining the number of clusters suggested that a 2-cluster solution was most appropriate (Table 3.6) and bootstrapping suggested that the two clusters were stable (mean Jaccard values of 0.846 and 0.762 over 100 bootstrapped iterations).

Solution	CH	Duda	Gamma	Beale	CIndex	Jaccard - Cluster 1	Jaccard - Cluster 2	Jaccard - Cluster 3	Jaccard - Cluster 4
2-Cluster	*18.496*	*1.070*	0.558	*-0.148*	0.381	*0.846*	*0.762*		
3-Cluster	17.745	1.327	*0.677*	-0.510	*0.362*	0.739	0.476	0.642	
4-Cluster	16.842	2.157	0.574	-1.036	0.437	0.736	0.581	0.596	0.719

Table 3.6: Clustering solution indices from CFMT constrained sample. Rows reflect clustering solutions, columns reflect individual indices. Italics indicate best solution according to a given index. Note that for the Jaccard similarity values, the clusters are not necessarily the same across solutions. Data comes from N = 56 who were within 0.75 standard deviations of the mean on the CFMT.

Outcome scores did not differ across clusters (CFMT: $t_{(53.248)} = 1.176$, p=0.245; PIM – Recognition Task: $t_{(42.708)} = 0.549$, p = 0.586; PIM – Overall MC Score: $t_{(44.65)} = 0.126$, p = 0.900) (Figure 3.6A). Cluster means revealed two distinct cognitive profiles: one with above average scores on the Face Processing and Episodic Long-Term Memory factors with below average scores on the Working Memory factor, and one with the opposite pattern of below average scores on the Face Memory and Episodic Long-Term Memory and above average scores on the Intelligence and Working Memory factors (Figure 3.6B).

For the sample constrained on the PIM – Recognition Task, 3 indices suggested that a 2-cluster solution was best, while 2 suggested a 4-cluster solution was best (Table 3.7). Bootstrapping suggested that for the 2-cluster solution, both clusters were stable (mean Jaccard values = 0.748 and 0.810 over 100 bootstrapped iterations). In contrast, the 4-cluster solution showed two clusters that were stable (average Jaccard values = 0.759 and 0.763), one was likely a true cluster but the assignment of individual data points was unreliable (average Jaccard value = 0.664) and one cluster was not stable (average Jaccard value = 0.565). Taken together, we opted to focus on the 2-cluster solution for the sample constrained on PIM – Recognition Score.

Solution	**CH**	**Duda**	**Gamma**	**Beale**	**CIndex**	**Jaccard - Cluster 1**	**Jaccard - Cluster 2**	**Jaccard - Cluster 3**	**Jaccard - Cluster 4**
2-Cluster	*16.256*	*0.887*	0.434	*0.296*	0.445	*0.748*	*0.810*		
3-Cluster	15.543	1.430	0.636	-0.670	0.382	0.581	0.712	0.771	
4-Cluster	14.684	1.035	*0.661*	-0.073	*0.348*	0.664	0.565	0.759	0.763

Table 3.7 Clustering solution indices from PIM - Recognition constrained sample. Rows reflect clustering solutions, columns reflect individual indices. Italics indicate the best solution according to a given index. Note that for the Jaccard similarity values, the clusters are not necessarily the same across solutions. Data comes from N = 49 who were within 0.75 standard deviations of the mean on the PIM - Recognition.

Just as with the sample constrained on the CFMT, there were no differences across group on the outcome measures across clusters (CFMT: $t_{(37.372)} = 0.319$, p = 0.418; PIM – Recognition

Task: $t_{(41.792)} = 0.6395$, $p = 0.526$; PIM – Overall MC Score: $t_{(36.621)} = 0.0720$, $p = 0.943$; Figure 3.6A). There was also a similar pattern of cluster means across the groups: one group with above average Face Processing and Episodic Long-Term Memory and below average Intelligence and Working Memory, with the other group showing the opposite pattern (Figure 3.6B).

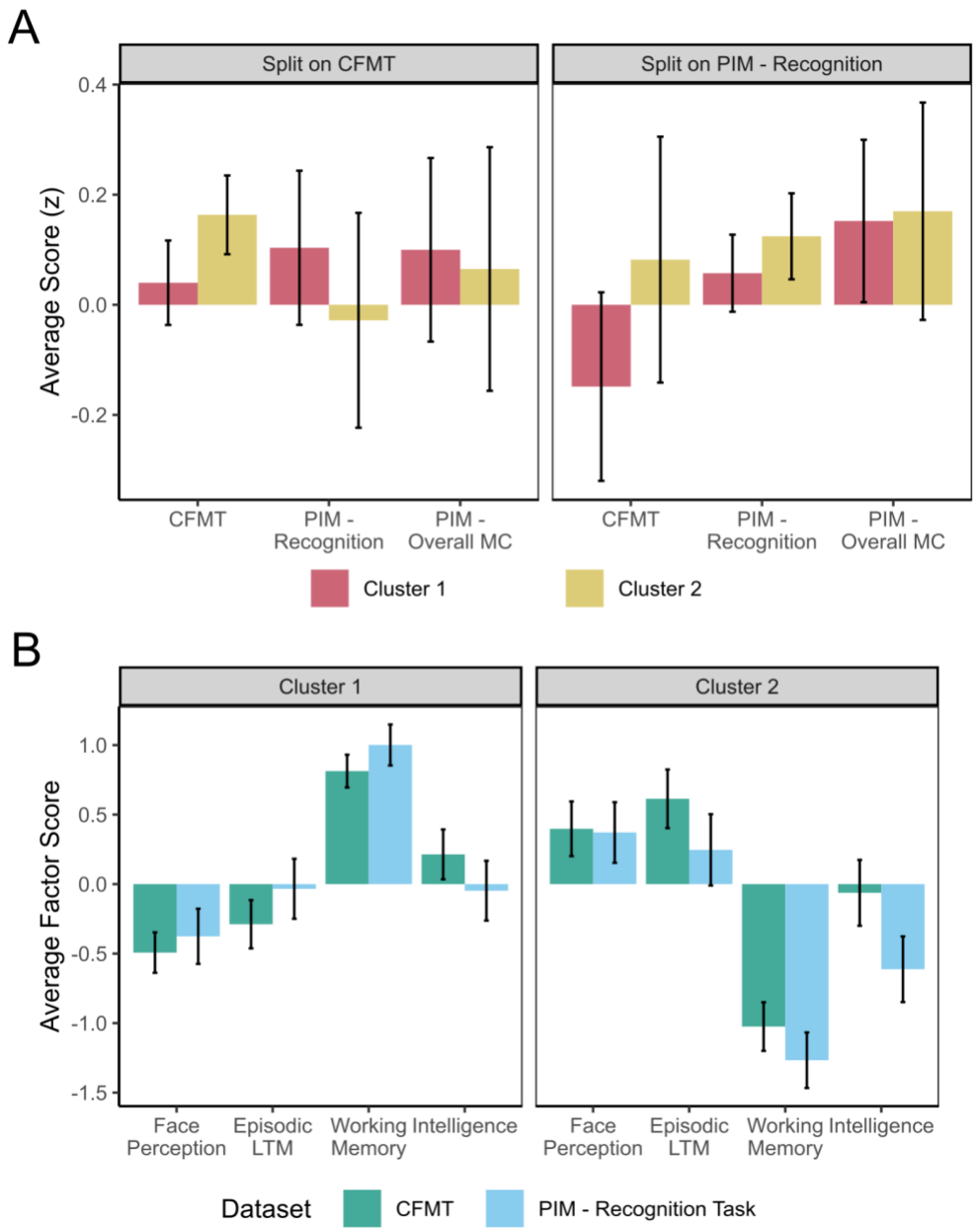


Figure 3.6: Cognitive profiles using constrained datasets. A. Outcome measures from the clusters in the 2-cluster solutions for each of the two subsampled datasets reveal that there were no significant differences in outcome measures across clusters. Left panel shows data from CFMT constrained dataset (N = 56); right panel shows data from PIM – Recognition constrained dataset (N = 49). Maroon reflects Cluster 1 and yellow reflects Cluster 2. B. Average factor scores across groups reveal two distinct cognitive profiles - one that is strong on the Working Memory factor and weak on the Face Perception and Episodic Long-Term Memory factors (left panel) and one that showed the opposite pattern of scores (right panel). Results were consistent across the CFMT constrained dataset (green; N = 56) and the PIM - Recognition constrained dataset (blue; N = 49). Error bars reflect standard error of the mean. Abbreviations: Episodic LTM: Episodic Long-Term Memory; CFMT: Cambridge Face Memory Test; PIM: Personal Identity Memory Test; Overall MC: Overall Multiple Choice Score.

Investigating Potential Trade-offs in Functioning

Our clustering results indicated the presence of two cognitive profiles, one which is more reliant on face perception and memory, and one which is more based on working memory. Although the dissociations in profiles may suggest trade-offs of strategy, the clustering algorithm that produced them was designed to maximize differences between groups so further testing of the resulting distributions would be circular. To that end, we returned to our full sample and added an interaction term between the Face Perception and Working Memory factors, as these were the most dissociated in both clustering solutions, to statistically test for an interaction of the factors that might indicate a tradeoff.

Linear regression analyses revealed a significant interaction term for the model predicting the PIM – Recognition Score (β = -0.155, $t_{(93)}$ = 2.157; p = 0.034; Table 3.8). Adding in the interaction term significantly improved the amount of variance explained by the model (Δ adjusted R^2 = 0.0291; $F_{(93,1)}$ = 4.655, p = 0.034).

Effect	**Estimate**	**SE**	**95% CI**	**t statistic**	**p**
Intercept	0.239	0.354	(-0.463 0.941)	0.677	0.5
Face Perception	0.232	0.087	(0.06 0.405)	2.67	0.009
Episodic LTM	0.344	0.075	(0.195 0.493)	4.585	$1.41x10^{-5}$
Working Memory	0.221	0.068	(0.086 0.356)	3.258	0.002
Intelligence	0.049	0.082	(-0.113 0.211)	0.601	0.549
Face Perception x Working Memory	-0.155	0.072	(-0.298 - 0.012)	-2.157	0.034
Age	-0.005	0.011	(-0.026 0.016)	-0.457	0.649
Sex	-0.202	0.182	(-0.564 0.16)	-1.109	0.27

Table 3.8: Regression coefficients for PIM – Recognition, including interaction terms

Decomposing the interaction for the PIM – Recognition score revealed that the effect of the Face Perception factor grew larger with lower Working Memory factor scores (Figure 3.7A).

A similar pattern was shown for the inverse effect – the effect of Working Memory grew larger for those with lower Face Perception factor scores (Figure 3.7B).

In contrast, the interaction term was not significant for the regressions predicting the CFMT (β = -0.0534, $t_{(93)}$ = 0.067; p = 0.425) or the PIM – Overall MC Score (β = 0.0382, $t_{(93)}$ = 0.079; p = 0.630). Full regression table for this analysis is reported in Supplementary Materials.

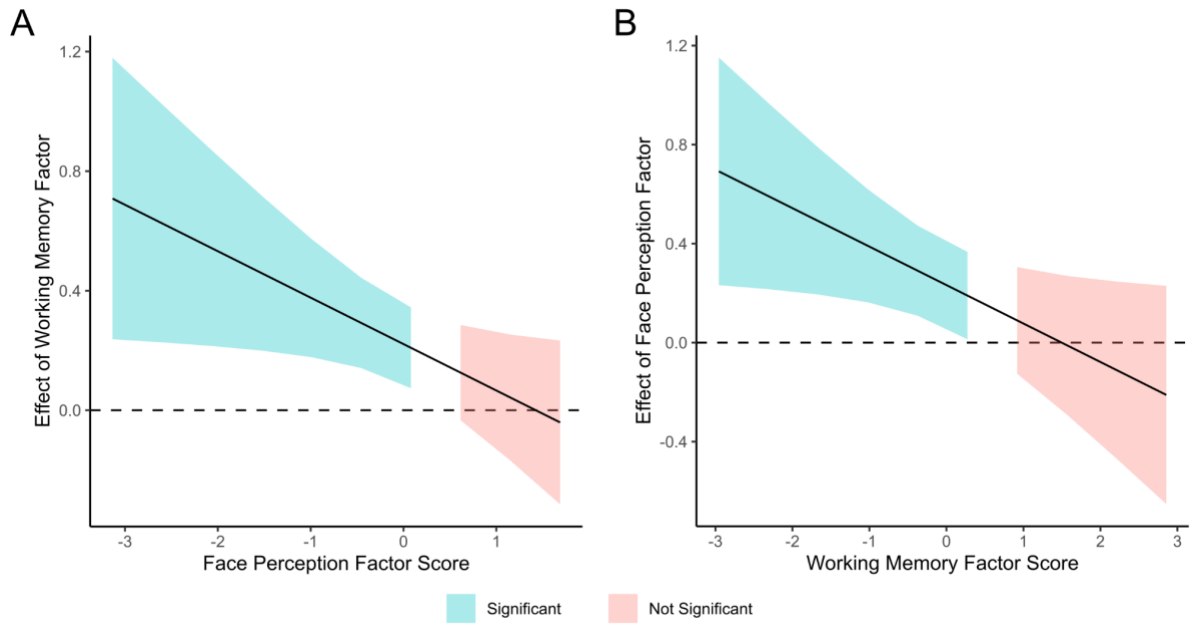


Figure 3.7: Johnson-Neyman regions of significance for the Face Perception x Working Memory factors interaction. Line reflects the simple effect of A. Working Memory, given a score on the Face Perception factor or B. Face Perception, given a score on the Working Memory factor. Turquoise regions reflect a statistically significant interactive effect, red regions reflect a non-significant effect.

Discussion

The goals of the current work were to understand how cognitive processes beyond visual face perception support face memory, and whether multiple combinations of these processes could produce the same level of performance on a face memory task using different strategies.

Before we could systematically investigate the cognitive processes that predict face memory performance, we needed to derive measures for these underlying processes. We used Exploratory Factor Analysis (EFA), a data-driven approach that searches for the common latent constructs that explain shared variance in measured variables (Fabrigar et al., 1999), to uncover four latent cognitive factors. Extant work using EFA on behavioral measures from the Human Connectome Project (Schöttner et al., 2023) produced factors that reflected processing speed and general cognitive processing, including processes like memory, working memory, fluid intelligence, executive functioning and language. Other meta-analytic work analyzing a vast amount of human neuroimaging literature to map brain circuits to function identified more nuanced factors including vision, memory and manipulation (which encompassed working memory, attention, and other executive functioning) (Beam et al., 2021). Our four factors – Intelligence, Episodic Long-Term Memory, Face Perception and Working Memory, are largely in agreement with these findings.

Our dissociation of a specific Face Perception factor from other cognitive factors suggests that face perception may not share variance with these other more domain-general processes. Although there may be some domain-general processing shared between face memory and memory for other visual objects, memory for faces is also largely served by separate, domain-specific face processes (McGugin et al., 2012; Van Gulick et al., 2016). This experimental work in healthy adults is supported by findings that both acquired and developmental prosopagnosics (DPs), who have profoundly impaired face memory, may retain normal object recognition (Barton et al., 2019) (although see Gerlach et al., 2016 for evidence suggesting impairment in object recognition in DPs). Other work has additionally highlighted how face processing ability is distinct

from general intelligence (Connolly et al., 2019; Verhallen et al., 2017; Wilmer et al., 2012, 2014). Both lines of work are consistent with our separate Face Perception factor.

Intriguingly, our factor structure also revealed that the Filled Delay condition of the Face Memory and Perception task, which was designed to tax face memory processes, did not load onto the Face Perception factor. There has been mixed support in the literature for whether face memory and face perception are distinct processes – one study found that the CFMT, a measure of face memory, and the Glasgow Face Matching Task (GFMT), which does not have a memory component, both load onto a single *f* factor (Verhallen et al., 2017). Although one individual differences approach has shown small but significant correlations between face memory and face perception in addition to shared regions that contribute to both processes in DPs, it also highlights distinct regions in the core and extended face processing networks that are selectively associated with either process (Liu et al., 2021). Other work, with both healthy adults (Hacker & Biederman, 2020) and DPs (Dalrymple et al., 2014) has suggested that face perception is largely separate from face memory, echoing the divisions that we found in our latent factors.

After establishing the latent cognitive processes that best explained variance in our task battery, we extracted factor scores that we then used to predict performance on the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006a) and the Personal Identity Memory (PIM) Task. We designed this novel task to be a more ecologically valid experience by introducing identities through dynamic videos that depicted multiple clips of each of our to-be-learned identities in a variety of contexts. Moreover, we paired each with personal semantic information, including a name, a location and personality traits. We indexed face recognition performance on this task using novel photos of each to-be-learned identity in entirely new contexts, much like

occurs outside the lab when meeting a new person. In addition to the face recognition task, the PIM task also probed recall of the associated personal semantic information.

We were able to significantly explain performance on all outcome measures (CFMT performance, PIM – Recognition Task performance and PIM – Overall Multiple Choice score), with the most variance explained on the CFMT. Although there were moderately strong correlations between the CFMT and the PIM measures, different patterns of factor scores predicted each outcome. While all four factors predicted performance on the CFMT, only the Face Perception, Episodic Long-Term Memory and Working Memory factors predicted performance on the PIM – Recognition Task. Moreover, standardizing the regression coefficients in each of our models revealed that our Working Memory factor more strongly predicted the PIM – Recognition score than it did performance on the CFMT.

One possible explanation for why our Intelligence factor predicts performance on the CFMT, but not the PIM – Recognition Task could come from how memory is probed in each task. The CFMT presents identities through relatively few images with only small variation - the same images from the learning phase are used as targets in the first phase, with only relatively minor changes to lighting and pose in the target images for the second and third phases. It is possible that the varied contexts during the dynamic videos from the learning phase of the PIM task allow participants to extract an "average" representation of the to-be-learned faces that can be generalized to new contexts (Burton et al., 2005; de Fockert & Wolfenstein, 2009), while the CFMT relies more on a more basic visual match-to-sample process. Although our Intelligence factor (which only predicts performance on the CFMT and not the PIM – Recognition Task) is mostly comprised of tasks measuring general fluid intelligence like Raven's Advanced Progressive Matrices and Letter Sets, the other tasks that load on the factor include the combined Emotion

Matching/Emotion Labeling score, Vanderbilt Expertise Tasks, and the Lure Discrimination Index from the Mnemonic Similarity Task. These tasks may reflect domain-general visual discrimination processes that could be more useful in the CFMT, where the targets are closer perceptual matches to the learned stimuli than in the PIM task.

Another crucial difference between the CFMT and the PIM – Recognition task is the addition of associated semantic details with each identity in the PIM task. Learning semantic information fundamentally changes how the brain represents identity information – one study that had participant learn unfamiliar faces with or without associated biographical information created a grouping effect, where identities with biographies were more similar to other identities learned with biographies than those learned without the associated information, and vice versa (Verosky et al., 2013). Other work has shown that encoding a face conceptually (i.e. encoding by asking about personality traits associated with the face) improves recognition over encoding a face perceptually (i.e. encoding by asking about the shape of the face) (Schwartz & Yovel, 2018). Together, these findings suggest that associating faces with conceptual information may prompt different mnemonic strategies, such as imagery or elaborative encoding, or facilitates verbal rehearsal, both of which rely on stronger working memory. Notably, our Working Memory factor was mostly comprised of tasks indexing verbal working memory (for example, Digit Span, Automated Operation Span) and as such, may be less useful for the CFMT.

We also show that the PIM – Overall Multiple Choice performance is only predicted by our Episodic Long-Term Memory and Working Memory factors. Although it was surprising that our Face Perception factor did not predict performance on the PIM – Overall Multiple Choice, it is possible that our choice to cue retrieval with images from the learning phase and select visually distinct identities within a block reduced this task to a cued retrieval task that did not rely as heavily

on face perception. That is, one could succeed on this task by binding the semantic information to a general representation of the face (i.e. "the blonde woman"), rather than encoding the specific and nuanced details that would be necessary to distinguish a to-be-learned identity from a lure as was necessary in the Recognition Task. In this manner, the PIM – Overall MC may have successfully encouraged binding of the semantic information together (i.e. remembering that the agent Mongoose was stationed in Colombia and bites their nails), but was not sufficiently difficult to measure individual differences in binding of the semantic information to a specific individual. Future work could provide insight by increasing the perceptual similarity of identities within a block.

While a linear regression provides insight into the connections between our cognitive factors and performance on face memory tasks, it cannot investigate whether there are multiple strategies that may achieve the same level of performance. To that end, we created two datasets selecting only subjects who were within 0.75 standard deviations of the mean for the CFMT and the PIM – Recognition Task to reduce variance in outcome measures related to overall performance. We then implemented k-means clustering (Hartigan & Wong, 1979), an unsupervised clustering algorithm, to investigate whether there were multiple cognitive profiles of individuals who had similar performance, which may indicate differential strategy use.

For both samples, we identified a 2-cluster solution and showed that the resulting groups were matched on performance on the outcome measures. Matched performance was not necessarily required by this analytic framework, as outcome measures were not included as features for the clustering algorithm. Despite being matched on performance, however, the two clusters revealed two distinct psychometric cognitive profiles - one with strengths in face perception and general episodic memory and weakness in general intelligence and working

memory, and one with the opposite pattern of abilities. These profiles were replicated in both sub-samples of our data; although the samples were not entirely independent, replication of the patterns across both samples suggested that these profiles may be reliable. An additional replication sample will be useful to further confirm the reproducibility of these psychometric profiles.

Although the dissociation between the Working Memory factor and Face Perception and Memory factors is striking, the mere existence of these profiles does not necessarily suggest distinct strategy usage across profiles. To further investigate this hypothesis, we conducted additional linear regression analyses where we included an interaction between the Face Perception and Working Memory factors (which showed the strongest and most reliable dissociation across both clustering solutions). This analysis showed a significant interaction between factor scores on the Working Memory and Face Perception factors when predicting the PIM – Recognition Score, but not the CFMT. Decomposing this interaction revealed that the effect of Face Perception ability was stronger when Working Memory ability was weaker, and vice versa. This result is consistent with work suggesting that strategy use during working memory tasks and moving viewing may depend on working memory capacity (Finn et al., 2020; Jafarpour et al., 2022).

A common approach to study face memory has been to compare typically developing individuals with either DPs (Duchaine & Nakayama, 2006; Guo et al., 2018) or super-recognizers (Nador et al., 2021) to identify processes or regions associated with differences in face memory ability. That is, this approach takes pre-defined groups that vary in face memory performance and looks for differences. While these analyses have been critical to our understanding of differences that relate to superior performance, our clustering analysis takes an alternative approach -

controlling performance to determine whether there are multiple strategies that may produce equivalent performance.

Using this approach, we identified a subset of individuals who rely on face perception and general memory ability to accomplish our PIM task. Although there is a relatively low correlation between face perception and face memory tasks, one study relating individual differences in face matching to those in face learning suggests that better face matchers (i.e. those with stronger face perception ability) can capitalize on within identity variability in appearance, which can in term better support face memory performance (Baker & Mondloch, 2023). This profile is also consistent with recent work identifying functional connectivity between face perception regions and regions associated with episodic memory at rest (Ramot et al., 2019) and during naturalistic movie watching (Levakov et al., 2023).

The other subset of individuals identified in our clustering analysis had weaker face perception ability but scored strikingly high on our Working Memory factor. One study suggested that increased modularity within the face processing network is correlated with better face memory performance (Levakov et al., 2022). Although this study only included regions within the ventral face processing pathway, they suggest that the increased modularity of the face processing network may serve to reduce interference from distractors, which may also be associated with increased working memory capacity (Gallen et al., 2023; Stanley et al., 2014; Zanto & Gazzaley, 2009). A separate line of work identified a functional connectivity profile characterized by connectivity between occipital visual regions and the posterior parietal cortex, medial prefrontal cortex and inferior frontal gyrus, which they suggest may be related to the effortful modulation of attention during visual processing to compensate for "sub-optimal" processing of faces (Levakov et al., 2023). Although this connectivity pattern was shown to be negatively correlated to face memory

performance, our results suggest that this strategy may still support at least average performance on face memory tasks, even if it is not the most "optimal."

One potential limitation of our work is the choice of tasks in our battery. Although we ensured that the tasks that we collected were reliable enough to use as stable measures for individual differences analyses (Kadlec et al., 2023), we had relatively few measures of social cognition, which may be implicated in individual differences in face memory (Ramot et al., 2019). We were able to include a composite measure reflecting emotion processing, which has been shown to be associated with face perception (Gobbini & Haxby, 2007); however, we lacked a stable measure capturing other aspects of social cognition, such as the use of social information or theory of mind (Bland et al., 2016). We additionally did not include measures of decision making, which may be implicated in the ability to reject unknown lure faces, a process that may be particularly relevant for our Recognition task (Devue et al., 2019). Factors reflecting other cognitive processes may improve the amount of variance explained in our outcome measures and provide additional insights into the ways to accomplish face memory tasks.

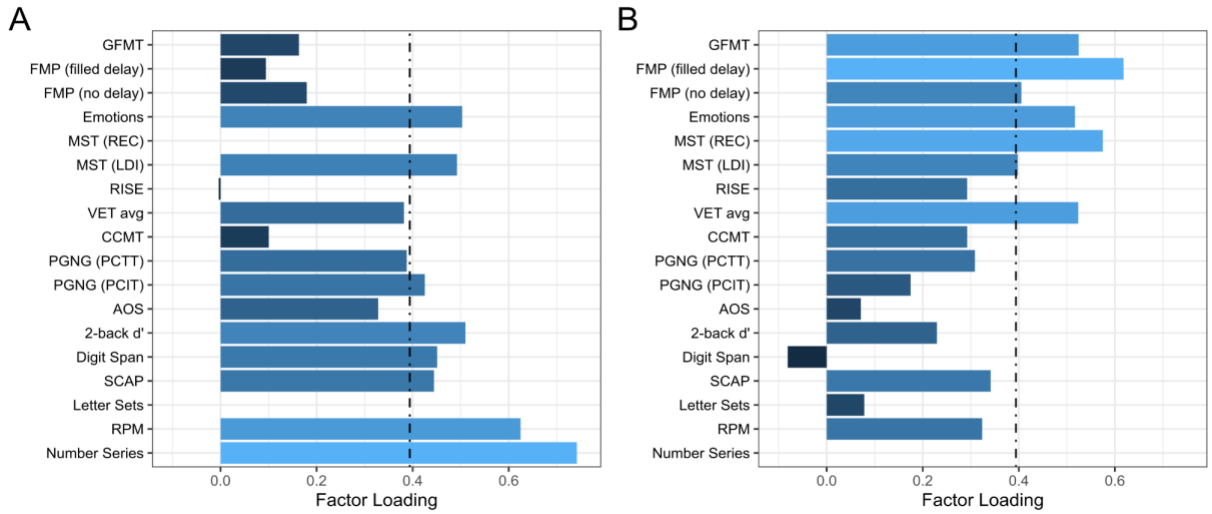
An additional limitation lies in our relatively small sample size for EFA (De Winter & Dodou, 2012; Mundfrom et al., 2005). Guidelines for sample size for EFA suggest a range of 110-240 subjects for our variable to factor ratio (Mundfrom et al., 2005). Although we have relatively few participants, we collected multiple forms of each task to ensure that we had enough trials to reliably measure individual differences in performance on each of the tasks (Kadlec et al., 2023), which may reduce the necessary sample size for EFA by reducing the noise present in the data. Although we report robust factor structures and well-fitting models, replicating our results in a larger sample would be helpful to validate our findings.

Taken together, our results highlight how face memory, despite being ubiquitous in our day-to-day lives, necessitates the use of more than just pure face perception. Moreover, the clustering results and significant interaction of the Face Perception and Working Memory factors highlight that there may be two potential strategies for accomplishing a face memory task – one that relies on face perception and memory ability, where individuals rely on their ability to perceive and remember nuanced visual details in a to-be-learned face, and one that utilizes working memory, such as elaborative encoding, imagery, increased rehearsal, goal maintenance or suppression of irrelevant information. Finally, our results suggest that the use of these strategies may depend on individual differences in cognitive ability, where individuals rely on their strengths. Future studies could extend this work by targeting the use of a particular strategy, either by specific behavioral interventions or non-invasive brain stimulation like covert real-time neurofeedback, to determine whether strategies may have different effects across the cognitive profiles. Successfully targeting processes separate from face perception may prove fruitful for novel targeted interventions for disorders like developmental prosopagnosia.

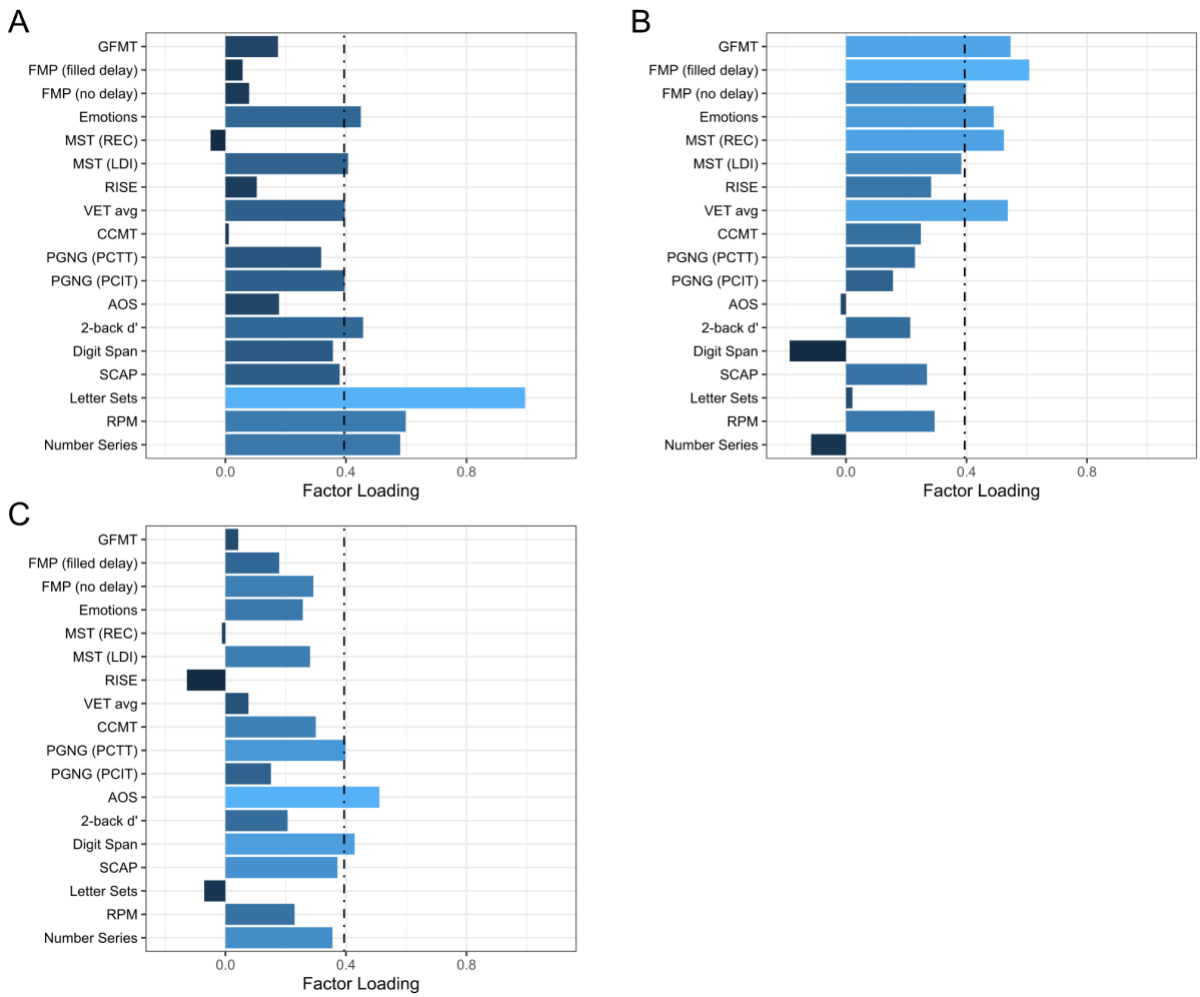
Supplementary Materials

Effect	Estimate	SE	95% CI	t statistic	p
Intercept	0.088	0.327	(-0.562 0.737)	0.268	0.789
Face Perception	0.348	0.081	(0.188 0.508)	4.326	0
Working Memory	0.167	0.063	(0.042 0.292)	2.659	0.009
Episodic LTM	0.377	0.069	(0.239 0.515)	5.42	0
Intelligence	0.238	0.075	(0.088 0.388)	3.153	0.002
Age	-0.004	0.01	(-0.023 0.016)	-0.39	0.697
Sex	0.052	0.169	(-0.283 0.386)	0.306	0.76
Face Perception x Working Memory	-0.053	0.067	(-0.186 0.079)	-0.802	0.425

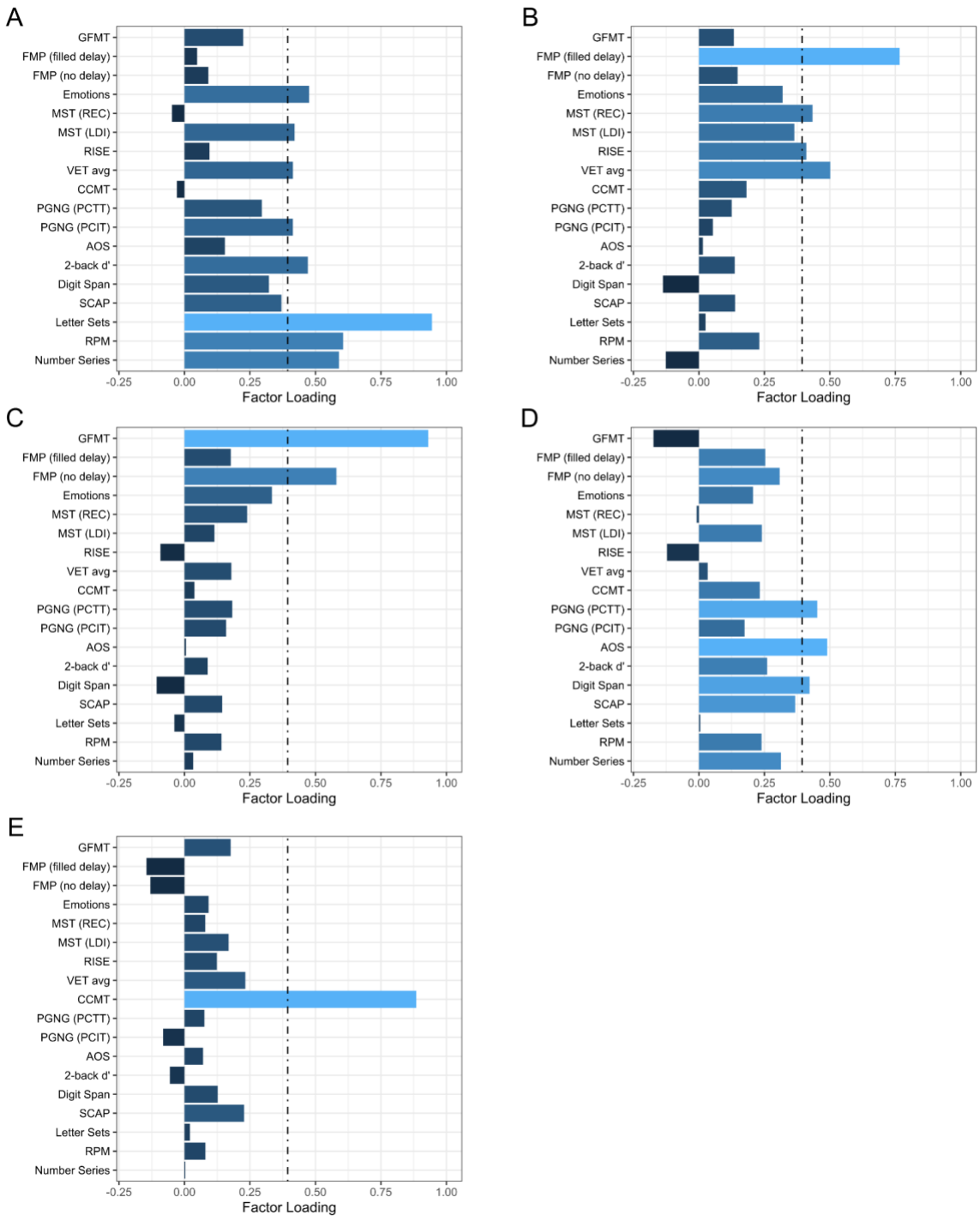
Supplementary Table 3.1: Regression coefficients for CFMT including interaction term.



Supplementary Figure 3.1: Factor loadings from the 2-factor solution. Data come from N = 101 subjects. Factors are ordered by their explained variance. Dotted line reflects a loading of 0.4, which was our threshold for saliency. Abbreviations: PIM: Personal Identity Memory Task; CFMT: Cambridge Face Memory Test; GFMT: Glasgow Face Matching Test; PI20: 20 question Prosopagnosia Index; FMP: Face Memory and Perception Task; CCMT: Cambridge Car Memory Task; VET: Vanderbilt Expertise Task; MST (REC): Mnemonic Similarity Task (Recognition Score); MST (LDI): Mnemonic Similarity Task (Lure Discrimination Index); RISE: Relational and Item Specific Encoding; PGNG (PCTT): Parametric Go/No-Go Task (Percent Correct to Targets); PGNG (PCIT): Parametric Go/No-Go task (Percent Correct to Inhibitory Trials); AOS: Automated Operation Span; SCAP: Spatial Capacity; RAPM: Raven's Advanced Progressive Matrices



Supplementary Figure 3.2: Factor loadings from the 3-factor solution. Data come from N = 101 subjects. Factors are ordered by their explained variance. Dotted line reflects a loading of 0.4, which was our threshold for saliency. Abbreviations: PIM: Personal Identity Memory Task; CFMT: Cambridge Face Memory Test; GFMT: Glasgow Face Matching Test; PI20: 20 question Prosopagnosia Index; FMP: Face Memory and Perception Task; CCMT: Cambridge Car Memory Task; VET: Vanderbilt Expertise Task; MST (REC): Mnemonic Similarity Task (Recognition Score); MST (LDI): Mnemonic Similarity Task (Lure Discrimination Index); RISE: Relational and Item Specific Encoding; PGNG (PCTT): Parametric Go/No-Go Task (Percent Correct to Targets); PGNG (PCIT): Parametric Go/No-Go task (Percent Correct to Inhibitory Trials); AOS: Automated Operation Span; SCAP: Spatial Capacity; RAPM: Raven's Advanced Progressive Matrices



Supplementary Figure 3.3: Factor loadings from the 5-factor solution. Data come from N = 101 subjects. Factors are ordered by their explained variance. Dotted line reflects a loading of 0.4, which was our threshold for saliency. Abbreviations: PIM: Personal Identity Memory Task; CFMT: Cambridge Face Memory Test; GFMT: Glasgow Face Matching Test; PI20: 20 question Prosopagnosia Index; FMP: Face Memory and Perception Task; CCMT: Cambridge Car Memory Task; VET: Vanderbilt Expertise Task; MST (REC): Mnemonic Similarity Task (Recognition Score); MST (LDI): Mnemonic Similarity Task (Lure Discrimination Index); RISE: Relational and Item Specific Encoding; PGNG (PCTT): Parametric Go/No-Go Task (Percent Correct to Targets); PGNG (PCIT): Parametric Go/No-Go task (Percent Correct to Inhibitory Trials); AOS: Automated Operation Span; SCAP: Spatial Capacity; RAPM: Raven's Advanced Progressive Matrices

Chapter 4: Integrating Task-based, Resting State, and Structural Neuroimaging Features to Predict Individual Differences in Working Memory and Psychiatric Outcomes

Introduction

Many of the complex actions and behaviors that make up our lives require the maintenance, manipulation and updating of information over time. It is widely accepted that there are limits to the capacity of working memory (Baddeley & Hitch, 1974; Engle, 2002; Unsworth & Engle, 2007a), and that this capacity varies across individuals (Cowan, 2010; Luck & Vogel, 2013; Miller, 1956). Individual differences in working memory capacity (WMC) have been shown to be related to higher order cognition, including intelligence (Chuderski et al., 2012; Johnson et al., 2013; Unsworth et al., 2014, 2015), reasoning ability (Kyllonen & Christal, 1990) and procedural and declarative learning (Kyllonen & Christal, 1990).

Deficits in WMC are correlated with dysfunction in psychiatric conditions such as schizophrenia (Perlstein et al., 2001), anxiety (Lapointe et al., 2013; Moriya & Sugiura, 2012), depression (Berman et al., 2011), bipolar disorder (Thompson et al., 2006), substance abuse (Grenard et al., 2008), and neurodevelopmental disorders, such as autism and ADHD (Gathercole & Alloway, 2006). One goal of the NIH Research Domain Criteria (RDoC) project is to formally characterize this relationship and relate neurocognitive markers of working memory to psychopathology across broad diagnostic categories (Bilder et al., 2013; Cuthbert & Insel, 2013).

Given the broad relevance of WMC, it is important to consider what neural processes may underlie individual differences in WMC. Early theoretical work suggested that differences in WMC may stem from differences in processes comprising executive attention, including inhibition of irrelevant information or processes, shifting focus as task demands change and updating of

actively maintained content (Engle, 2002). Other work has characterized working memory as interactions between Primary Memory (PM – short-term active maintenance of representations for a short period of time for ongoing processing) and Secondary Memory (SM – long term, stable storage). In this framework, WMC is defined by the amount of information that can be held in PM, the ability to effectively retrieve information from SM and the ability to selectively encode task-relevant information in SM (Unsworth & Engle, 2007b). A third theoretical perspective suggests that WMC is defined by the ability to build, maintain and update arbitrary bindings between information, and that the limitations of WMC stem from the ability to resolve interference between bindings (Oberauer et al., 2007; Wilhelm et al., 2013).

Neuroimaging work has provided preliminary insights into the neural underpinnings of individual differences in WMC (Minamoto et al., 2017). Univariate load-related activity (that is, BOLD signal levels modulated by the information maintenance demands of a task) in the prefrontal cortex (PFC) (Assem et al., 2020; Burgess et al., 2011; M. Osaka et al., 2003; N. Osaka et al., 2004; Rypma et al., 2002; Rypma & D'Esposito, 1999) and posterior parietal cortex (Todd & Marois, 2005) has been shown to vary with individual differences in working memory performance and capacity. Recent reviews have additionally highlighted the role of sustained univariate delay period activity in the prefrontal cortex as being crucial for working memory performance (Li et al., 2022; Riley & Constantinidis, 2016; Sreenivasan et al., 2014).

Other neuroimaging work has sought to characterize the mechanism of working memory using multivariate analytic techniques capable of capturing the informational content represented within distributed patterns of BOLD activity. Population coding of specific to-be-learned items during working memory may be reflected in multivariate pattern representations (Sreenivasan & D'Esposito, 2019). To that end, task-relevant visual information can be decoded from multivariate

pattern representations in visual, parietal and frontal cortices (Albers et al., 2013; Christophel et al., 2012; Harrison & Tong, 2009; Lewis-Peacock et al., 2012; Yu & Shim, 2017), and the fidelity of these representations in the face of distractions is associated with behavioral performance (Bettencourt & Xu, 2016; Hallenbeck et al., 2021).

Although performance on tasks involving working memory is certainly influenced by in-the-moment factors such as fluctuations in attention or arousal, WMC is considered to be a stable trait in individuals (Engle et al., 1999; Stevens et al., 2012). Thus, it is also worth considering measures of the brain that may reflect more stable trait-like properties, such as resting state functional connectivity, as indicators of WMC. Functional connectivity of the frontoparietal and default mode networks during task and rest has been shown to predict working memory performance (Avery et al., 2019; Murphy et al., 2020; Zhu et al., 2021) as has the strength of the connectivity between the lateral PFC and saliency network (including the anterior cingulate cortex and anterior insular cortex) (Fang et al., 2016) and the anticorrelation between the lateral and medial PFC (Keller et al., 2015) at rest. Resting state network structure, including measures of functional integration, such as modularity, network clustering and efficiency (Rubinov & Sporns, 2010), have been associated with better working memory performance (Alavash et al., 2015; Cohen & D'Esposito, 2016; Stevens et al., 2012). Moreover, the global connectivity of the lateral PFC at rest has been shown to be positively correlated with WMC (Cole et al., 2012).

Neuroimaging-based measures of macroanatomical brain structure may also help explain individual differences in working measures of capacity. For instance, recent work has suggested that a larger grey matter volume in the lateral occipital region is associated with higher WMC, whereas greater precision in the items held in working memory was associated with larger grey matter volume in the parietal lobe (Machizawa et al., 2020). Other work has implicated volumetric

differences in middle frontal gyrus, lateral frontal and posterior cingulate cortex in cognitive control and visual working memory capacity (Eayrs & Lavie, 2019; Xiao et al., 2021).

Although studying individual modalities has been fruitful for understanding the neural underpinnings of WMC, recent work attempting to predict individual differences in cognition has highlighted how models that integrate across modalities can explain more variance in individual differences in cognition than can any single modality on its own (Dhamala et al., 2021; Jiang et al., 2020; Ooi et al., 2022; Rasero et al., 2021; Tetereva et al., 2022). Moreover, recent work has shown that using task-based fMRI in these predictive models has the potential to improve predictive ability over measures such as resting state functional connectivity, structural connectivity, and anatomical features of the brain (Gao et al., 2019; Greene et al., 2018; Jiang et al., 2020; Tetereva et al., 2022).

In the current work, we seek to predict individual differences on in-scanner performance of a working memory task, working memory capacity measured outside the scanner, and overall psychiatric functioning in a large heterogeneous sample of adult participants who show a wide range of psychiatric disability. We identify neural regions that show significant group-level load-related differences in univariate activity, face-specific maintenance, or encoding-to-delay representational stability during a delayed match-to-sample task and use these measures, in addition to resting state functional connectivity, local and global resting state network structure, and anatomical features to train a series of machine learning algorithms to predict our outcome measures in a held-out test set. We explicitly compare models that use each modality individually to models that include all features and the relative predictive ability of each unique modality across outcome measures. Finally, we characterized the importance of each brain feature to investigate the regional contributions to our outcome measures. We expected to show that features from task

fMRI, including multivariate pattern features, across the regions in the prefrontal and parietal cortices, would be able to significantly predict performance on an in-scanner working memory task, working memory capacity, and indices of psychiatric function. Although we expected that each of our individual modalities would be able to predict our working memory outcomes, we expected that combining the feature sets into one full model would be able to predict more variance in our outcome measures than any one individual modality. Finally, we expected that there may be different patterns of features that predict our outcome measures, particularly in the models that predicted working memory performance and capacity relative to our measures of psychiatric functioning. Specifically, we predicted that our measures of working memory performance and capacity would be better predicted by the task-based measures, while the measures of psychiatric functioning would be better explained by stable, trait-like measures such as those from resting state and structural MRI.

Methods

Participants

Data from 200 participants (129 females, ages 21-40, mean age = 28.02, SD = 5.29) were collected as a part of a Research Domains Criteria inspired project, which sought to recruit adults across a wide range of trans-diagnostic mental health concerns. Participants were care-seeking (CS; n = 142) or non-care-seeking (NCS; n = 58) community volunteers from the greater Los Angeles area. The CS group was composed of individuals who were seeking treatment for mental or emotional problems and responded to advertisements or individuals who were referred to the UCLA Neuropsychiatric Behavioral Health Service. Non-care seeking participants were recruited through separate advertisements and had not sought behavioral health or substance-abuse

treatments within the year prior to enrollment. All participants provided informed consent as outlined by the University of California, Los Angeles Institutional Review Board and underwent extensive clinical and neurocognitive assessments. Relevant measures are reported below, with additional measures reported in Lenartowicz et al., 2021.

Inclusion/Exclusion

Participants were required to have had completed at least 8 years of formal education, showed sufficient general mental status, hearing, motor coordination and cooperation to be able to complete the session, were proficient in English, had an IQ estimate >70 (as measured on the Wechsler Adult Intelligence Scale - Fourth Edition (WAIS IV; (Wechsler, 1945) - Vocabulary and Matrix Reasoning). Participants were also required to have visual acuity better than 20/50 with each eye (tested separately) and no medical or neurological illness or treatment expected to have cognitive effects. Participants were excluded if they had taken any psychotropic or sedating drugs within 24 hours of exam, long-acting antipsychotics, electroconvulsive therapy, or diagnosis of substance abuse disorder (other than caffeine or nicotine) within the previous 6 months. Participants were required to have a negative urinalysis for THC, cocaine, amphetamine, opiates, and benzodiazepines. Finally, participants were excluded if they had any contraindications to MRI exam (metal in body, claustrophobia, or for females, pregnancy).

Additionally, we excluded participants from our analyses who did not have complete MRI data for our task of interest, participants who had a faulty pulse sequence (n=11), bad field of view (n=4), or excessive movement (n=1) in the scanner. Finally, we excluded 14 participants because they responded to the scanned task with below-chance accuracy according to a within-subject binomial test, and 1 participant who was missing a structural T1-weighted scan, leaving us with a

final sample of 169 participants (106 females, ages 21-40, mean age = 27.787, SD = 5.120; 116 CS, 53 NCS).

Behavioral Testing

Participants underwent extensive behavioral testing to measure cognitive ability and self-reported clinical symptomatology. Brief descriptions of the tests and the outcome measures included for further analyses are as follows, although see Lenartowicz et al., 2021 for a more in-depth description of the cognitive tasks:

Clinical Testing

- **World Health Organization Disability Assessment Schedule (WHODAS 2.0)**: measure of disability. Outcome measure: Total score.
- **Brief Psychiatric Rating Scale (BPRS)**: includes measures of Reality Distortion, Disorganization, Blunted Affect, Depression-Anxiety and Mania-Excitement. Outcome measure: scores summed over all domain scores to create a total score.

Cognitive Testing

- **WAIS-IV Vocabulary**: Participant is presented with individual words and must define them. Test measures crystallized intelligence. Outcome measure: Total raw score.
- **WAIS-IV Matrix Reasoning**: Participant is presented with a matrix of abstract images with one piece removed and must choose the missing piece from a selection of possible options. Test measures fluid intelligence. Outcome measure: Total raw score.

- **WAIS-IV Digit Span Forward**: Participant is read a sequence of numbers and must recall numbers in the order presented. Outcome measure: Total raw score.
- **WAIS-IV Digit Span Backwards**: Participant is read a sequence of numbers and must recall numbers in reverse order. Outcome measure: Total raw score.
- **WAIS-IV Digit Sequencing**: Participant is read a sequence of numbers and must recall numbers in ascending order. Outcome measure: Total raw score.
- **WAIS-IV Letter Number Sequencing**: Participant is read a sequence of letters and numbers and must recall letters in alphabetical order and numbers in ascending order. Outcome measure: Total raw score.
- **WMS-IV Spatial Addition**: Participant is shown two grids with blue and red circles and is asked to perform addition or subtraction operations based on a set of rules. Outcome measure: Total raw score.
- **WMS-IV Symbol Span**: Participant is shown a series of symbols and must recall the symbols in the order they were presented. Outcome measure: Total raw score.
- **Automated Operation Span**: total number correct (partial method): Participant presented with a series of letters that they must remember. Between successive letters presented, participants must complete math problems. Task from (Unsworth et al., 2005). Outcome measure: total number correct (partial method).
- **Lateralized Change Detection**: Participants must maintain paired color and location information of up to 5 colored objects. Outcome measure: Maximum storage capacity, calculated as $K = n * (HR – FA)/(1-FA)$, where n is the number of to be maintained objects, HR is the hit rate and FA is the false alarm rate. Task adapted from (Leonard et al., 2013). Outcome measure: Maximum K across all array sizes.

- **Spatial Working Memory Capacity**: Participant must maintain locations of up to 7 dots. Task adapted from (Glahn et al., 2002). Outcome measure: Maximum capacity across loads, as defined by Cowan's K (K = n * (H-FA), where n = load, H = hit rate, and FA = false-alarm rate).
- **Dot Probe Expectancy**: Participants see a series of cue and probe dot patterns and respond with either a target or non-target response. One cue ("A") is always considered valid, one ("X") is only considered valid when following "A" cue. All other cues are considered invalid cues. Task adapted from (Henderson et al., 2012). Outcome measure: d' (d-prime), defined as the z(HR) – z(FA) – that is, the difference of the z-transform of hit rate and false alarm rate.

Working Memory Capacity Measures

To maximize the effectiveness of our measurement of cognitive abilities, we created composite measures of working memory capacity. To do so, we submitted our behavioral indices of working memory to an exploratory factor analysis (EFA) (Fabrigar et al., 1999). We first tested whether our data were suitable for EFA using Bartlett's Test of Sphericity (Bartlett, 1950), which tests whether the correlation matrix is different from an identity matrix, and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1970; Kaiser & Rice, 1974) using the R *psych* package (Revelle, 2023). The KMO test determines the strength of the partial correlations of a variable, relative to the original correlations. When variables load onto distinct common factors, there will be high correlations but low partial correlations. According to Kaiser (1970), a KMO value of >0.9 is "marvelous," between 0.8-0.9 is "meritorious," between 0.7-0.8 is "middling," between 0.6-0.7 is "mediocre," between 0.5-0.6 is "miserable" and below 0.5 (when

the partial correlation matrix would be equivalent to the correlation matrix), is considered unacceptable.

We then determined the appropriate number of factors to extract using used Horn's parallel analysis (Horn, 1965), which creates randomly shuffled data (i.e. data where there is no underlying factor structure) that serves as a null hypothesis and compares the resulting eigenvalues to those extracted from the true data. The crossover point, where the eigenvalues of the measured data are larger than those from the shuffled data, indicated how many factors to extract.

After identifying the appropriate number of factors to extract, we implemented EFA with a maximum likelihood estimation (De Winter & Dodou, 2012) and using an oblimin rotation (Jackson, 2005), with mean imputation for the missing data. As has been recommended in guidelines from the literature, we used multiple indicators to measure model fit (Hu & Bentler, 1999). First, we used the Tucker-Lewis Index (TLI) (Bentler & Bonett, 1980; Tucker & Lewis, 1973), which compares the fit of a given model to a baseline model where all variables are uncorrelated. A TLI of greater than 0.95 is considered a good fit, while above 0.9 is considered an acceptable fit.

Next, we used two measures of absolute fit: root mean square of residuals (RMSR) (Bentler, 1995) and root mean square error of approximation (RMSEA) (Steiger & Lind, 1980). Following guidance from (Hu & Bentler, 1999), we considered both the point estimate as well as the 90% confidence interval of the RMSEA. These measures compare the model-specified fit to the sample data. RMSR compares the covariance matrix of the observed data to the covariance matrix specified by a hypothesized model, while RMSEA makes a similar comparison but takes into account the degree of misfit per degree of freedom. A smaller value on both measures indicates a better fit, with an RMSR < 0.05 and an RMSEA < 0.06 with a lower value of the confidence

interval being close to 0 and upper value being less than 0.08 being considered a good fit. An RMSR and RMSEA < 0.08 was considered an acceptable model fit.

We interpreted the resulting factors through the measured variables with loadings greater than 0.4 (Watkins, 2018). We extracted Bartlett factor scores for each participant for each factor. Factor scores extracted using this method are only impacted by the shared factors and minimize the influence of the unique variance of the set of variables, resulting in factor scores that are only correlated with their associated factor and not to the other factors. Moreover, this method produces unbiased estimates of the factor score parameters (DiStefano et al., 2009). These scores served as two of our five outcome measures that we sought to predict.

Additional Outcome Measures

In addition to the factor scores reflecting working memory span measures, we also sought to measure individual differences in average performance across loads on our delayed face recognition (DFR) task (described in detail below) and individual differences in psychiatric functioning, as measured by the WHODAS 2.0, and psychiatric disability, indexed by the total score from the BPRS.

fMRI Behavioral Protocols

Apparatus

Visual stimuli were shown with a custom-built MR-compatible rear projection system, viewed through a mirror mounted on the head coil above the participant's eyes. Responses were collected with an MR-compatible button box.

Resting State

Participants completed 6 minutes of resting state fMRI. During this time, participants were presented with a dark blue screen and were instructed to let their mind rest and wander freely while keeping their eyes open so that they did not fall asleep.

Face Localizer

Participants completed one approximately 6-minute run of a task designed to functionally localize face-specific representations in each individual participant's brain (Saxe et al., 2006). Participants were presented with 18 consecutive blocks of centered greyscale images of faces, objects, or scrambles. Blocks were presented in a pseudorandom order that was the same across all participants. Each block included 16 images presented for 600ms followed by a fixation cross for 400ms. The inter-block interval was 2,000ms. Participants performed a 1-back task and were instructed to judge whether the image was the same as the one presented directly before. Participants responded only if the image was the same as the one before using a right index button press. Participants were instructed to fixate on the cross at the center of the screen and not respond if the image was not the same as the one before. After the final block, participants were presented with a fixation cross for 9,000ms.

Delayed Face Recognition Task

Participants performed a delayed-match-to-sample task (modeled after Druzgal & D'Esposito, 2003; Morgan et al., 2008; Rissman et al., 2008) requiring maintenance of face stimuli across a brief delay period (Figure 4.1). Each trial began with the presentation of 1 face (low load) or 3 faces (high load) for 2,500ms (encoding period). Participants were tasked with remembering

the face(s), which appeared in random quadrants of the display. Bottom-up visual stimulation was balanced across load conditions by including a scrambled face that participants were not responsible for remembering in any quadrant not containing a real face. On each trial, to-be-remembered faces were drawn from a set of 32 unique greyscale face images (16 females and 16 males; genders were never mixed within a trial). Faces were unfamiliar to participants at the onset of the experiment and were cropped with a rectangular box. Each face was presented during the encoding period an average of 16 times across all runs and was used during the probe period an average 8.53 times. After presentation of the face(s), participants were presented with a fixation cross for 7,500ms (delay period) where they were required to maintain the face(s) they were presented with. Following the delay period, participants were presented with a single face for 2,000ms (probe period) and needed to make a decision using a response box as to whether this probe face had been encountered at the start of the trial. Participants used their index finger to respond that the probe matched a face from the encoding period or their middle finger to respond that the probe did not match a face from the encoding period and were encouraged to respond as quickly and as accurately as possible. Half of the trials included a matching face and half did not. Presentation of faces was randomized once and presented in the same order across participants.

The inter-trial interval (ITI) included an active baseline, where participants were presented with a fixation cross for 1,250-1,750ms, followed by a series of arrows pointing either left or right. Participants had to respond whether the arrow was pointing right or left by pressing a button with their middle or index finger, respectively. Arrows were presented on screen for 750ms, followed by a fixation cross at the center of the screen for 1,250ms. After the third set of arrows, the fixation cross was jittered to either 1,250 or 1,750ms. ITI jittering was set so that the MRI volume acquisition onset synced with the onset of face presentation 50% of the time for each load. Prior

to performing the task in the fMRI, participants completed a training task. In this training, they performed an abbreviated version of the same task albeit with different faces from those used in the scanning session. The only difference between the training and scanned task was that following each trial, participants received feedback for 1,500ms indicating whether they were correct or incorrect. Participants ran this training until they understood the task (usually 2-3 trials).

Each scanned run included 8 high load trials and 8 low load trials. Half of the trials included a match between the probe face and one of the faces in the encoding period. There were 4 runs of the task, meaning there were 32 high load trials and 32 low load trials in total. Runs began with an initial fixation cross for 4,500ms and ended with a fixation cross for 9,000ms.

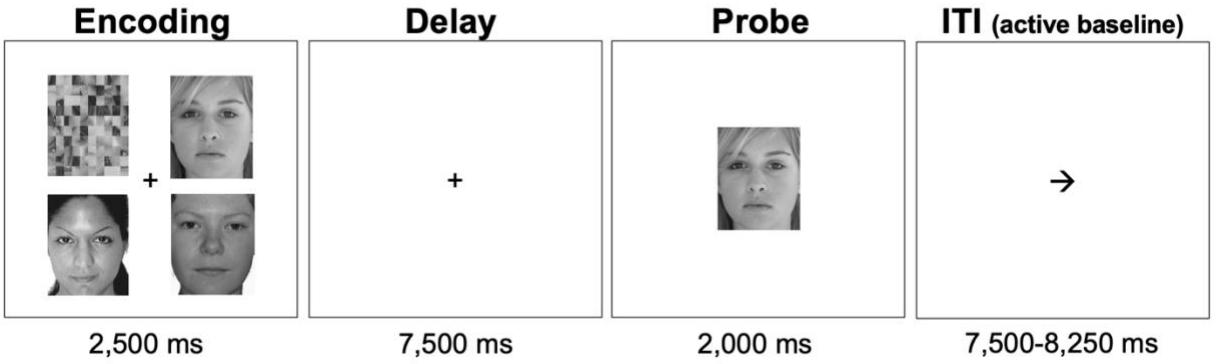


Figure 4.1: Delayed Face Recognition task design. Figure shows an example of a high load trial. Participants see three faces in a high load trial (depicted here) or one face in a low load trial for 2.5 seconds, must maintain the face(s) over a 7.5 second delay period, and then have 2 seconds to respond whether a probe face is a match to a face from the encoding period. The ITI is jittered with an active baseline.

MRI Data Acquisition

Whole-brain imaging was conducted on a Siemens 3.0T Tim Trio MRI scanner (n=65) or a Siemens 3.0T Prisma MRI scanner (n=105) at the Staglin One Mind Center for Cognitive Neuroscience at UCLA (the scanner was upgraded from a Trio to a Prisma Fit system part way through data collection). All scanning was performed with a 32-channel head coil. Functional

images were collected using a T2*-weighted echoplanar imaging (EPI) sequence (TR = 1.5 s; TE = 34.2 ms; flip angle = 80; FoV = 19.2 cm, voxel resolution = 2 mm^3, multiband acceleration factor = 4). Each functional volume consisted of 68 axial slices acquired in a temporally interleaved sequence. Functional data for the delayed facial recognition task were collected across four runs of 221 volumes each. Face localizer data were collected in one run of 226 volumes. Resting state data were collected in one run of 240 volumes lasting 6 minutes and 11 seconds. The three initial volumes from each run of the delayed face recognition task and four initial volumes of the face localizer task were discarded to allow for T1 stabilization. A B_0 field map image was obtained prior to functional scanning to facilitate subsequent unwarping of anterior frontotemporal regions that are prone to susceptibility-induced distortion. To aid in spatial registration of the functional data, a coplanar T2- weighted anatomical image was also collected, along with a high-resolution (1 mm^3) magnetization prepared rapid gradient echo (MPRAGE) T1-weighted image. Image pre-processing and univariate fMRI analysis were performed with SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/).

Structural MRI Measures

Cortical surface reconstruction from the T1-weighted images was conducted using the standard processing pipeline from the Freesurfer imaging analysis suite (version 5.3.0, http://surfer.nmr.mgh.harvard.edu/ (Dale et al., 1999; Fischl & Dale, 2000). In brief, the following steps were performed: intensity normalization, motion correction, registration to the Talaraich space, skull stripping, automatic subcortical segmentation, tessellation of the grey/white matter boundary, surface inflation, automated topology corrections and surface deformation. Cortical surfaces were automatically parcellated using the standard Deskian/Killiany atlas into 33 cerebral

structures per hemisphere (Desikan et al., 2006; Fischl, 2004) and average cortical thickness was computed for each region. Additionally, subcortical volume was extracted for 7 subcortical structures – corpus callosum, nucleus accumbens, caudate, hippocampus, pallidum, putamen and thalamus. Subcortical volume was extracted from each hemisphere separately for each region except for corpus callosum, which was divided into anterior, mid-anterior, central, mid-posterior and posterior subdivisions.

fMRI Preprocessing

Task fMRI Preprocessing

All functional images were corrected for differences in slice acquisition timing, unwarped based on the voxel-displacement field map to correct for distortions in static magnetic field, and motion-corrected using a six parameter rigid-body realignment procedure. Image co-registration involved a two-part procedure where the coplanar anatomical image was registered to the mean functional image and the MPRAGE was registered to the coplanar anatomical. The MPRAGE was then segmented into grey matter, white matter, and cerebrospinal fluid (CSF), and the grey matter image was warped to the SPM8 MNI grey matter template. The resulting nonlinear warping parameters were in turn applied to the functional images.

Resting State fMRI Preprocessing

The ultimate goal of this preprocessing pipeline was to extract an average de-noised, de-trended and filtered time series (240 time points) from each of the 400 ROIs from the Schaefer atlas used in the task-based analyses. The first step consisted of flagging each volume for excessive movements or spurious global signal variation. Variance of signal change from the average signal

(DVAR), excluding frames with signal intensities greater than absolute value of 8, was used to censor frames due to excessive head movement (Shannon et al., 2011). Artifact Detection Tools (Nieto-Castañón, 2015) was also applied to detect frame-to-frame movements exceeding 0.5mm. Participants who had more than 25% of their volumes flagged were removed for further analysis.

Multiple linear regressions were then applied on every brain voxel in order to remove the signal from 1st order linear trends and movement related noise. The movement related regressors were built using translational and rotational movements matrices R, R^2 and by shifting one TR the R matrix in both directions R_{t-1} and R_{t+1} (Friston, Williams, et al., 1996). The average signals from the WM, the CSF and the global brain signal (GM+WM+CSF) were also regressed out. The residual of the regressions (i.e. the signal of interest) was then normalized to mean of 0 and a standard deviation of 1. To avoid movement artifacts impacting the next steps of the filtering process, each flagged volume was replaced with an interpolated volume that preserved the frequency structure of the signal (Power et al., 2014). This composite new signal was then filtered to keep low frequencies between 0.009 and 0.08 Hz. Finally, each flagged volume was censored leaving 400 time courses with between 180 (75% of 240) and 240 time points.

First-level fMRI Analyses

Delayed Face Recognition Task - Univariate Analyses

At the single-subject level, fMRI data from the DFR task were spatially smoothed using a 6mm full width at half maximum (FWHM) Gaussian kernel, then were analyzed using the general linear model (GLM) framework with an event related design. The GLM included regressors for each of the task phases (encoding, delay, and probe), modeled using boxcars of 2.5, 7.5 and 2.0

seconds, respectively. Correct and incorrect trials were modeled separately. Task regressors were all convolved with a canonical hemodynamic response function. Several covariates of no interest were also entered into the model, including run means, 6-direction head movement parameters, and a variable number of stick-function regressors corresponding to artifact prone volumes to censor from analysis. Censored volumes were flagged using ArtRepair (http://www.cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html) as having translational movements exceeding 2 mm and/or global signal changes exceeding 6 SD from the mean. Serial autocorrelation and low-frequency drifts were accounted for using a first-order autoregressive model and a high-pass filter of 0.0078 Hz (cutoff period = 128s). Data were parcellated into the 400 region-of-interest (ROI) parcellation of the 17 Network Schaefer atlas (Schaefer et al., 2018) and the mean activation parameter estimate value was extracted from each ROI. Load effects were calculated by subtracting the mean activation from Load 1 from the mean activation from Load 3 for each task stage. For the following analyses, we will focus on the load effects from the delay period.

Delayed Face Recognition – Face-Specific Representational Content

In addition to the univariate task-level analyses, we also sought to probe face-specific representational content during the DFR task. To that end, we re-processed the DFR task data using a similar pipeline as the univariate task fMRI data, but with a 2mm FWHM Gaussian kernel, as has been recommended for RSA analyses (Dimsdale-Zucker & Ranganath, 2019). Additionally, we did not apply any drift model and used a 0.0078 Hz high-pass filter.

We used a similar processing pipeline for the face localizer task, which was used to define category-specific "templates" to characterize representational content during the DFR task. Spatial

smoothing was applied with a 2mm FWHM Gaussian kernel, and the data were analyzed using the GLM framework with a block design. The GLM included regressors for each condition (face, object, or scramble) modeled with boxcars of 16s. Task regressors were convolved with a canonical hemodynamic response function and covariates of no interest (6-direction head moment parameters and stick-function regressors corresponding to artifacts, defined in the same manner as above). Low frequency drift was accounted for using a high-pass filter of 0.0078 Hz (cutoff period = 128s).

Data from both the DFR task and the face localizer task were parcellated using the same 400 ROI parcellation of the 17 Network Schaefer atlas as the univariate analyses. Multivariate activity was extracted from each ROI from the t-map associated with each task phase of the DFR task, reflecting the representational content during the task, and the face vs baseline and object vs baseline t-maps from the localizer task. We then took the Pearson correlation between activity patterns from the DFR task and each of the localizer contrasts, resulting in a single continuous value measuring how similar the task content is to each category template. Correlation values were Fisher z-transformed prior to further analysis. Following Fisher's z-transformation, we took the difference between the correlation of task to the face template and the correlation of task to the object template to get a measure of face-specific maintenance at each load. Finally, we calculated the face-specific maintenance load effect by subtracting the measure of face-specific maintenance at Load 1 from that at Load 3.

Delayed Face Recognition Task - Trial-level Multivariate Pattern Analyses

In addition to the univariate task-related fMRI analyses, we also sought to investigate whether multivariate pattern representation across the duration of each trial was related to

individual differences in performance and in working memory capacity. To that end, we modeled data from the DFR using a least-squares single (LSS) approach (Mumford et al., 2012; Turner et al., 2012), which involves running a large series of separate GLM models, each of which includes one regressor for a single trial and task period (e.g., trial #1 encoding period) and another regressor encompassing all other trials and task periods. As with the univariate task fMRI data, we convolved task regressors with a canonical hemodynamic response function and entered covariates of no interest (run means, 6-direction head movement parameters, and a variable number of stick-function regressors corresponding to artifact prone volumes to censor from analysis) into the model. No autoregressive drift model was applied, and low-frequency drift was accounted for using a high-pass filter of 0.0078 Hz (cutoff period = 128s). The end result is a set of beta maps each reflecting the estimated activity level during a single task period of a single trial.

As with the representational content analyses, we extracted multivariate pattern activity from each of the 400 ROIs from the 17 Network Schaefer atlas. Instead of comparing these patterns to a template, however, we took the Pearson correlation of the multivariate patterns from the encoding period to those from the delay period at each load as a measure of the stability of representation across a single trial (hereafter called "encoding-to-delay pattern stability"). Correlation values were Fisher z-transformed prior to further analysis. Following Fisher z-transformation, we calculated a load effect measure for within-trial pattern stability by subtracting the encoding-to-delay pattern stability at Load 1 from the encoding-to-delay pattern stability at Load 3. Pattern similarity analyses (both face-specific representational content and encoding-to-delay pattern stability analyses) were conducted using custom scripts written in Python version 3.7.

Resting State fMRI Measures

A 400 x 400 correlation matrix was then obtained by computing pairwise Pearson's correlations between all ROIs. These matrices were Fisher's z-transformed prior to further analyses. We collapsed across subnetworks from the 17 Network Schaefer Atlas for all networks except for the fronto-parietal control network (i.e. Default A and Default B were collapsed into one Default network). Prior literature has suggested that there may be functionally distinct subnetworks within the frontoparietal control network (Dixon et al., 2018; Murphy et al., 2020; Yeo et al., 2011) and we wanted to maintain those distinctions in our analyses to best describe which of the FPCN subnetworks were most relevant for predicting working memory performance, WMC, and psychiatric function. We did not have comparable a priori predictions for the other networks so we collapsed across the subnetworks for ease of interpretation. Average across- and within-network connectivity was computed, resulting in a 10x10 connectivity matrix.

In addition to extracting network-average resting state functional connectivity, we also computed a series of graph-based connectomic measures that index the local and global network structure (Cao et al., 2019; Rubinov & Sporns, 2010). We thresholded the raw 400x400 Pearson correlation matrices to include only the top 20% of all edges (Luppi et al., 2023) and extracted the following measures for each of the 400 ROIs using the *igraph* (Csardi & Nepusz, 2006) and *brainGraph* (Watson, 2020) R packages:

- **characteristic path length**: a global network measure indexing the average shortest path length between all pairs of nodes in the network. A shorter characteristic path length may indicate more efficient communication across nodes in a network.
- **Louvain's modularity**: a global network measure indexing the extent to which the nodes in a network can be divided into non-overlapping communities. This measure was calculated using the

Louvain greedy algorithm (Blondel et al., 2008). A lower modularity value may indicate more integration across the brain.

- **participation coefficien**t: a local network measure indexing the relative number of intermodular versus intramodular connections of a given node. A high participation coefficient usually indicates that a node may be important for between-network communication.
- **strength**: a local network measure indexing the weighted sum of all edges for a given node. A large strength value indicates that a node is highly connected within a network.
- **betweenness centrality**: a local network measure defined as the fraction of all shortest paths in a network that pass through a node. A high betweenness centrality value for a node often indicates the node connects disparate parts of a network.

We additionally calculated global brain connectivity (Cole et al., 2010) from the unthresholded Fisher's z-transformed 400x400 correlation matrices. This measure splits functional connectivity by sign and computes the average positive and negative connectivity for each of the 400 ROIs.

Feature Set Definition

We identified three main classes of neural activity that may be associated with performance on a working memory task and working memory capacity: task fMRI measures, resting state fMRI measures and structural MRI measures. Specifically, for the task fMRI, we included data from the GLM univariate delay period activation, face-specific representational content, and encoding-to-delay pattern stability. For the task fMRI measures, we included both the values from Load 3 and the load effect for each measure.

From the resting state fMRI, we calculated average network-level functional connectivity, measures of global network structure (modularity, characteristic path length), and measures of local network structure (strength, betweenness centrality, participation coefficient, and positive and negative global brain connectivity) for each of the 400 ROIs. All graph theoretical measures, except for global brain connectivity, were calculated from the top 20^{th} percentile of connections; global brain connectivity was calculated from the full, unthresholded 400x400 connectivity matrix. Given previous literature suggesting that the strength of brain-behavior correlations within ROIs was strongly correlated with the strength of the between-subjects load effect (Y. P. Li et al., 2021), we opted to constrain the ROIs we included to those which showed a load effect at the group level for at least one of our three task fMRI measures. To that end, we performed a within-sample two-tailed t-test to test for differences between high-load and low-load trials in each of the 400 ROIs for each measure with corrections for multiple comparisons using FDR corrections (Benjamini & Hochberg, 1995) within each task stage and task measure.

Regions that showed a significant positive load effect (Load 3 > Load 1) during the delay period after corrections for multiple comparisons were selected for further analysis of the task fMRI measures, in addition to the measures of local network structure from resting state (strength, betweenness, participation coefficient, and positive and negative global brain connectivity). From the structural MRI measures, we included mean cortical thickness and subcortical volume of all regions in the Deskian/Killiany atlas.

Using these measures, we created 7 distinct feature sets – one for each of the individual task fMRI measures (univariate – 164 features, face-specific maintenance – 164 features, encoding-to-delay pattern stability – 164 features), one that included all task fMRI measures (492 features), one that included all the resting state fMRI measures (467 features), one that included

the structural MRI measures (85 features) and one set that included all measures combined (1044 features). Each feature set additionally included age, sex, and scanner model (Trio vs. Prisma) as covariates.

Prediction of Outcome Measures

For each of our feature sets, we used ElasticNet regression (Zou & Hastie, 2005), a linear and additive algorithm that has been previously used to predict cognition from brain measures (Pat et al., 2022; Tetereva et al., 2022). ElasticNet has the benefit of being interpretable (Molnar, 2020; Pat et al., 2022; Tetereva et al., 2022), with the magnitude of the coefficient reflecting the importance of each feature. Additionally, ElasticNet allows us to have more parameters (i.e. features) than the number of observations (i.e. participants), as it minimizes the squared distance between a fitted plane and the data points (James et al., 2021; Kuhn & Johnson, 2013) by simultaneously minimizing the weighted sum of the features' slopes. The degree to which the sum of the slopes is penalized is characterized by the shrinkage hyperparameter $\boldsymbol{\lambda}$ (higher $\boldsymbol{\lambda}$ values result in higher shrinkage of slopes). ElasticNet also includes a mixing parameter α, which indexes the degree to which the sum of either the squared (reflecting "ridge" regularization; $\alpha = 0$) or absolute (reflecting "LASSO" regularization; $\alpha = 1$) slopes is penalized (Zou & Hastie, 2005). We opted for a fixed mixing hyperparameter of $\alpha = 0.05$, as that was the most common best mixing hyperparameter in previous work predicting individual differences in working memory performance and cognitive ability (Pat et al., 2022) and favors less severe shrinkage of feature weights, allowing a more thorough interrogation of informative features. We tuned the shrinkage hyperparameter within each cross-validation fold (see below for cross-validation scheme details) using 1000 values of $\boldsymbol{\lambda}$.

We implemented the ElasticNet algorithm using the *eNetXplorer* R package (Candia & Tsang, 2019; https://github.com/cran/eNetXplorer) to fit two sets of models. The first set of models predicted the true response variable (DFR average accuracy, our two WMC factors, WHODAS score, BPRS; target models). The models were run in a 10-fold cross validation framework, where the data were split into 10 folds such that the model was trained on 9 folds and tested on the held-out fold. For each run, we defined the coefficient for each feature as the average of the non-zero coefficients across all folds. This process was repeated over 100 runs where the new random folds were sampled for each run; we additionally calculated an average coefficient over all runs, weighted by the frequency of obtaining a non-zero coefficient across runs. Each model's predictive capacity was measured by the median Spearman's correlation between the predicted and actual values across all folds in all runs.

The second set of models permutes the response variable and uses the same cross validation splits and overall procedure as the models predicting the actual response values to predict these permuted values (permuted null models). This permutation occurs 25 times (eNetXplorer default) in each of the 100 runs. Because the permutation of the response variables breaks the relationship between the features and the responses, these models serve as a null distribution that we could statistically compare the target models to. Specifically, eNetXplorer uses the proportion of runs where the target model performs better than the null models to establish an exact measure of statistical significance for the overall models and for each explanatory feature's coefficient (see Candia & Tsang, 2019 for details about this implementation).

Comparing the relative predictive ability across outcome measures

We evaluated the relative predictive ability of each feature set across outcome by statistically comparing the difference in correlation (Myers & Sirois, 2006). Specifically, we calculated the test statistic

$$z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

Where Z_r is the Fisher's z-transformation of the correlation. This test statistic can be compared to the normal distribution to measure statistical significance of the difference.

Comparing the relative predictive ability within outcome measures

To index the relative predictive ability of models within outcome measure, we created 5,000 bootstrap distributions of the Spearman's correlation between the predicted and actual values of the target variables between each pair of models predicting a given outcome measure. If the 95% confidence interval of the differences between these distributions did not include 0, we concluded that these models were statistically different from one another.

Decomposing unique variance in each model

Although our previous analyses investigated the relative predictive ability of the different kinds of feature sets, they cannot show whether the different kinds of feature sets uniquely predict variance in the outcome measure. Therefore, we next turned to the models that combined all feature sets to better characterize and decompose the specific features that specifically predicted each

outcome measure. To do so, we investigated and report significant coefficients from the full models.

Results

Exploratory Factor Analysis - Working Memory Span Measures

Bartlett's test of Sphericity ($\chi^2_{(55)} = 410.842$, $p = 6.490x10^{-56}$) and the Kaiser-Meyer-Olkin test (overall KMO = 0.82) indicated that our data (described in Table 4.1) were appropriate for exploratory factor analysis. Horn's parallel analysis (Figure 4.2A) suggested that a 2-factor solution was the best fit for our data and explained 34% of the variance in the data; the extracted 2-factor structure resulted in a good model fit (TLI = 0.969, RMSR = 0.05, RMSEA = 0.034 [0 0.069]). The first factor (which explained 17% of the variance in the data; Figure 4.2B) had salient loadings from Letter/Number Sequencing, Digit Span (Forward and Backward) and Digit Span (Sequencing); together, we interpreted this as Verbal Working Memory Capacity (Verbal WMC). The second factor (which explained an additional 17% of the variance in the data; Figure 4.2C), had salient loadings from the Lateralized Change Detection Task, Dot Probe Expectancy Task, Spatial Working Memory, Symbol Span Task, and Spatial Addition Task; we interpreted this factor as measuring Visual Working Memory Capacity (Visual WMC).

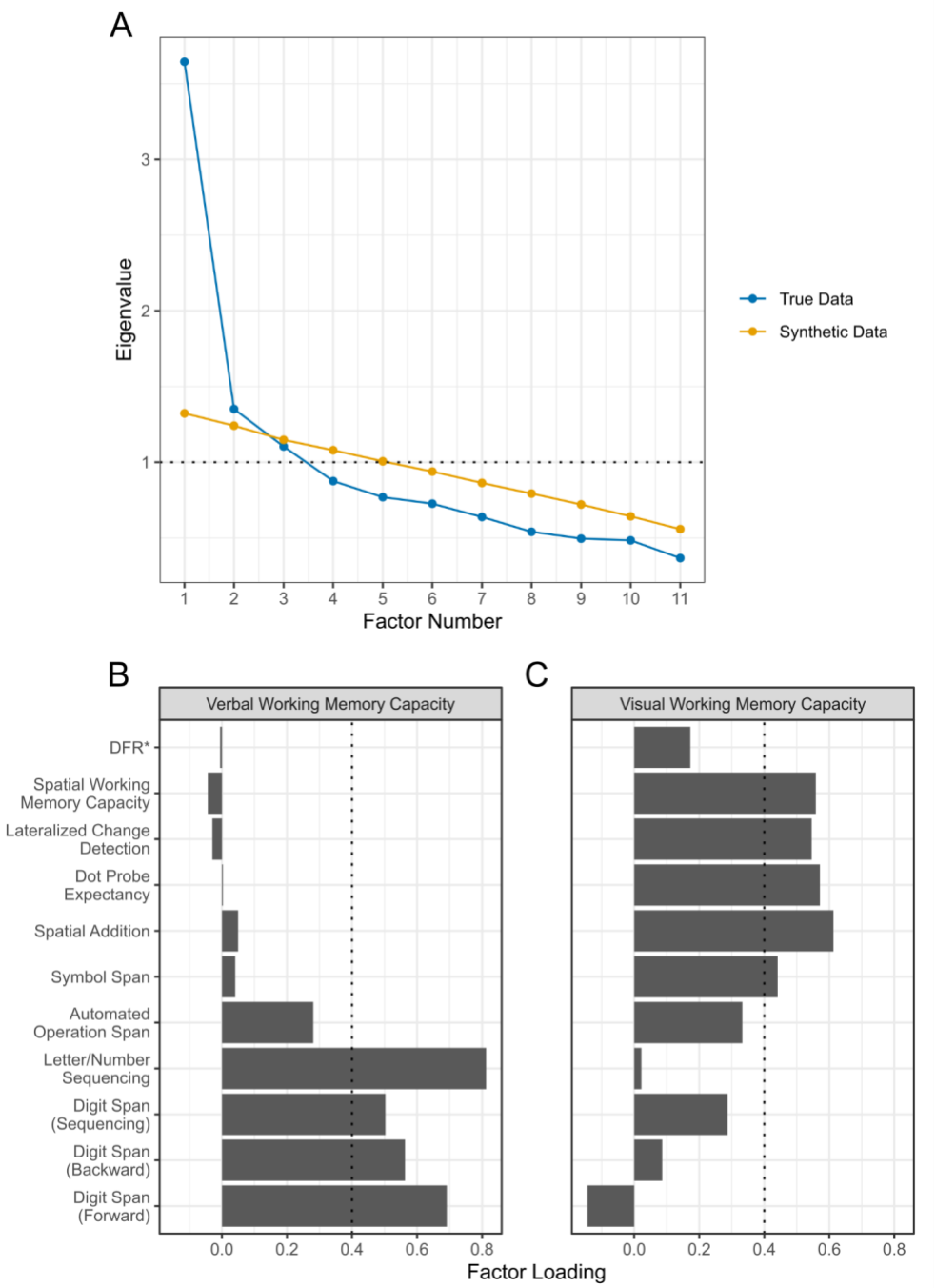


Figure 4.2: Exploratory factor analysis results. A: Results from Horn's parallel analysis suggested that a 2-factor solution was appropriate. B. The 2-factor solution revealed a Verbal Working Memory Capacity factor (including salient loadings for all Digit Span tasks and Letter Number Sequencing) and a Visual Working Memory Capacity factor (which included salient loadings of Symbol Span, Spatial Addition, Dot Probe Expectancy, Lateralized Change Detection, and Spatial Working Memory tasks). Dotted line at 0.4 reflects threshold for salient loadings. *Delayed Face Recognition score was collected from a separate session than the scanned fMRI DFR task.

Task Name	Missing	Median	Minimum	Maximum	Skew	Kurtosis	MSA
Digit Span (Forward)	0	0.143	-2.183	2.470	0.335	-0.581	0.763
Symbol Span	0	-0.026	-2.434	2.382	0.150	-0.491	0.801
Digit Span (Backward)	0	-0.139	-1.911	2.962	0.425	-0.376	0.858
Digit Span (Sequencing)	0	0.025	-2.328	3.319	0.517	0.560	0.847
Letter/Number Sequencing	0	-0.233	-2.854	3.138	0.093	0.069	0.807
Spatial Addition	0	0.251	-3.135	1.814	-0.771	-0.137	0.845
Lateralized Change Detection	2	0.033	-3.529	2.372	-0.131	0.102	0.821
Spatial Working Memory Capacity	0	0.353	-3.168	1.279	-0.771	0.092	0.836
Automated Operation Span	0	0.260	-3.380	1.203	-1.376	1.875	0.884
Dot-Probe Expectancy	0	0.146	-2.938	1.413	-0.822	0.426	0.832
Delayed Face Recognition*	1	0.002	-2.615	2.614	0.075	-0.348	0.535

Table 4.1: Descriptive statistics of behavioral measures. Note that all variables are z-scored before entry into the factor analysis, so each task has mean=0 and standard deviation=1. MSA = Measure of Sampling Adequacy. *Delayed Face Recognition score was collected from a separate session than the scanned fMRI DFR task.

There were moderate correlations between our outcome measures (Figure 4.3); Visual WMC and Verbal WMC were significantly correlated ($r = 0.37$, $t_{(167)} = 5.220$, $p = 9.450 \times 10^{-6}$). Additionally, performance on the Delayed Face Recognition (DFR) task was significantly correlated with both Visual WMC ($r = 0.25$, $t_{(167)} = 3.351$, $p = 0.0192$) and Verbal WMC ($r = 0.23$, $t_{(167)} = 2.997$, $p = 0.0377$). Visual WMC was significantly correlated with the BPRS ($r = -0.27$, $t_{(167)} = 3.661$, $p = 0.00539$). Finally, WHODAS and BPRS were significantly correlated ($r = 0.60$, $t_{(167)} = 9.740$, $p = 9.88 \times 10^{-17}$). All other correlations were not significant (all $p > 0.05$).

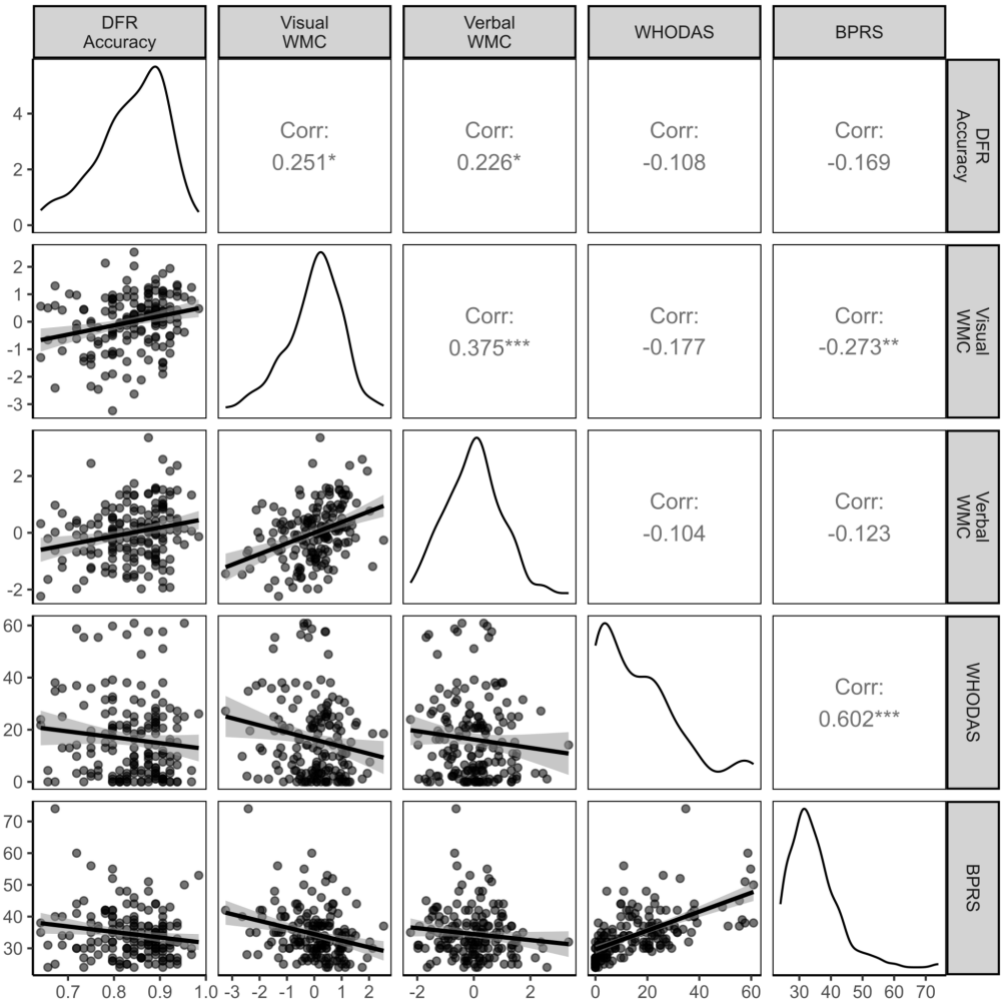


Figure 4.3: Relationship of outcome measures. Lower triangle shows scatter plots of each pair of outcome measures, upper triangle reports Pearson correlation value of the pairs of variables. Diagonal shows the distribution of values. Data are from N = 169 participants. Symbols in upper triangle reflect statistically significant correlations with Holm-Bonferroni corrections for multiple comparisons (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Delay Period Group Level Effects

Our first goal was to reduce our feature space for our downstream analyses by identifying regions that showed group-level effects in each of our task modalities, as previous work has shown that regions that show the largest group-level univariate effects also show the strongest brain-behavior correlations (Li et al., 2021).

Task fMRI: Univariate

We identified 50 regions that showed a significant Load 3 > Load 1 univariate effect during the delay period of our Delayed Face Recognition Task at the group level (Figure 4.4A). These regions were mostly found in the Control B Network (17 regions) and Control A network (10 regions), followed by the Dorsal Attention subnetworks (8 regions), Default subnetworks (6 regions), Salience/Ventral Attention subnetworks (5 regions), Control C network (3 regions) and Visual subnetworks (1 region). Notably, even though only 3/50 regions came from the Control C network, that constituted 25% of the total regions in the Control C network.

Task fMRI: Face-Specific Content

We next identified 45 regions that showed a significant Load 3 > Load 1 effects in face-specific representational content at the group level during the delay period (Figure 4.4B). That is, these regions show more face-specific content, as indexed by the relative difference in correlation between face and object "templates" when there are more faces to be held in memory. These regions were mostly found in the Visual subnetworks (14 regions), followed by the Dorsal Attention subnetworks (10 regions), Control A network (8 regions), Default subnetworks (4 regions), Control C network (3 regions), Salience/Ventral Attention subnetworks (3 regions), Control B network (2 regions) and Somato-Motor subnetworks (1 region).

Task fMRI: Encoding-to-Delay Pattern Stability

Finally, we identified 3 regions that showed a significant Load 3 > Load 1 representational stability from encoding-to-delay period at the group level (Figure 4.4C). Specifically, these regions showed stronger correlations between multivariate representations across task periods when there

were more faces to hold in memory. This analysis only resulted in 3 regions with significantly greater correlations at Load 3 during delay – two located in the Control A network (in prefrontal and parietal cortices), and one located in the Default subnetworks (in the medial parietal cortex).

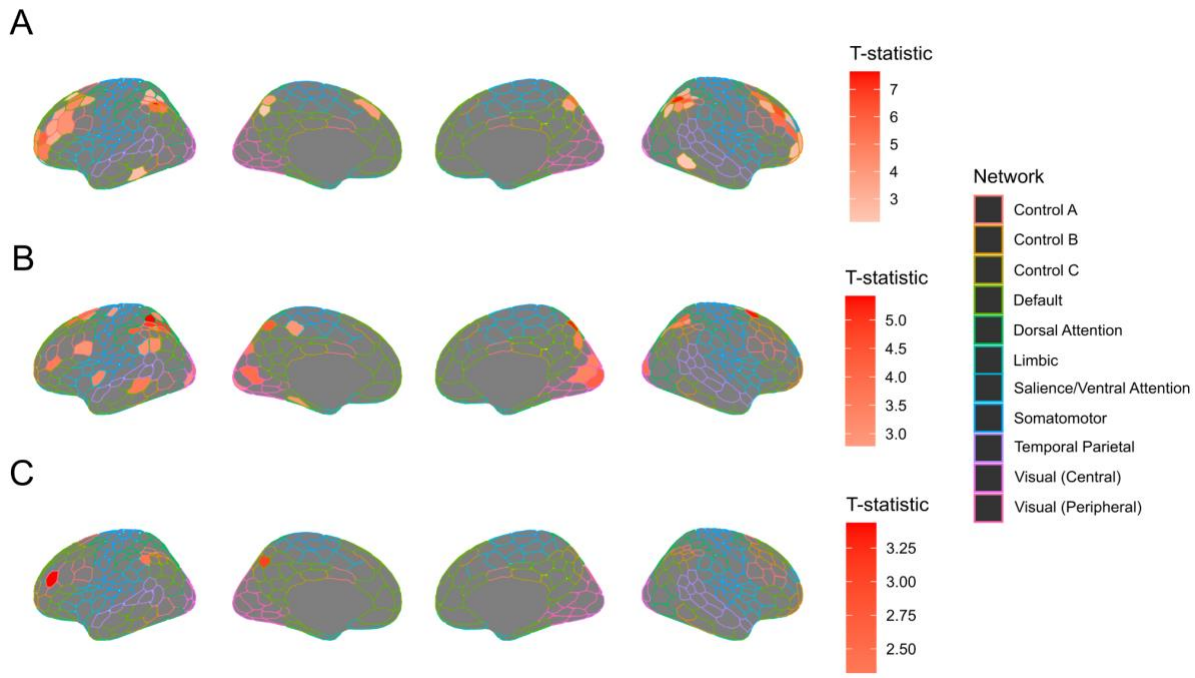


Figure 4.4: Group Load 3 > Load 1 effects during delay period for each task fMRI measure. A. Univariate beta values. B. Face-specific representational content. C. Encoding-to-Delay Pattern Stability. Values for each measure were extracted for N = 169 participants and a paired t-test was computed for each ROI. P-values were FDR corrected within each measure and regions with an FDR-corrected $p < 0.05$ are shown. The fill color of each non-grey ROI reflects t-statistic of the Load 3 > Load 1 t-test; outline color of each ROI reflects network assignment from the 17 Network Schaefer atlas, with subnetworks in Default, Dorsal Attention, Limbic, Salience/Ventral Attention, and Somatomotor networks collapsed.

Union of Regions

One intermediate goal was to identify a relatively restricted set of working memory related regions to use to predict behavior and cognition. To that end, we took the union of the regions that showed significant Load 3 > Load 1 effects across all three task fMRI analyses. This resulted in a set of 82 unique regions (Figure 4.5). These regions were mostly located in the Control B network (17 regions), Visual subnetworks (15 regions), Dorsal Attention subnetworks (14 regions) and

Control A network (14 regions), followed by the Default subnetworks (9 regions), Salience/Ventral Attention subnetworks (8 regions), Control C network (4 regions) and Somato-Motor subnetworks (1 region).

There were no individual regions that showed all three task fMRI effects – 14 regions showed significant effects in both univariate and face-specific representational content, while 2 regions showed significant effects in univariate and encoding-to-delay pattern stability. Intriguingly, the majority only showed a single effect – 34 regions only showed a univariate effect, 31 regions only showed an effect in face-specific representational content and 1 region only showed an effect in encoding-to-delay pattern stability. No region showed significant effects for both face-specific representational content and encoding-to-delay pattern stability.

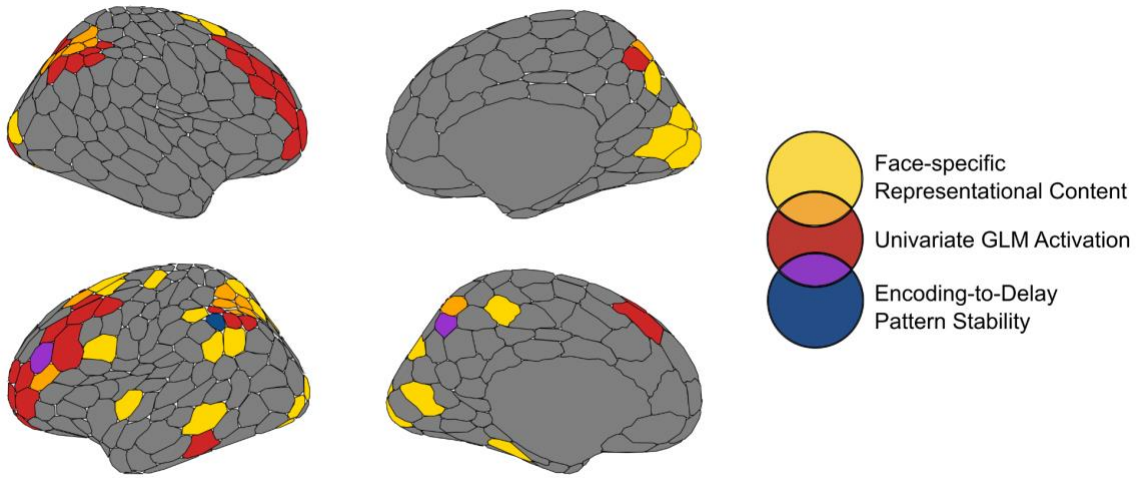


Figure 4.5: Union of task fMRI group level effects. Colored regions reflect ROIs that show at least one significant group-level task fMRI load effect. Yellow reflects Face-specific representational content effects, red reflects univariate activation effects, blue reflects encoding-to-delay pattern stability effects. Orange reflects regions that showed both face-specific representational content and univariate activation effects, while purple reflects regions that show both encoding-to-delay pattern stability and univariate activation effects. Data are from N = 169 participants.

Prediction for Modality Specific Models

Once we had established a set of candidate regions that are potentially involved in working memory, we developed a series of modality-specific models predicting each of our five outcome measures (Figure 4.6A).

Models including only task fMRI measures were able to significantly predict variance in DFR accuracy (Spearman's rho = 0.422, p = 0.000400), Visual WMC (Spearman's rho = 0.371, p = 0.00160), and BPRS (Spearman's rho = 0.181, p = 0.0464).

Models including only resting state fMRI measures were able to significantly predict variance in DFR accuracy (Spearman's rho = 0.186, p = 0.0400) and WHODAS (Spearman's rho = 0.199; p = 0.0232).

Models including only structural MRI measures were able to significantly predict variance in WHODAS (Spearman's rho = 0.199; p = 0.0148) and BPRS (Spearman's rho = 0.256, p = 0.002).

The model using only structural MRI measures captured more variance than the model using only resting state fMRI measures when predicting Visual WMC (mean bootstrapped difference = 0.022, 90% CI = [0.0187 0.0245]), WHODAS score (mean bootstrapped difference = 0.0361, 90% CI = [0.0335 0.0388]) and BPRS (mean bootstrapped difference = 0.0880, 90% CI = [0.0851 0.0908]). In contrast, the model using only structural MRI predicted less variance than the model using only resting state fMRI when predicting DFR performance (mean bootstrapped difference = 0.242, 90% CI = [0.240 0.245]). All other single modality models (including all models predicting Verbal WMC) were not significant (all p's < 0.05).

The task fMRI model could predict more variance than the resting state fMRI model (mean bootstrapped difference = 0.234, 90% CI = [0.233 0.238]) and structural MRI model (mean bootstrapped difference = 0.475, 90% CI = [0.472 0.478]) for DFR performance. A similar pattern of results was found for predicting Visual WMC, where the task fMRI could predict more variance than resting state fMRI (mean bootstrapped difference = 0.309, 90% CI = [0.306 0.312]) and structural fMRI (mean bootstrapped difference = 0.287, 90% CI = [0.284 0.289]).

The opposite pattern of results was found for predicting WHODAS where the task fMRI model predicted less variance than the resting state models (mean bootstrapped difference = 0.0631, 90% CI = [0.0600 0.0661]) and structural MRI models (mean bootstrapped difference = 0.0977, 90% CI = [0.0949 0.1004]).

The task fMRI model could predict more variance in BPRS than the resting state fMRI model (mean bootstrapped difference = 0.0135, 90% CI = [0.0107 0.0164]), but less than the structural MRI model (mean bootstrapped difference = 0.0757, 90% CI = [0.0729 0.0785]).

Decomposing task fMRI models

Decomposing the task fMRI modality into the specific kinds of measures can provide additional insight into the specific neural processes that are involved in predicting each outcome measure (Figure 4.6B). All individual task fMRI modality models could significantly predict DFR performance (univariate: Spearman's rho = 0.200, p = 0.0004; face-specific representational content: Spearman's rho = 0.166, p = 0.04998; encoding-to-delay pattern stability: Spearman's rho = 0.473, p = 0.0004). The univariate fMRI model was able to predict significantly more variance in the DFR performance than the face-specific representational content model (mean bootstrapped difference = 0.234, 90% CI = [0.231 0.236]), but less than the encoding-to-delay pattern stability

model (mean bootstrapped difference = 0.072, 90% CI = [0.0702 0.0742]). Additionally, the face-specific representational content model predicted significantly more variance in DFR performance than the encoding-to-delay pattern stability model (mean bootstrapped difference = 0.304, 90% CI = [0.301 0.306]).

Visual WMC was only predicted by the univariate fMRI (Spearman's rho = 0.201, p = 0.0340) and face-specific representational content (Spearman's rho = 0.405, p = 0.0004) models. Univariate fMRI features predicted significantly less variance in Visual WMC than the face-specific representational content model (mean bootstrapped difference = 0.196, 90% CI = [0.194 0.199]), but more than the encoding-to-delay pattern stability model (mean bootstrapped difference = 0.0690, 90% CI = [0.0662 0.0710]). Additionally, the face-specific representational content model predicted significantly more variance in Visual WMC than the encoding-to-delay pattern stability model (mean bootstrapped difference = 0.269, 90% CI = [0.266 0.272]).

WHODAS was only significantly predicted by the face-specific representational content model (Spearman's rho = 0.182, p = 0.0488). Univariate fMRI features were able to predict significantly less variance in the WHODAS score than the face-specific representational content model (mean bootstrapped difference = 0.140, 90% CI = [0.137 0.144]) and the encoding-to-delay pattern stability model (mean bootstrapped difference = 0.0467, 90% CI = [0.0440 0.0494]). Additionally, the face-specific representational content model predicted significantly more variance in WHODAS than the encoding-to-delay pattern stability model (mean bootstrapped difference = 0.0938, 90% CI = [0.0907 0.0968]).

Finally, BPRS was significantly predicted by face-specific representational content (Spearman's rho = 0.215, p = 0.0144) and encoding-to-delay pattern stability (Spearman's rho = 0.171, p = 0.0348). As with the WHODAS score, the univariate fMRI model was able to predict

significantly less variance in the BPRS than the face-specific representational content model (mean bootstrapped difference = 0.201, 90% CI = [0.199 0.204]) and the encoding-to-delay pattern stability model (mean bootstrapped difference = 0.155, 90% CI = [0.151 0.157]). Additionally, the face-specific representational content model predicted significantly more variance in BPRS than the pattern stability model (mean bootstrapped difference = 0.0440 90% CI = [0.0468 0.0412]).

All other specific task fMRI modality models (including all models predicting Verbal WMC) were not significant (all p's < 0.05).

Comparing prediction strength across outcome measures

Our univariate task fMRI measures could predict performance on the DFR better than Visual WMC (z = 1.962, p = 0.0497), WHODAS (z = 3.491, p = 0.00048) and BPRS (z = 3.744, p = 0.000191). Univariate task fMRI also trended towards predicting Visual WMC better than BPRS (z = 1.781, p = 0.0748). The face-specific representational content model was able to predict Visual WMC better than DFR performance (z = 2.385, p = 0.0170) and WHODAS (z = 2.241, p = 0.0250), with a trend towards being able to predict better than BPRS (z = 1.924, p = 0.0566).
Encoding-to-delay pattern stability was able to predict performance in the DFR better than Visual WMC (z = 3.441, p = 0.00058), WHODAS (z = 3.869, p = 0.00011) and BPRS (z = 3.107, p = 0.0019). When combining all individual task fMRI measures, we could predict DFR performance better than WHODAS (z = 2.860, p = 0.0042) and BPRS (z = 2.433, p = 0.0150).

Additionally, the model including task fMRI task measures could predict Visual WMC better than WHODAS (z = 2.310, p = 0.021) and trended towards predicting Visual WMC better than BPRS (z = 1.882, p =0.060). Finally, structural MRI measures were able to predict both

WHODAS and BPRS better than performance on the DFR task (WHODAS: $z = 2.692$, $p = 0.0071$; BPRS: $z = 2.899$, $p = 0.00375$). All other comparisons were not significant (all p-values > 0.05).

Prediction for Full Models

Although the individual modality models were each able to predict a range of variance for each outcome measure, it is possible that each modality utilizes similar sources of variance. One way to test whether each modality is representing unique variance in the outcome measure is to include all features in a single full model – if these full models predict more variance than the individual unique models, it suggests that the predictive ability of different modalities captures different aspects of variance in the outcome measures.

When features from all modalities were included in a single model (Figure 4.6C), we were able to significantly predict variance in DFR performance (Spearman's rho = 0.412, $p = 0.0008$), Visual WMC (Spearman's rho = 0.267, $p = 0.0132$), WHODAS (Spearman's rho = 0.403, $p = 0.0004$) and BPRS (Spearman's rho = 0.255, $p = 0.012$). There were no significant differences in predictive accuracy across outcomes (all p values > 0.1).

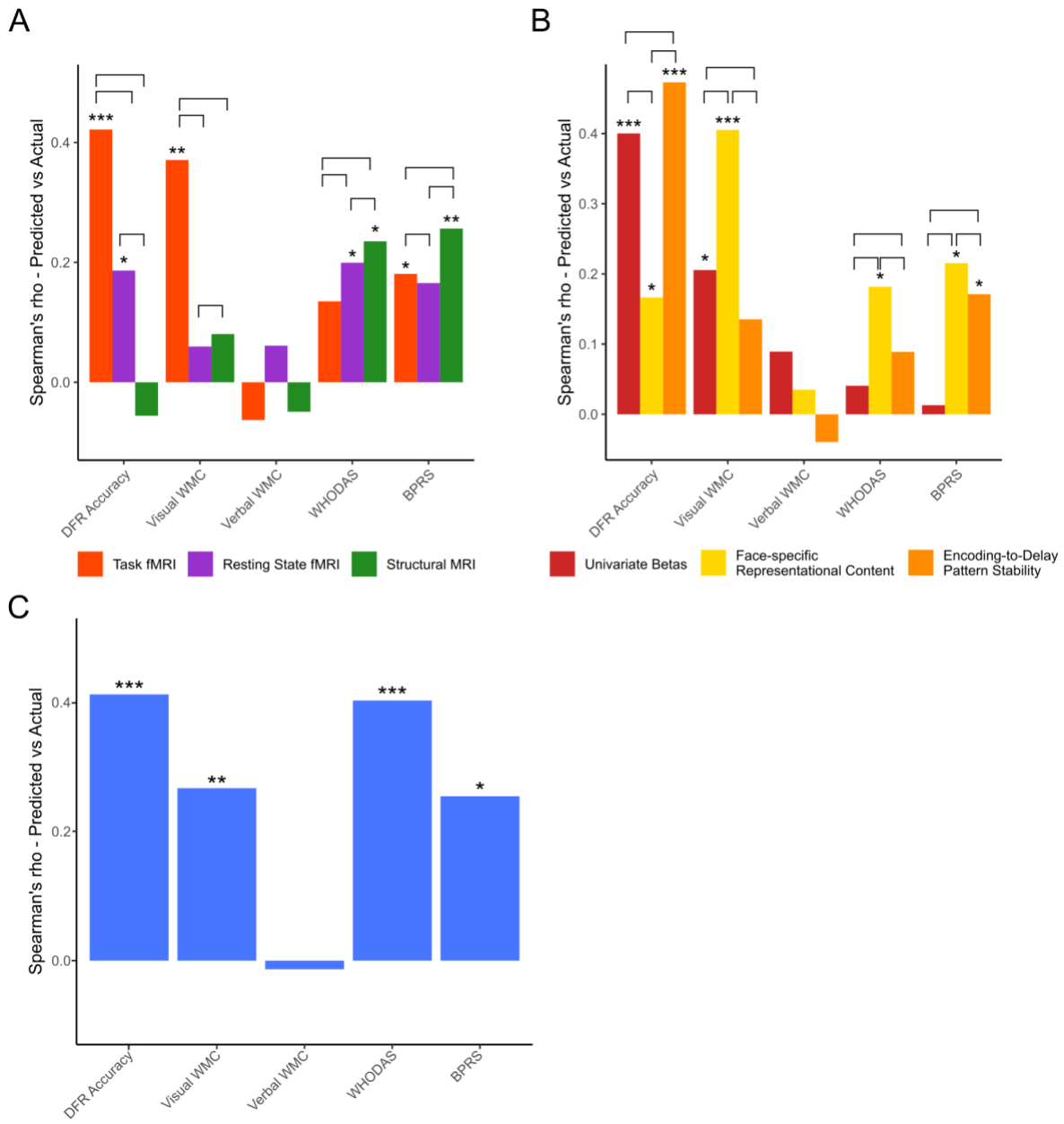


Figure 4.6: Predictive ability of models. Models using A. individual modalities, B. task fMRI measures, and C. all features. Each model was run in a 10-fold cross validation framework, where nine folds were used to train the models and one fold was held out as an independent testing set to ensure that the model generalized to new data. This cross-validation was repeated 100 times. Values depicted are the median Spearman's rho between the predicted and actual values across runs. Correlation values are calculated on data from all N = 168 participants concatenated across folds. Symbols reflect statistical significance derived from permutation testing, where the response variables were permuted 25 times per run to break the relationship between features and outcomes (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Brackets reflect significant differences across 5,000 bootstrapped iterations. Note that this figure does not include comparisons between outcome measures; see Figure 4.7 for comparisons of individual modalities to full models.

The full task model predicted significantly more variance in the DFR performance than the individual resting state fMRI model (mean bootstrapped difference = 0.224, 90% CI = [0.222 0.226] and the structural MRI model (mean bootstrapped difference = 0.466, 90% CI = [0.463 0.468]) (Figure 4.7A). Similarly, the full model could predict more variance in Visual WMC than the individual resting state fMRI model (mean bootstrapped difference = 0.205, 90% CI = [0.203 0.208] and the structural MRI model (mean bootstrapped difference = 0.185, 90% CI = [0.183 0.188]) (Figure 4.7B).

Intriguingly, the task fMRI model was able to predict significantly more variance than the full model for DFR performance (mean bootstrapped difference = 0.00823, 90% CI = [0.00712 0.00935]) and Visual WMC (mean bootstrapped difference = 0.1033, 90% CI = [0.101 0.104]). This indicates that adding additional feature modalities beyond task fMRI data (i.e. adding resting state and structural MRI measures) did not improve prediction of our two key indices of visual working memory ability, and indeed these additional features slightly diminished the models' predictive power.

The full model was able to predict more variance in the WHODAS (Figure 4.7C) than the task fMRI model (mean bootstrapped difference = 0.267, 90% CI = [0.265 0.269]), resting state fMRI model (mean bootstrapped difference = 0.203, 90% CI = [0.201 0.205]), and structural MRI model (mean bootstrapped difference = 0.168, 90% CI = [0.166 0.170]). Similarly, the full model predicted more variance in BPRS than the task fMRI model (mean bootstrapped difference = 0.0741, 90% CI = [0.0723 0.0760]) and resting state fMRI model (mean bootstrapped difference = 0.0883, 90% CI = [0.0863 0.0904]) (Figure 4.7D).

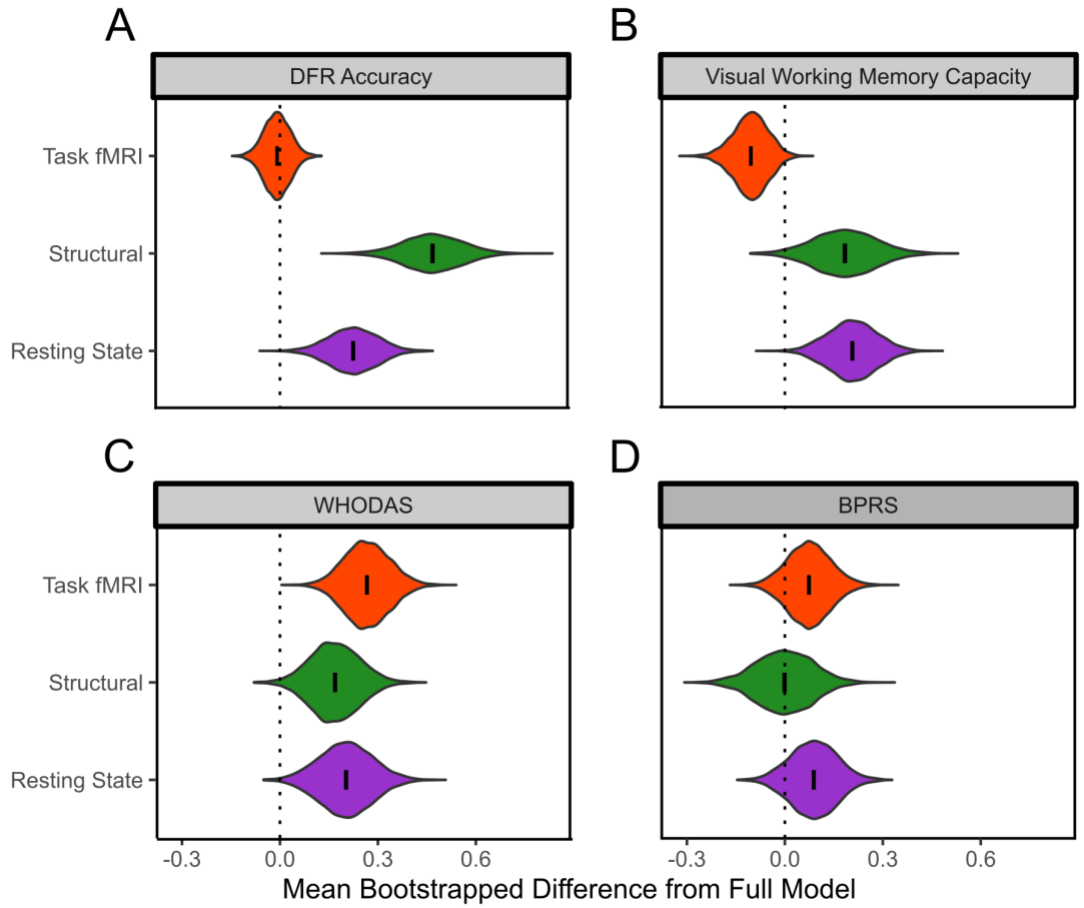


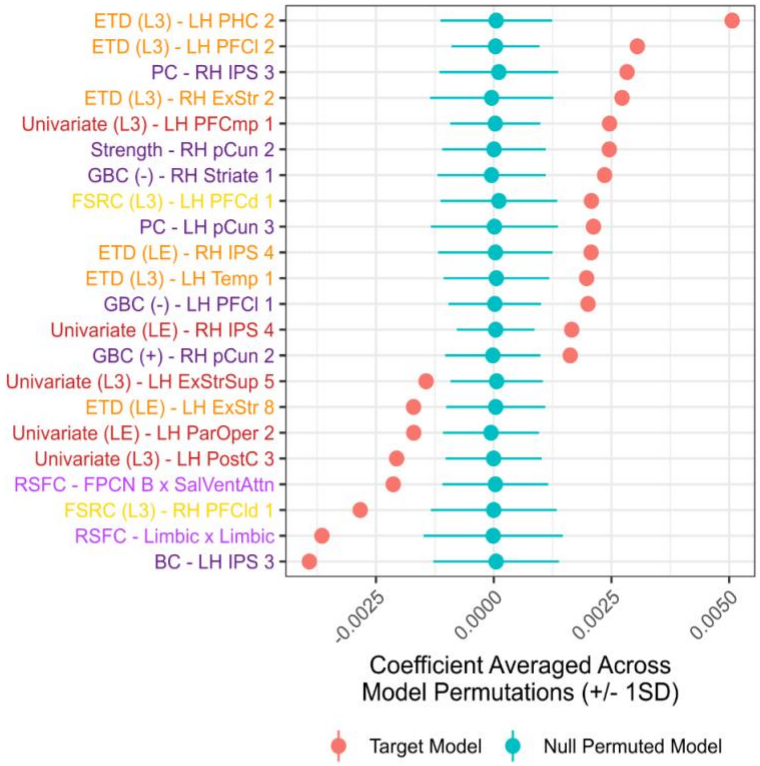
Figure 4.7: Bootstrap distributions of difference between individual modalities and full model for each outcome measure. Five thousand bootstrapped distributions of the Spearman's correlation between predicted and actual values of the target variables were created and the difference between the predictive ability of each outcome and the full model was calculated for each iteration. Violin plots reflect the distribution of these difference values across iterations; error bars reflect 95% confidence intervals of the difference values (all too small to be visible in these plots). Distributions where the full confidence interval was above zero (dotted line) show that the full model performed better than the individual modality; values where the full confidence interval was below zero show that the individual modality model performed better than the full model. All effects shown are significant except for the comparison of the structural model and full model's ability to predict BPRS scores (panel D; green plot).

Feature Importance in Full Models

All significant features for each model are visualized in Figures 4.8-4.11. For the full model predicting DFR performance (Figure 4.8), we saw that univariate activity and pattern stability task fMRI features in the Control and Visual networks showed significant predictive capacity. This included regions in the medial parietal and lateral prefrontal cortex, intraparietal sulcus, and

extrastriate and parahippocampal cortex. Measures of resting state network structure, mostly from the Control networks, also showed significant predictive capacity. These features included regions from the precuneus, lateral prefrontal cortex and striate visual cortex.

A


ETD (L3) - LH PHC 2
ETD (L3) - LH PFCl 2
PC - RH IPS 3
ETD (L3) - RH ExStr 2
Univariate (L3) - LH PFCmp 1
Strength - RH pCun 2
GBC (-) - RH Striate 1
FSRC (L3) - LH PFCd 1
PC - LH pCun 3
ETD (LE) - RH IPS 4
ETD (L3) - LH Temp 1
GBC (-) - LH PFCl 1
Univariate (LE) - RH IPS 4
GBC (+) - RH pCun 2
Univariate (L3) - LH ExStrSup 5
ETD (LE) - LH ExStr 8
Univariate (LE) - LH ParOper 2
Univariate (L3) - LH PostC 3
RSFC - FPCN B x SalVentAttn
FSRC (L3) - RH PFCld 1
RSFC - Limbic x Limbic
BC - LH IPS 3
-0.0025
0.0000
0.0025
0.0050
Coefficient Averaged Across
Model Permutations (+/- 1SD)
Target Model
Null Permuted Model

B

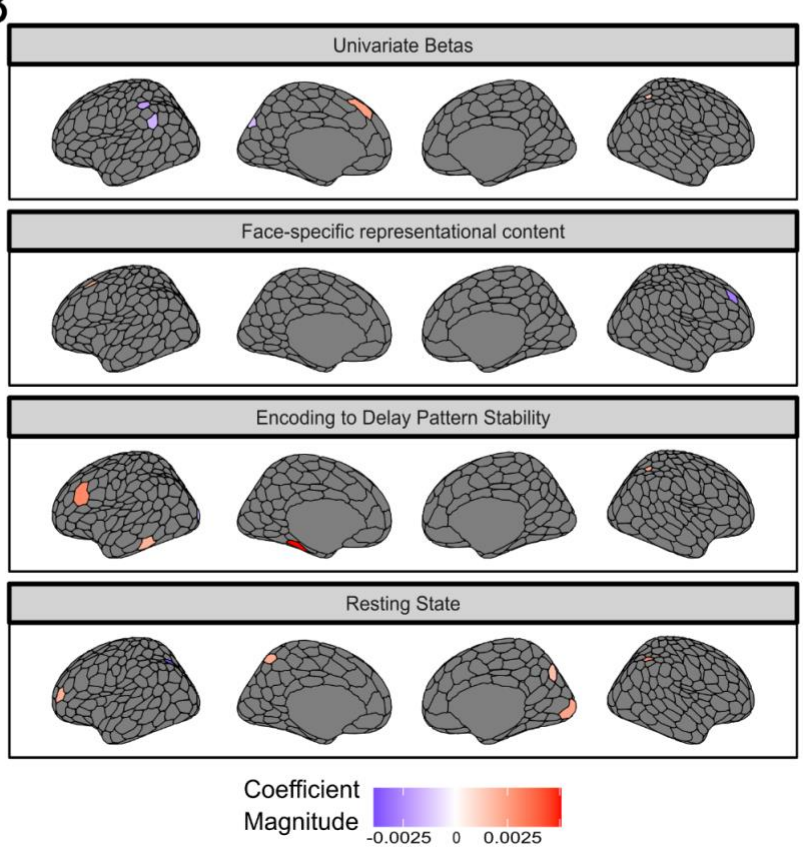

Univariate Betas
Face-specific representational content
Encoding to Delay Pattern Stability
Resting State
Coefficient
Magnitude
-0.0025
0
0.0025

Figure 4.8: Significant features predicting DFR performance. A. Significant feature coefficients from full models that included all potential features. Values reflect the average of non-zero model coefficients across folds and runs, weighted by the proportion of models across runs that the feature was included in. Statistical significance determined through permutation testing. Error bars reflect +/- 1 standard deviation of the coefficients across runs. Pink reflects the target models and turquoise reflects the coefficients from the null models. ROI labels are labels from 400 ROI parcellation of the 17 Network Schaefer atlas (Schaefer et al., 2018). Color of labels represent feature modality - red: univariate GLM activity, orange: encoding-to-delay pattern stability, yellow: face-specific representational content, dark purple: resting state network structure measures, light purple: resting state network-average functional connectivity, green: structural MRI measures, pink: demographic measures. Abbreviations: FSRC: face-specific representational content; ETD: encoding-to-delay pattern stability; PC: participation coefficient; BC: betweenness centrality; GBC (+): global brain connectivity, positive connections; GBC (-): global brain connectivity, negative connections. B. Significant coefficients from individual ROIs projected onto the brain using the ggseg R package.

The full model predicting Visual WMC (Figure 4.9), in contrast, was strongly dominated by face-specific representational content features across the Control (particularly the Control B network), Visual, and Dorsal Attention subnetworks. Specifically, this included regions in the lateral prefrontal and superior parietal cortices, in addition to the extrastriate visual cortex. Additionally, network structure measures derived from resting state fMRI from the Control and Dorsal Attention subnetworks also significantly predicted variance in Visual WMC. This included regions in the dorsal and lateral prefrontal cortex, parietal cortex, striate visual cortex and post-central gyrus. Age was also a significant negative predictor of Visual WMC.

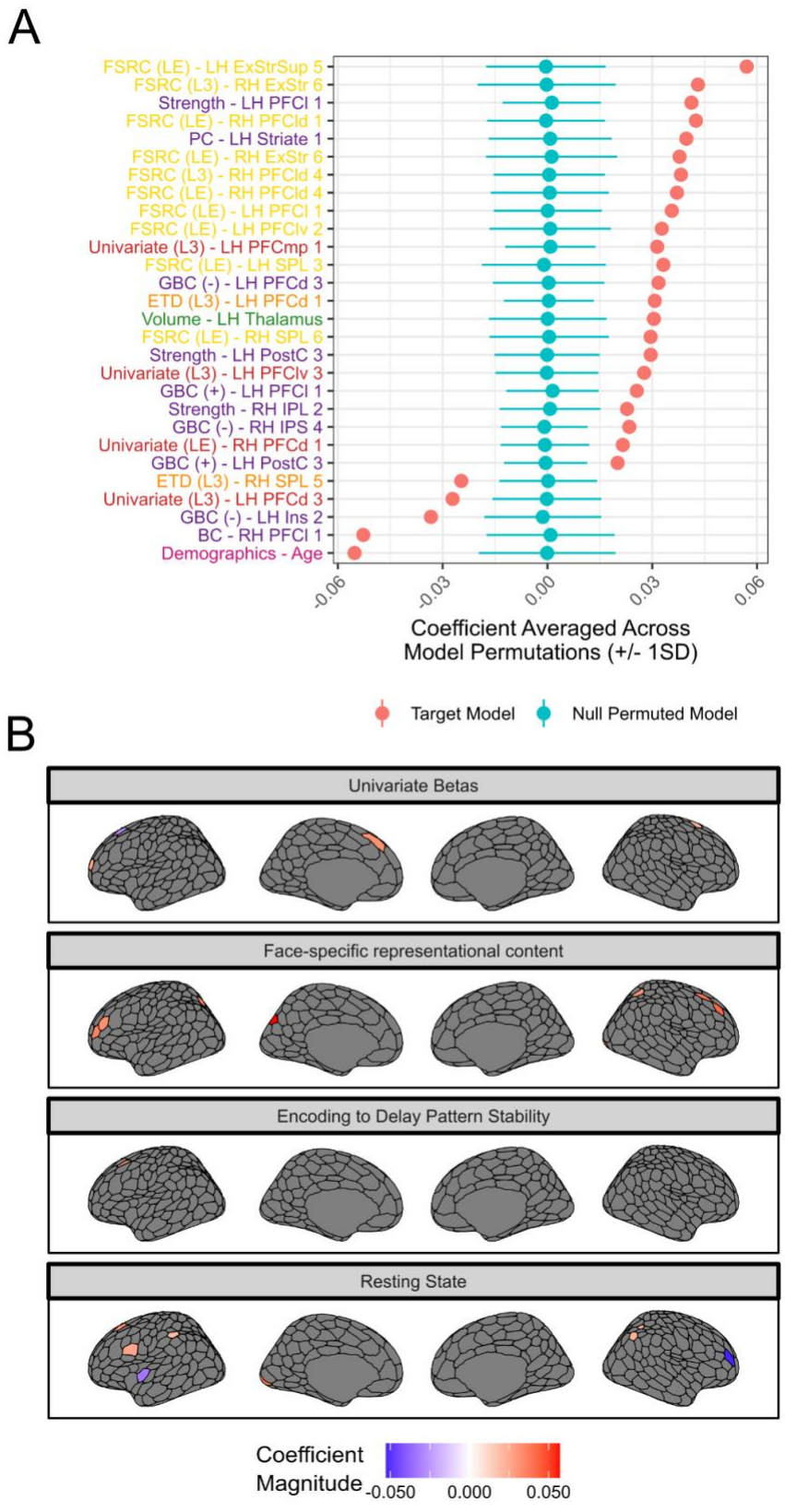
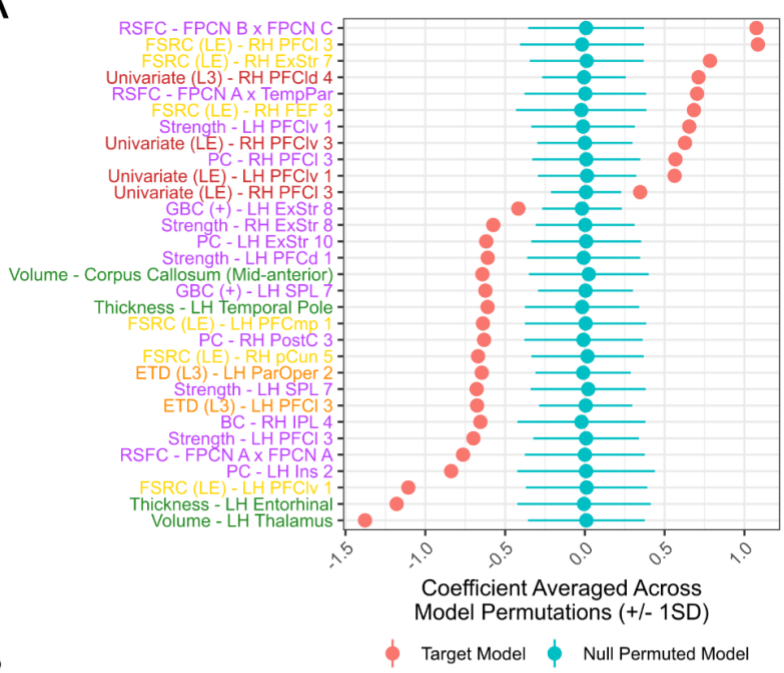

A
FSRC (LE) - LH ExStrSup 5
FSRC (L3) - RH ExStr 6
Strength - LH PFCl 1
FSRC (LE) - RH PFCld 1
PC - LH Striate 1
FSRC (LE) - RH ExStr 6
FSRC (L3) - RH PFCld 4
FSRC (LE) - RH PFCld 4
FSRC (LE) - LH PFCl 1
FSRC (LE) - LH PFClv 2
Univariate (L3) - LH PFCmp 1
FSRC (LE) - LH SPL 3
GBC (-) - LH PFCd 3
ETD (L3) - LH PFCd 1
Volume - LH Thalamus
FSRC (LE) - RH SPL 6
Strength - LH PostC 3
Univariate (L3) - LH PFClv 3
GBC (+) - LH PFCl 1
Strength - RH IPL 2
GBC (-) - RH IPS 4
Univariate (LE) - RH PFCd 1
GBC (+) - LH PostC 3
ETD (L3) - RH SPL 5
Univariate (L3) - LH PFCd 3
GBC (-) - LH Ins 2
BC - RH PFCl 1
Demographics - Age
-0.06
-0.03
0.00
0.03
0.06
Coefficient Averaged Across
Model Permutations (+/- 1SD)
Target Model
Null Permuted Model
B
Univariate Betas
Face-specific representational content
Encoding to Delay Pattern Stability
Resting State
Coefficient
Magnitude
-0.050
0.000
0.050

Figure 4.9: Significant features predicting Visual Working Memory Capacity. A. Significant feature coefficients from full models that included all potential features. Values reflect the average of non-zero model coefficients across folds and runs, weighted by the proportion of models across runs that the feature was included in. Statistical significance determined through permutation testing. Error bars reflect +/- 1 standard deviation of the coefficients across runs. Pink reflects the target models and turquoise reflects the coefficients from the null models. ROI labels are labels from 400 ROI parcellation of the 17 Network Schaefer atlas (Schaefer et al., 2018). Color of labels represent feature modality - red: univariate GLM activity, orange: encoding-to-delay pattern stability, yellow: face-specific representational content, dark purple: resting state network structure measures, light purple: resting state network-average functional connectivity, green: structural MRI measures, pink: demographic measures. Abbreviations: FSRC: face-specific representational content; ETD: encoding-to-delay pattern stability; PC: participation coefficient; BC: betweenness centrality; GBC (+): global brain connectivity, positive connections; GBC (-): global brain connectivity, negative connections. B. Significant coefficients from individual ROIs projected onto the brain using the ggseg R package.

WHODAS score (Figure 4.10) was mostly predicted by features indexing network structure from regions in the Visual, Dorsal Attention, Salience/Ventral Attention and Control subnetworks during resting state fMRI. Specifically, this included regions in the extrastriate visual, dorsal and lateral prefrontal and parietal cortices. Network average functional connectivity within the Control subnetworks was also significantly predictive of WHODAS. Additionally, structural MRI measures (both subcortical volume of the corpus callosum and left thalamus and cortical thickness of the entorhinal cortex and temporal pole) were significant predictors of WHODAS. While the features derived from task fMRI were less important for predicting WHODAS, face-specific representational content in the Control subnetworks and pattern stability in the Salience/Ventral Attention networks were also significant predictors. This included regions in the medial parietal and ventrolateral prefrontal cortex and precuneus. Univariate task fMRI in the lateral prefrontal cortex also significantly positively predicted WHODAS scores (although note that a positive WHODAS score reflects worse functioning).

A


RSFC - FPCN B x FPCN C
FSRC (LE) - RH PFCl 3
FSRC (LE) - RH ExStr 7
Univariate (L3) - RH PFCld 4
RSFC - FPCN A x TempPar
FSRC (LE) - RH FEF 3
Strength - LH PFClv 1
Univariate (LE) - RH PFClv 3
PC - RH PFCl 3
Univariate (LE) - LH PFClv 1
Univariate (LE) - RH PFCl 3
GBC (+) - LH ExStr 8
Strength - RH ExStr 8
PC - LH ExStr 10
Strength - LH PFCd 1
Volume - Corpus Callosum (Mid-anterior)
GBC (+) - LH SPL 7
Thickness - LH Temporal Pole
FSRC (LE) - LH PFCmp 1
PC - RH PostC 3
FSRC (LE) - RH pCun 5
ETD (L3) - LH ParOper 2
Strength - LH SPL 7
ETD (L3) - LH PFCl 3
BC - RH IPL 4
Strength - LH PFCl 3
RSFC - FPCN A x FPCN A
PC - LH Ins 2
FSRC (LE) - LH PFClv 1
Thickness - LH Entorhinal
Volume - LH Thalamus
-1.5
-1.0
-0.5
0.0
0.5
1.0
Coefficient Averaged Across
Model Permutations (+/- 1SD)
Target Model
Null Permuted Model

B

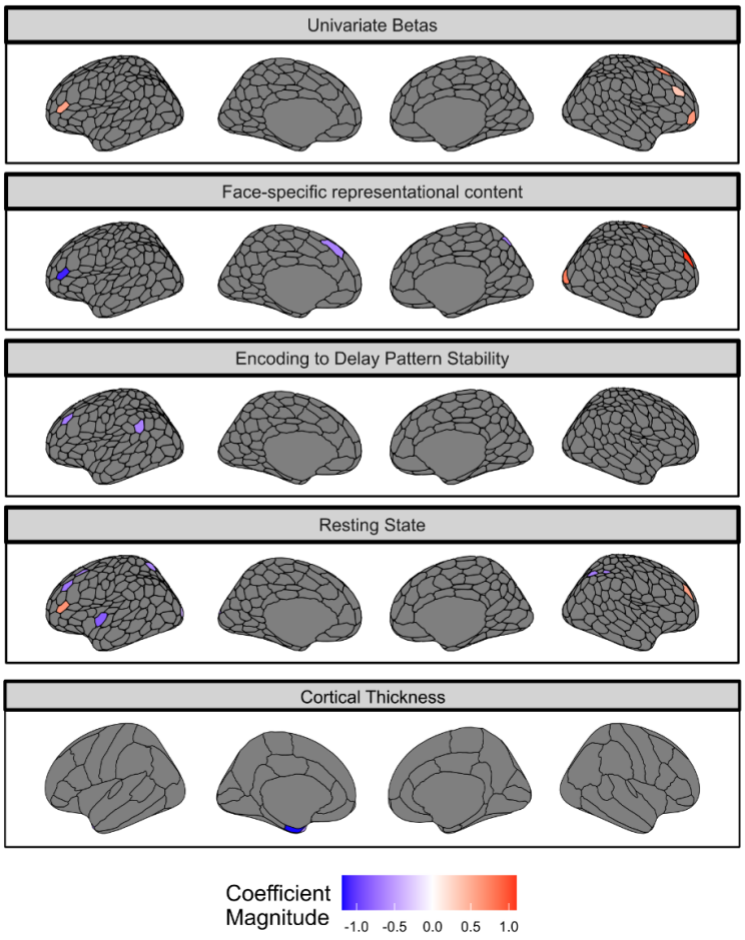

Univariate Betas
Face-specific representational content
Encoding to Delay Pattern Stability
Resting State
Cortical Thickness
Coefficient
Magnitude
-1.0
-0.5
0.0
0.5
1.0

Figure 4.10: Significant features predicting WHODAS. A. Significant feature coefficients from full models that included all potential features. Values reflect the average of non-zero model coefficients across folds and runs, weighted by the proportion of models across runs that the feature was included in. Statistical significance determined through permutation testing. Error bars reflect +/- 1 standard deviation of the coefficients across runs. Pink reflects the target models and turquoise reflects the coefficients from the null models. ROI labels are labels from 400 ROI parcellation of the 17 Network Schaefer atlas (Schaefer et al., 2018). Color of labels represent feature modality - red: univariate GLM activity, orange: encoding-to-delay pattern stability, yellow: face-specific representational content, dark purple: resting state network structure measures, light purple: resting state network-average functional connectivity, green: structural MRI measures, pink: demographic measures. Abbreviations: FSRC: face-specific representational content; ETD: encoding-to-delay pattern stability; PC: participation coefficient; BC: betweenness centrality; GBC (+): global brain connectivity, positive connections; GBC (-): global brain connectivity, negative connections. B. Significant coefficients from individual ROIs projected onto the brain using the ggseg R package. Note that for WHODAS, higher values reflect worse functioning, so negative coefficients should be interpreted as being associated with better psychiatric functioning.

Similarly, BPRS (Figure 4.11) was strongly predicted by network structure measures derived from resting state fMRI, particularly the participation coefficient and betweenness in the Control subnetworks. This included regions in the intraparietal sulcus, inferior parietal and lateral prefrontal cortices, the precuneus and the insula. Subcortical volume in the corpus callosum and left thalamus, pattern stability in the parietal cortex and insula, and face-specific representational content features in the extrastriate visual cortex, frontal eye fields and precuneus also significantly predicted BPRS.

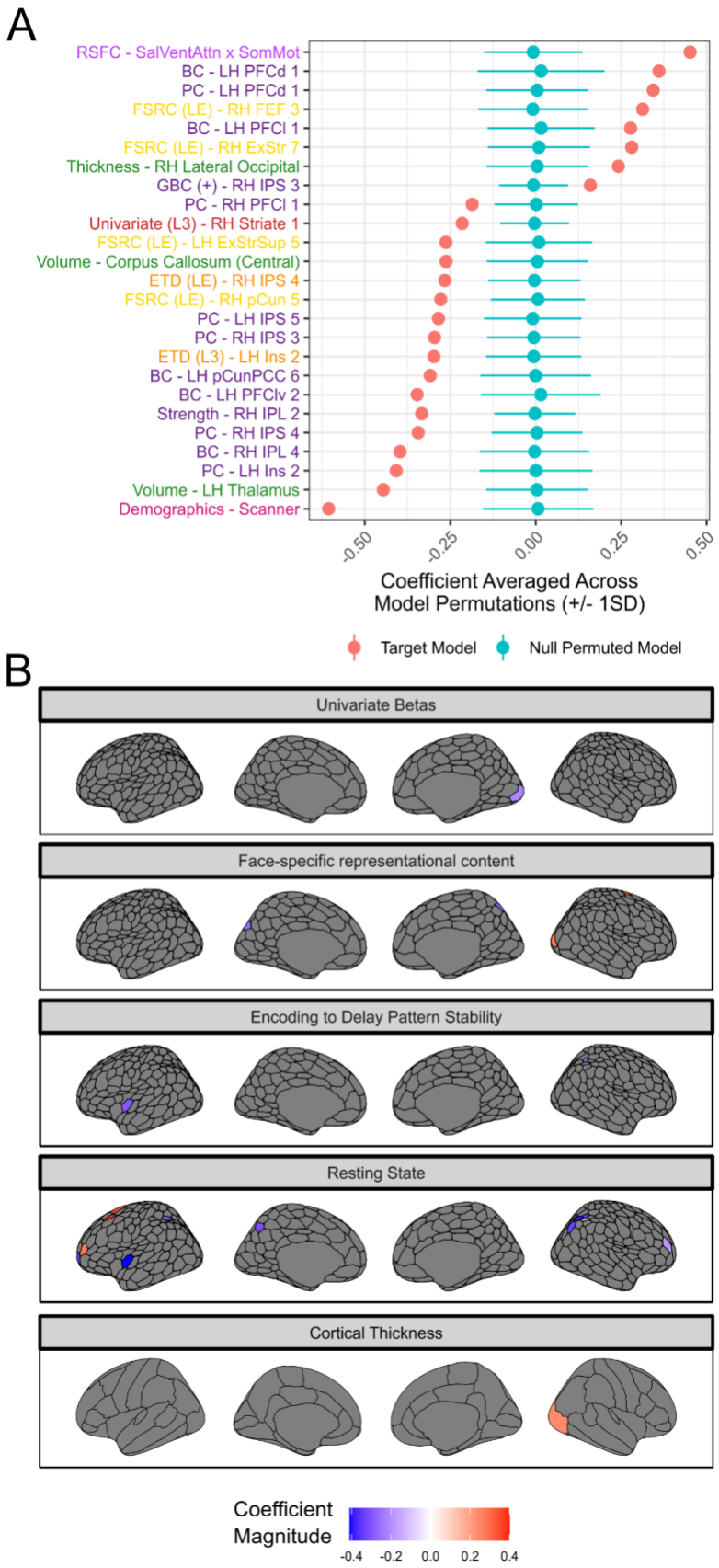

A
RSFC - SalVentAttn x SomMot
BC - LH PFCd 1
PC - LH PFCd 1
FSRC (LE) - RH FEF 3
BC - LH PFCl 1
FSRC (LE) - RH ExStr 7
Thickness - RH Lateral Occipital
GBC (+) - RH IPS 3
PC - RH PFCl 1
Univariate (L3) - RH Striate 1
FSRC (LE) - LH ExStrSup 5
Volume - Corpus Callosum (Central)
ETD (LE) - RH IPS 4
FSRC (LE) - RH pCun 5
PC - LH IPS 5
PC - RH IPS 3
ETD (L3) - LH Ins 2
BC - LH pCunPCC 6
BC - LH PFClv 2
Strength - RH IPL 2
PC - RH IPS 4
BC - RH IPL 4
PC - LH Ins 2
Volume - LH Thalamus
Demographics - Scanner
-0.50
-0.25
0.00
0.25
0.50
Coefficient Averaged Across
Model Permutations (+/- 1SD)
Target Model
Null Permuted Model
B
Univariate Betas
Face-specific representational content
Encoding to Delay Pattern Stability
Resting State
Cortical Thickness
Coefficient
Magnitude
-0.4
-0.2
0.0
0.2
0.4

Figure 4.11: Significant features predicting BPRS. A. Significant feature coefficients from full models that included all potential features. Values reflect the average of non-zero model coefficients across folds and runs, weighted by the proportion of models across runs that the feature was included in. Statistical significance determined through permutation testing. Error bars reflect +/- 1 standard deviation of the coefficients across runs. Pink reflects the target models and turquoise reflects the coefficients from the null models. ROI labels are labels from 400 ROI parcellation of the 17 Network Schaefer atlas (Schaefer et al., 2018). Color of labels represent feature modality - red: univariate GLM activity, orange: encoding-to-delay pattern stability, yellow: face-specific representational content, dark purple: resting state network structure measures, light purple: resting state network-average functional connectivity, green: structural MRI measures, pink: demographic measures. Abbreviations: FSRC: face-specific representational content; ETD: encoding-to-delay pattern stability; PC: participation coefficient; BC: betweenness centrality; GBC (+): global brain connectivity, positive connections; GBC (-): global brain connectivity, negative connections. B. Significant coefficients from individual ROIs projected onto the brain using the ggseg R package. Note that for BPRS, higher values reflect worse functioning, so negative coefficients should be interpreted as being associated with better psychiatric functioning.

Comparison of the relative proportions of significant features in the full models for each outcome measure (Figure 4.12) shows similar results to the predictive ability of the individual modality models. For instance, DFR performance showed relatively more significant predictors from the encoding-to-delay pattern stability task fMRI modality, while the full model predicting Visual WMC showed more significant features from the face-specific representational content task fMRI modality. An important caveat for this analysis, however, is that there are a different number of features from each modality included in the full models (left-most bar in Figure 4.12), so the presence of more features from one such category may be biased by the over-representation in the base rates.

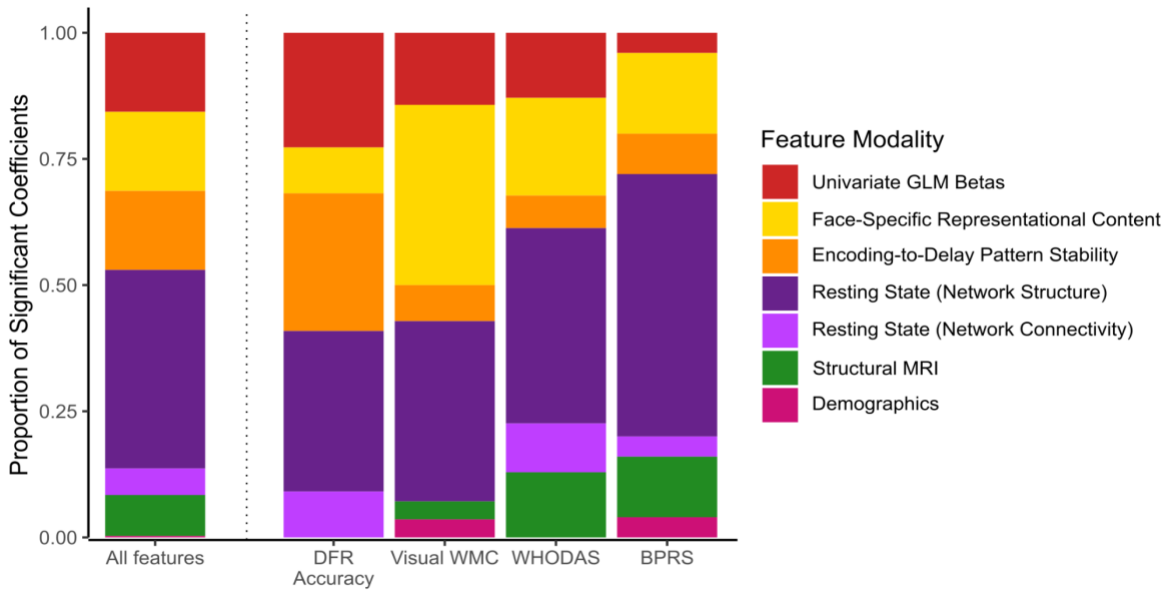


Figure 4.12: Distribution of feature modalities in full models. Proportion of significant features from full models for each outcome measure belonging to each feature modality. Statistical significance determined through permutation testing. Proportions relative to number of features for each individual model. Bar to left of dotted line depicts distribution of feature modalities across all potential features. Colors represent feature modality - red: univariate GLM activity, orange: encoding-to-delay pattern stability, yellow: face-specific representational content, dark purple: resting state network structure measures, light purple: resting state network-average functional connectivity, green: structural MRI measures, pink: demographic measures.

Discussion

In the present study, our overarching goal was to be able to predict individual differences across in-scanner working memory performance, as measured by a delayed match-to-sample task with faces, and out-of-scanner measures of working memory capacity and psychiatric functioning. To accomplish this, we took two approaches – first, we compared the predictive ability of different aspects of brain structure and function, including multivariate pattern representations during a working memory task. Second, we combined all features into a single model for each outcome measure to determine whether the different kinds of features reflected unique variance and investigated the regional distributions of each kind of feature.

Before we could compare the relative utility of each modality, we first had to define a set of candidate regions that may be implicated in working memory processes. Following existing

work that shows that the strongest brain-behavior correlations are found in regions with the largest between-subjects group effects (Li et al., 2021), we identified regions that showed group level Load 3 > Load 1 effects during the delay period of a delayed match-to-sample task. We additionally included measures of multivariate pattern similarity, including an index of face-specific representational content (i.e. how much a region's BOLD activity pattern during the delay period matched a face-specific template derived from an independent face localizer scan) and a trial-by-trial level measure of pattern stability across the task (i.e. how much a region's BOLD activity pattern during face encoding was preserved during the delay period). Although there has been a range of literature showing that task-relevant visual information can be decoded during maintenance across the brain (Albers et al., 2013; Christophel et al., 2012; Harrison & Tong, 2009; Yu & Shim, 2017) and that the specificity of this information when presented with distractors is related to behavioral performance (Bettencourt & Xu, 2016; Hallenbeck et al., 2021), it is still unclear whether the differences in this fidelity across working memory loads (i.e. as there are more items to hold in memory) is related to behavioral performance.

Both univariate and face-specific representational content analyses showed a wide array of load effects across the brain. We identified univariate load effects in bilateral parietal and PFC, consistent with regions previously shown to be modulated by load during working memory tasks (D'Esposito & Postle, 2012; Li et al., 2022; Manoach et al., 1997; Rypma & D'Esposito, 1999). Load effects in face-specific representational content during delay period were mostly found in occipital and parietal cortex, with some additional regions in dorsal and lateral PFC; these regions are largely in accordance with prior work decoding stimulus-specific information (Bettencourt & Xu, 2016; Velenosi et al., 2020; Yu & Shim, 2017).

Although there was a widespread distribution of regions for the univariate and face-specific representational content effects, only three regions showed load effects in pattern stability across encoding and delay. That is, only three regions (two in the bilateral parietal cortex and one in left lateral PFC) showed significantly stronger preservation of encoding-induced activity patterns during the ensuing delay period on trials requiring the maintenance of three faces rather than a single face. Intriguingly, there was also no region that showed all three task-related load effects, although there were regions across the bilateral parietal cortex and, to a lesser extent, the left lateral PFC that showed two out of the three effects. The potential roles of each of these regions will be discussed in detail below.

It is important to note, however, that even though we focus on load effects, there may be aspects of these processes (particularly the multivariate pattern effects) that do not scale with increasing memory load. For instance, even though there were relatively few regions that showed load effects in pattern stability across the phases of the task, it is possible that the stability of pattern information, irrespective of load, may still play an important role in predicting performance and visual working memory capacity. To explore this possibility, we also included univariate activation, face-specific representational content, and pattern stability from high load trials as additional features for each of our models.

We developed a series of models with three sets of features (task fMRI features, resting state fMRI features, and structural MRI features) using ElasticNet regularized regression (Zou & Hastie, 2005) in a 10-fold cross-validation framework to compare the relative predictive ability of each kind of feature for five different outcomes: (1) performance on a delayed face recognition (DFR) task, (2) Visual Working Memory Capacity (WMC), (3) Verbal WMC, (4) overall well-

being, as indexed by the World Health Organization Disability Assessment Scale 2.0 (WHODAS), and (5) psychiatric function, as indexed by the Brief Psychiatric Rating Scale (BPRS).

With respect to working memory ability, we found that DFR performance could be predicted by both task-based and resting state fMRI measures, while Visual WMC could only be predicted by task-based fMRI measures. The degree to which we were able to predict variance in working memory ability from task-based fMRI measures was in line with reports from several recent studies that attempted similar analyses on large data sets (Pat et al., 2022; Tetereva et al., 2022). Our somewhat stronger prediction of DFR performance (Spearman's rho = 0.42) than Visual WMC (Spearman's rho = 0.37) is likely attributable to the fact that our task-based fMRI measures were derived from the DFR task itself and thus intrinsically linked to participants' performance on that task.

Notably, resting state fMRI measures were only predictive of DFR performance and WHODAS scores. Although the predictive ability of these models is in line with the predictive ability from other studies using resting state functional connectivity (Rasero et al., 2021; Tetereva et al., 2022), task fMRI still was significantly better at predicting DFR performance than was resting state fMRI. It is likely that task-based measures may be better at highlighting individual differences in cognitive behavior and as such, may be more optimal for use in predicting behavior (Finn, 2021; Finn et al., 2017; Geerligs et al., 2015).

We showed mixed patterns of results when it came to predicting overall well-being and psychiatric function. WHODAS scores could be significantly predicted by resting state fMRI and structural MRI models while BPRS scores were predicted by task fMRI and structural MRI measures. Recent attempts to predict psychiatric function and mental health outcomes have found varied results – one study suggested that resting state fMRI outperforms task-based fMRI when

predicting mental health measures (Chen et al., 2022), while another found that only structural connectivity predicted mental health (Mansour et al., 2021). Future work will be necessary to better characterize the dissociations between these two highly correlated (Pearson's r = 0.60), yet still meaningfully distinct constructs.

Intriguingly, none of our models were able to predict significant variance in Verbal WMC. There has been a wide literature debating the functional organization of working memory and whether there is an underlying domain-general working memory process that may subserve the maintenance of both visual and verbal stimuli (Baddeley & Hitch, 1974; Li et al., 2014). Recent theoretical work has suggested that working memory capacity may be partially explained by the ability of lateral PFC to create task-relevant functional networks (Minamoto et al., 2017). Although our Visual WMC and Verbal WMC measures showed a moderate correlation with each other (Pearson's r = 0.37), it is possible that by identifying regions of interest from a visual working memory task, we excluded potential regions that may be uniquely crucial for predicting Verbal WMC. It is also possible that our Verbal WMC construct was not as robustly estimated from our battery of working memory tasks as was our Visual WMC construct. Our Verbal WMC factor scores were mostly driven by performance on digit span tests and letter/number sequencing, but our task battery did not include other commonly used verbal WMC assessments like reading span and sentence span (Waters & Caplan, 1996, 2003).

We next decomposed the task fMRI models into separate models for each of our three classes of task-based effects (univariate activity levels, face-specific representational content, and encoding-to-delay pattern stability) to determine whether these different ways of indexing working memory related brain signals predicted distinct outcome measures. In this analysis, we showed a striking dissociation – although all 3 indices could predict DFR performance, it was most strongly

predicted by the univariate activity and encoding-to-delay pattern stability models. In contrast, Visual WMC was most strongly predicted by the univariate activity and face-specific representational content models, but not the pattern stability model.

Although our Visual WMC and DFR performance measures were significantly correlated (Pearson's r = 0.25), there are important differences between these two outcome measures. An individual's underlying trait working memory capacity will certainly influence their performance on a given working memory task; however, DFR performance (which was captured during the fMRI scan) may also have been influenced by idiosyncratic variation related to the task and session (Stevens et al., 2012). Our Visual WMC measure was derived from an exploratory factor analysis, which seeks to identify the latent shared variation across multiple tasks and as such, will be less influenced by idiosyncratic sources of variance. It is possible that DFR performance is better predicted by the pattern stability across trials because this measure is derived at the trial level, so it may better reflect these idiosyncratic sources of variance that are particularly relevant for predicting performance. In contrast, Visual WMC was more strongly predicted by face-specific representational content. This measure (which is not derived at the individual trial level), may reflect more efficient representation of task-relevant information and/or better suppression of distractors.

Our final set of analyses included all potential modalities in a single model to determine whether they each reflected unique variance, attempt to maximize predictive ability, and investigate the importance of individual features. Intriguingly, the only analysis where the full model performed better than any of its individual parts was the model predicting WHODAS. Even though modalities beyond task fMRI were able to significantly predict DFR performance and Visual WMC, the full model did not predict significantly more variance than the task fMRI models

for these two outcomes. Previous literature has suggested that combining modalities improves prediction performance (Dhamala et al., 2021; Jiang et al., 2020; Ooi et al., 2022; Rasero et al., 2021; Tetereva et al., 2022). However, most of these studies only contain measures derived from resting state fMRI and structural measures, such as structural connectivity or cortical surface area and thickness. To date, the only study that combined task-based univariate activity with other measures found that combining univariate task fMRI, resting state fMRI and structural measures including cortical thickness, cortical surface area and subcortical volume performs better than a model only including task fMRI measures. One potential explanation for the apparent discrepancy between this result and our findings may stem from the fact that Tetereva and colleagues (2022) only included univariate task fMRI activations (albeit from multiple tasks). It is possible that the addition of multivariate pattern features captured variance that is also reflected in other measures but not present in univariate fMRI and thus increased our task fMRI model's predictive performance to be comparable with that of the full models. A second key distinction between the existing literature and our study is that we utilized within- and across-network average resting state functional connectivity to reduce computational complexity, rather than including the entire 400x400 connectivity matrix as features. It is possible that in doing so, we diminished our ability to explain variance in our outcome measures from resting state fMRI.

Our full models also allow us to investigate the importance of specific features in the models. The mixture parameter in our ElasticNet models allows for robust clusters of correlated variables to be maintained, while still providing some degree of feature selection through the L1 regularization. Considering the features that remain in our models can provide insight into the processes that underlie working memory performance, working memory capacity, and psychiatric function.

Our model predicting DFR performance showed a sizable cluster of task fMRI pattern stability features, across the ventral visual stream (extrastriate and parahippocampal cortex), in addition to parietal and prefrontal cortex. There has been considerable debate in the working memory literature about where and how representations are maintained across a delay period (Riley & Constantinidis, 2016; Rose, 2020; Serences, 2016; Sreenivasan et al., 2014; Sreenivasan & D'Esposito, 2019; Xu, 2020), with some work suggesting that the short term maintenance of mnemonic content is subserved by the activation of the same neural regions that are implicated during perception of those stimuli (Courtney et al., 1997; Emrich et al., 2013; Harrison & Tong, 2009; O'Craven & Kanwisher, 2000; Ranganath et al., 2004; Ranganath & D'Esposito, 2005). Others have suggested that stimulus maintenance in the parietal cortex is more efficient, robust and important than sensory reactivation, as we are constantly bombarded with visual content that may bias representations held in the visual cortex (Bettencourt & Xu, 2016; Xu, 2020). There is mixed literature on whether stimulus representations are maintained in the PFC, with some work showing an inability to decode task-specific content from the PFC (Christophel et al., 2012; Emrich et al., 2013; Lee et al., 2013), while others showing that it is possible (Ester et al., 2015; Lewis-Peacock et al., 2012). Our results suggest that the stability of task representations of task content from the encoding period across the visual, parietal, and prefrontal cortex are all important for DFR performance.

Although we have shown that encoding-to-delay pattern stability is strongly predictive of performance on the DFR task, having heightened similarity across phases of a task does not necessarily suggest that the representations themselves contain more task-specific content. A more direct test of whether the content represented within a given region's delay period activity reflects the actual information being maintained in working memory comes from our face-specific

representational content features, which compared the multivariate pattern representations during the delay period to face and object "templates" from an independent localizer. Using this approach, we show that larger load effects in face-specific representational content in the visual, parietal, and prefrontal cortices are related almost exclusively to Visual WMC, rather than DFR performance. It could be that having similar pattern reinstatement across task phases may be sufficient for successful DFR performance, but that individuals with high Visual WMC are able to extract and maintain specific task-relevant features more efficiently. These results are consistent with work using electroencephalography (EEG) that shows that individuals with higher WMC are more efficient at maintaining only goal-relevant information, while lower WMC individuals tend to maintain more task-irrelevant information (McNab & Klingberg, 2008; Vogel et al., 2005).

A complementary aspect to successfully maintaining mnemonic content over a delay is protecting the maintained representations against interference from distractors. Distractors during the delay period can systematically bias maintained mnemonic content (Bettencourt & Xu, 2016; Huang, 2010; Magnussen, 2000; Magnussen & Greenlee, 1992, 1999; Rademaker et al., 2015). It is possible that this process is mediated by lateral PFC (Postle, 2005) – Chao & Knight (1998) demonstrated that patients with lateral PFC lesions showed worse performance on a delayed match-to-sample task and showed more EEG markers of distraction from irrelevant sensory stimuli. Other work has suggested that using theta-burst transcranial magnetic stimulation to disrupt lateral PFC functioning leads to impaired performance on delayed match-to-sample tasks and less precision in content-specific representations in visual cortex (Feredoes et al., 2011; Higo et al., 2011; Miller et al., 2011; Zanto et al., 2011).

Although our DFR task did not include the presentation of visual distractors, it is likely that distractor suppression mechanisms may still play a role in preserving the fragile contents of

visual working memory from distraction by task-irrelevant thoughts or other sensory experiences in the scanner. We show that increased functional connectivity between the Control B and Salience/Ventral Attention subnetworks at rest, larger load effects in the prefrontal region Pars Opercularis (part of the Salience/Ventral Attention A subnetwork), and increased univariate activation at high load in the superior extrastriate visual cortex are all negatively associated with DFR performance. It is possible that the increased activation in the Salience/Ventral Attention subnetwork and visual cortex may reflect processing of salient distractors (including from internally-generated thoughts) during the delay period. The Control B network has been associated with the regulation of perceptual attention (Dixon et al., 2018); stronger connectivity between this subnetwork and Salience/Ventral Attention subnetworks may result in more influence of the distractors on the PFC representations or control processes.

We also show that betweenness centrality, which measures how many of the shortest paths in a network must pass through a given node and thus indexes integration of modular networks (Rubinov & Sporns, 2010), of the left intraparietal sulcus (IPS) is negatively associated with DFR performance. We also found that negative global brain connectivity of the right IPS is positively associated with Visual WMC. Consistent with these results, one recent study showed that lower degree centrality (a summary measure of the connectivity strength of a given region to all other regions), and stronger negative functional connections in the parietal cortex are associated with higher working memory capacity (Markett et al., 2018).

We also show that the participation coefficient of the left IPS is positively associated with DFR performance, and the node strength of the right inferior parietal lobule (IPL) is positively associated with Visual WMC. A larger participation coefficient suggests a higher ratio of intermodular hubs to intramodular hubs, indicating that the node may be a connector hub between

networks (Rubinov & Sporns, 2010). Node strength reflects the weighted sum of the neighboring connections; a higher strength value reflects a more connected region (Rubinov & Sporns, 2010). Previous work has shown that stronger indices of local processing in the bilateral parietal cortex at rest is associated with better visual working memory (Alavash et al., 2015). Other work has shown that stronger centrality in the parietal cortex at rest is associated with better working memory performance and that increases in centrality in the parietal cortex are associated with greater gains following working memory training (Langer et al., 2013).

Together, these results suggest two possibilities. On one hand, less functional integration of the parietal cortex at rest may help protect maintained representations from distractors. On the other, having more connections of the parietal cortex to other modules in the brain may allow for more efficient transfer of information across the working memory system. Future work may seek to investigate the connectivity of the parietal cortex with individual networks, as opposed to with all other regions, to help understand its role in the transfer of information across networks.

Our results also highlight the role of the lateral prefrontal cortex in both performance and Visual WMC. The lateral PFC has been implicated as a flexible hub that controls a wide range of processing (Cole et al., 2012; Menon & D'Esposito, 2022; Serences, 2016). Other work has suggested that the lateral PFC is particularly important for explaining individual differences in working memory capacity (Minamoto et al., 2017). Our results highlight how the univariate activation at high load and degree of connectedness of the lateral PFC, as measured by the global connectivity and strength of the lateral PFC, are predictive of Visual WMC. Prior work has shown similar effects, where the global connectivity of the lateral PFC is correlated with fluid intelligence (Cole et al., 2010, 2012). The authors interpret these results by suggesting that the lateral PFC is a flexible functional hub and that higher global connectivity at rest could reflect multiple possible

routes that the lateral PFC can dynamically reconfigure to perform different tasks (Cole et al., 2012).

Intriguingly, we also show that the *negative* global correlation is positively associated with DFR performance – that is, stronger functional anticorrelations of the lateral PFC are associated with better performance on the DFR task. Previous work has shown similar findings, where the anticorrelation of the lateral PFC and DMN is associated with fluid intelligence (Cole et al., 2012; Hampson et al., 2010; Keller et al., 2015). It is possible that this anticorrelation may reflect the disengagement of the internally-directed functioning of the DMN in favor of the task-related processing from the PFC .

One additional important implication of our work lies in our ability to directly predict psychiatric functioning. There has been a recent explosion of work seeking to predict behavioral outcomes (including cognitive performance and psychiatric outcomes) using brain-based measures (Sui et al., 2020). Given that cognition and behavior comprise one of the transdiagnostic pillars identified by the National Institute of Health's Research Domain Criteria (RDoC) framework (Cuthbert & Insel, 2013), and the existing literature demonstrating deficits in working memory across a spectrum of psychiatric disorders (Berman et al., 2011; Gathercole & Alloway, 2006; Grenard et al., 2008; Lapointe et al., 2013; Moriya & Sugiura, 2012; Perlstein et al., 2001; Thompson et al., 2006), identifying the specific aspects of neural processing that are uniquely related to psychiatric outcomes may provide more direct path for identifying effective targeted interventions. Accurate predictive modeling of cognition and behavior may provide insight into neuropsychiatric underpinnings of mental health and illness and provide benchmarks for success of new interventions.

Although self-report measures are typically more difficult to predict than objective measures of behavioral performance (Kong et al., 2019, 2021; Li et al., 2019; Sui et al., 2020), we were able to predict both WHODAS (Spearman's rho = 0.403) and BPRS (Spearman's rho = 0.255) relative well. In comparison, existing work has shown relatively low predictive ability of mental health outcomes, around a Pearson's correlation of 0.1 between predicted and actual mental health scores (Chen et al., 2022; Ooi et al., 2022). Our more successful predictions of psychiatric function may stem from our unique sample, which includes participants with a wide range of psychiatric functioning. Most of the work predicting cognitive performance utilizes large, open-source neuroimaging datasets such as the Human Connectome Project (Van Essen et al., 2013) or the Adolescent Brain and Cognitive Development Study (Jernigan et al., 2018), which have the benefit of large sample sizes necessary for machine learning prediction algorithms. Despite the benefit of their size, however, these datasets intentionally threshold their participants with respect to psychiatric risk, thus potentially hurting their predictive ability for mental health outcomes (Ooi et al., 2022).

We additionally show that similar patterns of neuroimaging-derived features predict psychiatric outcomes and working memory performance and capacity. Specifically, we show that stronger functional integration of the parietal cortex and lateral PFC, in addition to face-specific representational content in the prefrontal cortex are related to better psychiatric outcomes. Intriguingly, we also show that stronger face-specific representational content in the frontal eye fields, right lateral PFC and visual cortex, in addition to larger load effects in the lateral PFC are associated with poorer psychiatric outcomes. It is possible that these findings could reflect less efficient processing or sub-optimal maintenance of task-relevant stimuli. We also show that stronger pattern stability in the Salience/Ventral Attention networks is associated with better

psychiatric outcomes, potentially reflecting better suppression of salient distractors, or other domain-general attention related processes.

Although the present work provides novel insight into the utility of different aspects of task fMRI in predicting working memory performance, working memory capacity, and psychiatric outcomes, there are limitations to this study. Previous work has suggested that other measures of brain structure and anatomy, such as cortical surface area and volume, and structural connectivity (e.g., diffusion tensor imaging tractography) may also be important predictors of cognitive ability and psychiatric functioning (Choi et al., 2008; Dhamala et al., 2021; Ekman et al., 2016; Mansour et al., 2021; Ooi et al., 2022). Similarly, task-related connectivity measures have also been shown to be related to cognitive performance (Chen et al., 2022; Greene et al., 2018; Jiang et al., 2020). Inclusion of such additional measures could potentially improve our overall predictive performance and provide additional insight into the underpinnings of individual differences in working memory performance and capacity.

Similarly, although the ElasticNet algorithm we used for our predictive models has the benefit of being easily interpretable by virtue of its linearity, it is unable to capture non-linear relationships between features and outcomes or interactions between features. Other algorithms, such as support vector regression with a radial basis function or polynomial kernels, random forest or XGBoost algorithms, have recently been used to predict cognition (Pat et al., 2022; Tetereva et al., 2022) and may allow for potential non-linear and interactive effects.

It is also possible that our choice to restrict our regions of interest to those showing load effects during the delay period at the group level in our task fMRI measures may have potentially obscured interesting effects. Even though we retained 82 ROIs out of 400 potential regions in our parcellation, the 318 regions we excluded from our task fMRI analyses might have contained

signals that could have been diagnostic in predicting working memory ability. Likewise, our choice to use network-average resting state functional connectivity measures rather than the full connectivity matrix as features may have deprived the model of finer-grained connectivity signatures of cognitive ability and psychological well-being. We made these choices to reduce computational complexity as we combined modalities, as predictive models can be easily overwhelmed by an excessive features to examples ratio. Future work may wish to include all potential regions of interest in a parcellation, potentially with data-reduction techniques (Rasero et al., 2021; Sripada et al., 2020) to determine whether effects are indeed restricted to regions that are modulated by working memory load and whether there are additional processes that occur during encoding or retrieval that may improve predictive performance.

Finally, although our sample is large relative to many fMRI studies and fairly unique in its heterogeneity of psychiatric functioning, we may not have had a sufficient sample size to adequately power our analyses. Recent work has shown that thousands of participants may be necessary to reproducibly estimate brain-behavior relationships (although the use of multivariate fMRI measures may reduce the number of participants necessary to reliably estimate these relationships) (Marek et al., 2020). We also did not have a large enough sample size to use alternative machine learning approaches, such as stacking ensembles, which have been recently used with data from the Human Connectome Project (with individual studies using around 500-800 participants) to quantify the relative importance of different modalities (Ooi et al., 2022; Rasero et al., 2021; Tetereva et al., 2022). These models necessitate the training and cross-validation of two layers of algorithms; given our sample size, we would be unable to have large enough folds for both layers to produce models with acceptable amounts of error.

Taken together, our findings highlight the roles of distinct aspects of brain structure and function, including multivariate pattern information from task fMRI, in predicting performance on a delayed match-to-sample task involving faces, visual working memory capacity, and psychiatric outcomes. Task fMRI measures were most predictive of working memory ability/performance; specifically, encoding-to-delay pattern stability was most predictive of DFR performance, and face-specific representational content was most predictive of Visual WMC. Psychiatric outcomes were mostly predicted by resting state fMRI and structural MRI measures. Investigation of the significant features in models including all modalities highlights how better working memory performance is associated with the fidelity of representations across visual, parietal, and prefrontal cortex and the protection of these representations from distraction. Our results additionally suggest that Visual WMC may be supported by the flexible and efficient control of information within the working memory network, as well as the ability to selectively maintain task-relevant features. Future work can expand upon these results by including additional regions, task phases or data modalities to improve predictive ability and investigating the potential of different machine learning algorithms to identify and harness non-linear effects. Ultimately, these findings could serve as the basis for novel targeted interventions for boosting working memory capacity and improving psychiatric outcomes.

Chapter 5: General Discussion

The three studies presented so far in this dissertation have used behavioral and neuroimaging approaches to investigate how cognitive processes are integrated to support human memory. Together, the findings highlight the complexity of the mind and the brain – across three domains, behavior was best described by the interactions of processes, especially as behavior became more complex. In the following chapter, I will summarize and discuss the central experimental findings.

Chapter 2 utilized a paired associate learning task with semantically related and unrelated words to investigate the bidirectional interactions of episodic and semantic memory. The testing effect, or the relative benefit of actively retrieving to-be-remembered materials over passively restudying them, has been a robust and long-standing finding in the episodic memory literature (Delaney et al., 2010; Karpicke & Roediger, 2008; Kornell & Vaughn, 2016; Rowland, 2014). A separate line of research has shown that semantic information can either facilitate (Antony et al., 2022; Payne et al., 2012; van Kesteren et al., 2020; Wing et al., 2022) or impair (Antony & Bennion, 2022; Craig et al., 2013) episodic memory performance. We showed that semantic relatedness decreased the magnitude of the testing effect after a delay by improving recall of semantically related restudied pairs, meaning that the relative benefit for testing over restudying was decreased for semantically related pairs.

Accuracy alone, however, only provided evidence that there was an effect of semantic relatedness; it could not give insight into the mechanism underlying the effects. To that end, we developed a novel extension of a multi-arrangement paradigm (Kriegeskorte & Mur, 2012) to measure the semantic similarity of all to-be-learned words before and after learning. Using this paradigm allowed us to show that successfully recalled to-be-learned pairs of words were drawn

together closer in space. Importantly, this paradigm allowed us to index the full semantic space, from which we could make inferences about *how* the semantic representations of each word changed. Here, we showed that for related pairs, cue words changed more than target words, while cues and targets in unrelated pairs changed an equivalent amount. Moreover, we adapted an analysis from the neuroimaging literature (Bein et al., 2020) to show that learning drew cues towards the targets in representational space for related pairs, suggesting that prior knowledge about the relationship between the words sculpted the cue to become more predictive of the target word and improve later recall. We also used this paradigm to show that moderately related lure words (i.e., all other words in our target set that were not part of the to-be-learned pair) were pushed further away from cues in representational space while the representations of other strongly or not related lures were not systematically changed.

Taken together, these results are consistent with predictions from neurobiologically inspired computational modeling accounts of memory, such as the Non-Monotonic Plasticity Hypothesis (NMPH) (Ritvo et al., 2019). This hypothesis posits that the relative co-activation of two items changes their relationship in memory, with strong co-activation strengthening the relationship, moderate co-activation weakening it, and weak co-activation not having an effect. Shared spreading semantic information may cause more co-activation between semantically related pairs (Anderson, 1983; Collins & Loftus, 1975) and thus strengthen them more than semantically unrelated pairs. Additionally, while testing and restudy both strongly co-activate representations, testing activates representations of other moderately related concepts during the search process for the correct answer and this moderate co-activation weakens these relationships to prevent future interference.

Finally, we showed that recall of a to-be-learned pair was more likely when the cue word in the pair showed more representational change. Representational change in the target, however, was only related to recall of the pair when the pair was unrelated and tested. These findings suggest that how predictive the cue is of the target is always important for later recall, but it may only be necessary to sculpt the semantic representation of the target if elaborative links do not already exist. That is to say, the impact that episodic learning has on recall depends on the presence or absence of prior knowledge about the to-be-learned associations, highlighting the interactive nature of episodic and semantic memory systems.

While Chapter 2 showed intriguing interactions *within* declarative memory systems, it did not investigate any factors *outside* of memory. In Chapter 3, we turned to the investigation of face memory ability, which previous work has characterized as being supported by both general memory, face-specific processing, and other cognitive systems (Ramot et al., 2019). In this study, we collected an extensive battery of behavioral tasks that spanned multiple domains of cognition and applied exploratory factor analysis to identify four latent cognitive factors that reflected general fluid intelligence, episodic long-term memory, face perception, and working memory. These factors are consistent with prior work identifying latent cognitive processes (Beam et al., 2021; Schöttner et al., 2023), and highlights the separation between face perception and more domain-general object processing and memory abilities (Connolly et al., 2019; McGugin et al., 2012; Van Gulick et al., 2016; Verhallen et al., 2017).

Once we had identified latent cognitive factors, we used them as predictors in a regression to explain individual differences in performance on the Cambridge Face Memory Task (CFMT) (Duchaine & Nakayama, 2006a) and the Personal Identity Memory (PIM) task – a novel task that we developed in order to index face memory in a more ecologically valid manner. In the PIM task,

we had participants learn about identities using short, dynamic videos of the person where they interacted with the world across multiple contexts. Identities were also associated with personal semantic information – a name, a location, a hobby, and a vice. Hobby and vice were chosen to be socially relevant and congruent with the video content for the identity to maximize ecological validity of the task. The PIM task queried two aspects of personal identity memory – first, it investigated recognition memory using new photos of each identity in contexts that were not present during the learning video; this procedure necessitated the ability to generalize knowledge about the faces to new contexts, as might happen in real life. Second, we tested recall of the associated personal semantic information through both free recall and multiple-choice questions.

We were able to significantly predict performance on all three outcomes (CFMT score, PIM Recognition Score, PIM Overall Multiple Choice Score). All four cognitive factors contributed to the prediction of the CFMT score; in contrast, only the episodic long-term memory, face perception and working memory factors predicted the PIM Recognition Score. We also showed that the working memory factor more strongly predicted the PIM Recognition Score than the CFMT score. It is possible that the dissociation of significant predictors across outcome measures that putatively measure the same underlying ability could stem from the differences between how memory is tested – in the CFMT, the test stimuli are much more visually similar to each other and to the images used in learning, which could allow individuals to rely more on the detailed visual processing, rather than verbal rehearsal processes that may be supported by working memory.

In the previous linear regression analyses, we leveraged the across outcome measures to determine what features linearly combine to support face memory. We complemented this analysis with a data-driven k-means clustering analysis on a sample of our dataset that was constrained to

have similar performance on the outcome measures. This approach allowed us to investigate whether there were multiple potential strategies that supported similar performance on our tasks. To that end, we identified two cognitive profiles – one that was strong on our face perception and general episodic memory factors and weak on the working memory and general intelligence factors, and one that showed the opposite pattern of strengths and weaknesses.

Given the apparent dissociation between face perception and working memory, we turned back to the prediction of our outcome measures and included an interaction term between the face perception and working memory factors. This interaction significantly predicted the PIM Recognition Score, but not the CFMT score, and showed that the effect of face perception was stronger when working memory was weak and vice versa. These results suggest that there are multiple ways to succeed at an ecologically valid face memory task – one that may rely on the visual processing and memory of faces, and another that may rely more on working memory, including the use of strategies like elaborative encoding or better suppression of distractors and irrelevant information.

In Chapter 3, we showed that the factors that predict face memory performance may depend on working memory capacity. In Chapter 4, we explored the neural underpinnings of working memory capacity. To do so, we characterized working memory performance during a scanned delayed face recognition (DFR) match-to-sample task and visual and verbal working memory capacity (WMC) measured outside the scanner on a battery of independent tasks. We additionally captured three main modalities of brain features – structural MRI (consisting of cortical thickness and subcortical volume), resting state fMRI (consisting of within- and across-network resting state functional connectivity in addition to graph theoretical measures of local and global network structure), and task fMRI measures tracking activation levels and patterns during the maintenance

of to-be-learned faces in the DFR task. Specifically, we extracted traditional general linear model (GLM) univariate contrasts of high and low load (that is, regions that are sensitive to the difficulty of the task; this measure is commonly used in fMRI studies of working memory (Li et al., 2022) and may be an indicator of sustained attention or cognitive control; Minamoto et al., 2017), a measure of encoding-to-delay pattern stability (which may reflect the recruitment of the sensory processing during encoding; Serences, 2016), and an index of face-specific representational content (which may reflect better extraction of task-relevant features or protection of the maintained representations from distractors; Bettencourt & Xu, 2016; Vogel et al., 2005).

Next, we used these features (extracted from regions that were modulated by load in any of the three task fMRI measures) as predictors of the three working memory measures (DFR task performance, Visual WMC, and Verbal WMC), in addition to two measures of overall psychiatric disability across a series of ElasticNet regression models (Zou & Hastie, 2005). When each modality of features was used in separate models, we showed that working memory performance could be predicted by the task fMRI and resting state, Visual WMC could only be predicted by task fMRI measures, and none of our modalities could significantly predict Verbal WMC. Next, we decomposed the task fMRI measures to further investigate which task-related processes specifically contributed to each outcome measure. We found a striking dissociation – while encoding-to-delay pattern stability was most predictive of DFR task performance measured during the scanner, the measure of face-specific representational content predicted Visual WMC. Together, these results suggest that while visual working memory task performance is characterized by the reactivation of the same cortical regions that were involved during encoding, higher Visual WMC may be characterized by more efficient representation of task-relevant information.

Our final analyses combined all possible features into a single model to determine whether each of the modalities utilized unique variance to predict our outcome measures and to investigate the regional contributions of each kind of features. In this analysis, we found similar results to our previous analysis, where mostly encoding-to-delay pattern stability features were predictive of DFR task performance while mostly face-specific representational content features were predictive of Visual WMC. Our models showed these effects across the visual, parietal and prefrontal cortices, which is consistent with previous work decoding task-relevant content in these regions (Bettencourt & Xu, 2016; Harrison & Tong, 2009; Lewis-Peacock et al., 2012). Some have suggested that the parietal cortex is more efficient at maintaining visual representations as the visual cortex is constantly bombarded with distracting content (Bettencourt & Xu, 2016; Y. Xu, 2020); others have debated whether task-specific content can be decoded from prefrontal cortex (Christophel et al., 2012; Emrich et al., 2013; Lee et al., 2013). Our results suggest that task-specific content is present across all three cortical regions and that the strength of these representations is related to both DFR performance and Visual WMC.

The full models also revealed a number of significant predictors from resting state fMRI. Specifically, the models included predictors that suggested the overall connectivity of the lateral prefrontal cortex and the integration of the parietal cortex at rest were predictive of both performance and capacity. These findings are consistent with accounts that posit that the lateral PFC exerts flexible top-down control over to-be-maintained content-specific representations (Chao & Knight, 1998; Cole et al., 2012; Feredoes et al., 2011; Higo et al., 2011; Menon & D'Esposito, 2022; Serences, 2016; Zanto et al., 2011), and ones that suggest the functional integration of the parietal cortex may allow for efficient transfer of information across the working

memory system (Alavash et al., 2015; Cohen & D'Esposito, 2016; Langer et al., 2013) and the protection of representations against distractors (Markett et al., 2018).

Finally, we show that we could predict a striking amount of variance in our outcome measures indexing psychiatric dysfunction, and that the most important features for predicting these measures were largely overlapping with those that predict working memory performance and capacity. Specifically, we showed that stronger functional integration of the parietal cortex and lateral PFC, in addition to face specific representational content in the PFC, are related to better psychiatric outcomes. It is possible that these findings reflect an inefficient maintenance of task-relevant stimuli or less effective use of domain-general attention-related processes.

Future Directions

Although these studies have laid important groundwork establishing evidence for the integration of cognitive processes across domains of human memory, future work could extend these findings by considering other potential sources of variance, either by expanding feature sets or employing alternate analytic frameworks. For example, Chapter 2 might be extended by utilizing to-be-learned pairs of words that span a wider range of semantic relatedness, as the effect of semantic relatedness and prior knowledge may depend on the range of strength of associations across the entire stimulus set (Antony et al., 2022). Creating more variance in the range of semantic relatedness within sets may impact the role to which semantic knowledge impacts episodic learning. The work in Chapter 3 could be extended by measuring social cognition, which has been shown to be implicated in face memory (Ramot et al., 2019), and the work in Chapter 4 could be extended by including other brain measures such as task-related connectivity or structural connectivity (Ekman et al., 2016; Jiang et al., 2020), features from the encoding or probe phases

(Bocincova & Johnson, 2019), or using algorithms that allow for more interactive or non-linear effects such as random forests, support vector regressions with polynomial or RBF kernels, or XGBoost (Pat et al., 2022; Tetereva et al., 2022).

Additionally, future studies could systematically manipulate the interactions of the processes identified in this dissertation to further characterize their influence and their boundary conditions. We could extend the work presented in Chapter 3 by training participants to improve performance using strategies that focus on enriching face perception or working memory abilities, depending on their particular cognitive profile. This could occur through either behavioral interventions or using non-invasive covert real-time neurofeedback (Ramot et al., 2016). The results in Chapter 4 are particularly relevant to the treatment of mental illness, given the recent push from the National Institute of Health's Research Domain Criteria (RDoC) to include cognition as a key pillar of characterizing mental illness in a more transdiagnostic fashion (Bilder et al., 2013; Cuthbert & Insel, 2013). Utilizing the specific predictors of working memory capacity here could prove fruitful for the development of targeted interventions to improve psychiatric outcomes.

Conclusion

Taken together, the studies in this dissertation use three domains of memory to demonstrate how cognitive processes integrate to support complex behavior. They have shown that complex behavior is best described by the interactive of cognitive processes. As large-scale behavioral and neuroimaging datasets become available and more sophisticated analytical frameworks are developed, it will be critical to refine how we think about cognition and to better parse the underlying cognitive subprocesses that together contribute to any given task. Tightly controlled

lab-based experiments will always be important for advancing our knowledge, but expanding our understanding of real-world, complex behavior will necessitate more complex, integrative hypotheses as well.

References

Alavash, M., Doebler, P., Holling, H., Thiel, C. M., & Gießing, C. (2015). Is functional integration of resting state brain networks an unspecific biomarker for working memory performance? *NeuroImage*, *108*, 182–193. https://doi.org/10.1016/j.neuroimage.2014.12.046

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Current Biology*, *23*(15), 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065

Anderson, J. R. (1983). A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior*, *22*. https://doi.org/10.1016/B978-1-4832-1446-7.50016-9

Anderson, J. R., Pyke, A. A., & Fincham, J. M. (2016). Hidden Stages of Cognition Revealed in Patterns of Brain Activation. *Psychological Science*, *27*(9), 1215–1226. https://doi.org/10.1177/0956797616654912

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The Discovery of Processing Stages: Extension of Sternberg's Method. *Psychological Review*, *123*(5), 481–509. https://doi.org/10.1037/rev0000030

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*, *23*(3), 905–913. https://doi.org/10.1016/j.neuroimage.2004.07.060

Antony, J. W., & Bennion, K. A. (2022). Semantic associates create retroactive interference on an independent spatial memory task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001216

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route for consolidation. *Trends in Cognitive Sciences, 21*(8), 573–576. https://doi.org/10.1016/j.tics.2017.05.001.Retrieval

Antony, J. W., Romero, A., Vierra, A. H., Luenser, R. S., Hawkins, R. D., & Bennion, K. A. (2022). Semantic relatedness retroactively boosts memory and promotes memory interdependence across episodes. *eLife*, *11*, e72519. https://doi.org/10.7554/eLife.72519

Anzellotti, S., & Young, L. L. (2020). The acquisition of person knowledge. *Annual Review of Psychology*, *71*, 613–634. https://doi.org/10.1146/annurev-psych-010419-050844

Arrington, M., Elbich, D., Dai, J., Duchaine, B., & Scherf, K. S. (2022). Introducing the female Cambridge face memory test – long form (F-CFMT+). *Behavior Research Methods*, *54*(6), 3071–3084. https://doi.org/10.3758/s13428-022-01805-8

Assem, M., Blank, I. A., Mineroff, Z., Ademoğlu, A., & Fedorenko, E. (2020). Activity in the fronto-parietal multiple-demand network is robustly associated with individual differences in working memory and fluid intelligence. *Cortex*, *131*, 1–16. https://doi.org/10.1016/j.cortex.2020.06.013

Audrain, S., & McAndrews, M. P. (2022). Schemas provide a scaffold for neocortical integration of new memories over time. *Nature Communications*, *13*(1), 5795. https://doi.org/10.1038/s41467-022-33517-0

Avery, E. W., Yoo, K., Rosenberg, M. D., Greene, A. S., Gao, S., Na, D. L., Scheinost, D., Constable, T. R., & Chun, M. M. (2019). Distributed Patterns of Functional Connectivity Predict Working Memory Performance in Novel Healthy and Memory-impaired Individuals. *Journal of Cognitive Neuroscience*, 1–15.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of Learning and Motivation—Advances in Research and Theory* (Vol. 8, pp. 47–89). https://doi.org/10.1016/S0079-7421(08)60452-1

Badham, S. P., Estes, Z., & Maylor, E. A. (2012). Integrative and semantic relations equally alleviate age-related associative memory deficits. *Psychology and Aging*, *27*(1), 141–152. https://doi.org/10.1037/a0023924

Baey, C, & Kuhn, E. (2019). *varTestnlme: Variance components testing in mixed-effect models.* [Computer software]. https://github.com/baeyc/varTestnlme

Baker, F. B., & Hubert, L. J. (1975). Measuring the Power of Hierarchical Cluster Analysis. *Journal of the American Statistical Association*, *70*(349), 31–38. https://doi.org/10.1080/01621459.1975.10480256

Baker, K. A., & Mondloch, C. J. (2023). Unfamiliar face matching ability predicts the slope of face learning. *Scientific Reports*, *13*(1), 5248. https://doi.org/10.1038/s41598-023-32244-w

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*. https://doi.org/10.1016/j.neuron.2017.06.041

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience*, *38*(45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Ballard, I. C., Wagner, A. D., & McClure, S. M. (2019). Hippocampal pattern separation supports reinforcement learning. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-08998-1

Bartlett, M. S. (1950). TESTS OF SIGNIFICANCE IN FACTOR ANALYSIS. *British Journal of Statistical Psychology*, *3*(2), 77–85. https://doi.org/10.1111/j.2044-8317.1950.tb00285.x

Barton, J. J. S., Albonico, A., Susilo, T., Duchaine, B., & Corrow, S. L. (2019). Object recognition in acquired and developmental prosopagnosia. *Cognitive Neuropsychology*, *36*(1–2), 54–84. https://doi.org/10.1080/02643294.2019.1593821

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bays, P. M., Catalao, R. F. G., & Husain, M. (2011). *The precision of visual working memory is set by allocation of a shared resource*. *9*(10), 1–14. https://doi.org/10.1167/9.10.7.The

Beale, E. M. L. (1969). *Euclidean Cluster Analysis*. Scientific Control Systems Limited.

Beam, E., Potts, C., Poldrack, R. A., & Etkin, A. (2021). A data-driven framework for mapping domains of human neurobiology. *Nature Neuroscience*, *24*(12), 1733–1744. https://doi.org/10.1038/s41593-021-00948-9

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the effect of semantic congruency on episodic memory: The role of integration and relatedness. *PLoS ONE*, *10*(2), 1–24. https://doi.org/10.1371/journal.pone.0115624

Bein, O., Reggev, N., & Maril, A. (2020). Prior knowledge promotes hippocampal separation but cortical assimilation in the left inferior frontal gyrus. *Nature Communications*, *11*(4590), 1–13. https://doi.org/10.1038/s41467-020-18364-1

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Berman, M. G., Nee, D. E., Casement, M., Kim, H. S., Deldin, P., Kross, E., Gonzalez, R., Demiralp, E., Gotlib, I. H., Hamilton, P., Joormann, J., Waugh, C., & Jonides, J. (2011). Neural and behavioral effects of interference resolution in depression and rumination. *Cognitive, Affective and Behavioral Neuroscience*, *11*(1), 85–96. https://doi.org/10.3758/s13415-010-0014-x

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150–157. https://doi.org/10.1038/nn.4174

Bilder, R. M., Howe, A. G., & Sabb, F. W. (2013). Multilevel models from biology to psychology: Mission impossible? *Journal of Abnormal Psychology*, *122*(3), 917–927. https://doi.org/10.1037/a0032263

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation.

Cognitive Neuropsychology, *33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, *56*.

Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Boland, H., Campbell-Meiklejohn, D. K., Emsley, R. A., Munafo, M. R., Penton-Voak, I. S., Seara-Cardoso, A., Viding, E., Voon, V., Sahakian, B. J., Robbins, T. W., & Elliott, R. (2016). EMOTICOM: A Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition. *Frontiers in Behavioral Neuroscience*, *10*. https://doi.org/10.3389/fnbeh.2016.00025

Blazhenkova, O., & Kozhevnikov, M. (2008). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, *23*(5), 638–663. https://doi.org/10.1002/acp.1473

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

Bocincova, A., & Johnson, J. S. (2019). The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory. *Cortex*, *111*, 196–209. https://doi.org/10.1016/j.cortex.2018.10.013

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Bulevich, J. B., Thomas, A. K., & Parsow, C. (2016). Filling in the gaps: Using testing and restudy to promote associative learning. *Memory*, *24*(9), 1267–1277. https://doi.org/10.1080/09658211.2015.1098706

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*(4), 674–692. https://doi.org/10.1037/a0024695

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

Burianova, H., McIntosh, A. R., & Grady, C. L. (2010). A common functional brain network for autobiographical, episodic, and semantic memory retrieval. *NeuroImage*, *49*(1), 865–874. https://doi.org/10.1016/j.neuroimage.2009.08.066

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, *23*(1), 1–31. https://doi.org/10.1016/s0364-0213(99)80050-0

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*(3), 256–284. https://doi.org/10.1016/j.cogpsych.2005.06.003

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

Buuren, S. V., & Groothuis-Oudshoorn, K. (2011). **mice**: Multivariate Imputation by Chained Equations in *R. Journal of Statistical Software*, *45*(3). https://doi.org/10.18637/jss.v045.i03

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Candia, J., & Tsang, J. S. (2019). eNetXplorer: An R package for the quantitative exploration of elastic net families for generalized linear models. *BMC Bioinformatics*, *20*(1), 189. https://doi.org/10.1186/s12859-019-2778-5

Cao, H., McEwen, S. C., Forsyth, J. K., Gee, D. G., Bearden, C. E., Addington, J., Goodyear, B., Cadenhead, K. S., Mirzakhanian, H., Cornblatt, B. A., Carrión, R. E., Mathalon, D. H., McGlashan, T. H., Perkins, D. O., Belger, A., Seidman, L. J., Thermenos, H., Tsuang, M. T., Van Erp, T. G. M., … Cannon, T. D. (2019). Toward Leveraging Human Connectomic Data in Large Consortia: Generalizability of fMRI-Based Brain Graphs Across Sites, Sessions, and Paradigms. *Cerebral Cortex*, *29*(3), 1263–1279. https://doi.org/10.1093/cercor/bhy032

Caplan, J. B., Boulton, K. L., & Gagné, C. L. (2014). Associative asymmetry of compound words. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(4), 1163–1171. https://doi.org/10.1037/a0036588

Carpenter, S. K. (2009). Cue Strength as a Moderator of the Testing Effect: The Benefits of Elaborative Retrieval. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(6), 1563–1569. https://doi.org/10.1037/a0017021

Carpenter, S. K. (2011). Semantic Information Activated During Retrieval Contributes to Later Retention: Support for the Mediator Effectiveness Hypothesis of the Testing Effect.

Journal of Experimental Psychology: Learning Memory and Cognition, *37*(6), 1547–1552. https://doi.org/10.1037/a0024140

Carpenter, S. K., & Delosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testng effect. *Memory*, *34*(2), 268–276.

Carpenter, S. K., & Kelly, J. W. (2012). Tests enhance retention and transfer of spatial learning. *Psychonomic Bulletin and Review*, *19*, 443–448. https://doi.org/10.3758/s13423-012-0221-2

Carpenter, S. K., Pashler, H., & Cepeda, N. J. . (2009). Using Tests to Enhance 8th Grade Students' Retention of U.S. History Facts. *Applied Cognitive Psychology*, *23*, 760–771. https://doi.org/10.1002/acp

Chadwick, M. J., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2010). Decoding Individual Episodic Memory Traces in the Human Hippocampus. *Current Biology*, *20*(6), 544–547. https://doi.org/10.1016/j.cub.2010.01.053

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology*. https://doi.org/10.1016/j.cub.2017.06.057

Chanales, A. J. H., Tremblay-McGaw, A. G., Drascher, M. L., & Kuhl, B. A. (2021). Adaptive Repulsion of Long-Term Memory Representations Is Triggered by Event Similarity. *Psychological Science*, *32*(5), 705–720. https://doi.org/10.1177/0956797620972490

Chao, L. L., & Knight, R. T. (1998). Contribution of Human Prefrontal Cortex to Delay Performance. *Journal of Cognitive Neuroscience*, *10*(2), 167–177. https://doi.org/10.1162/089892998562636

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). **NbClust**: An *R* Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*(6). https://doi.org/10.18637/jss.v061.i06

Chen, J., Tam, A., Kebets, V., Orban, C., Ooi, L. Q. R., Asplund, C. L., Marek, S., Dosenbach, N. U. F., Eickhoff, S. B., Bzdok, D., Holmes, A. J., & Yeo, B. T. T. (2022). Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. *Nature Communications*, *13*(1), 2217. https://doi.org/10.1038/s41467-022-29766-8

Choi, Y. Y., Shamosh, N. A., Cho, S. H., DeYoung, C. G., Lee, M. J., Lee, J.-M., Kim, S. I., Cho, Z.-H., Kim, K., Gray, J. R., & Lee, K. H. (2008). Multiple Bases of Human Intelligence Revealed by Cortical Thickness and Neural Activation. *The Journal of Neuroscience*, *28*(41), 10323–10329. https://doi.org/10.1523/JNEUROSCI.3259-08.2008

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. *The Journal of Neuroscience*, *32*(38), 12983–12989. https://doi.org/10.1523/JNEUROSCI.0184-12.2012

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*(3), 278–295. https://doi.org/10.1016/j.intell.2012.02.010

Clark, V. P., Maisog, J. M., Haxby, J. V, Vincent, P., Maisog, J. M., & Haxby, J. V. (1998). fMRI Study of Face Perception and Memory Using Random Stimulus Sequences. *Journal of Neurophysiology*, *79*, 3257–3265.

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*(September), 4–6.

Cohen, J. R., & D'Esposito, M. (2016). The Segregation and Integration of Distinct Brain Networks and Their Relationship to Cognition. *The Journal of Neuroscience*, *36*(48), 12083–12094. https://doi.org/10.1523/JNEUROSCI.2965-15.2016

Cole, M. W., Pathak, S., & Schneider, W. (2010). Identifying the brain's most globally connected regions. *NeuroImage*, *49*(4), 3132–3148. https://doi.org/10.1016/j.neuroimage.2009.11.001

Cole, M. W., Yarkoni, T., Repovš, G., Anticevic, A., & Braver, T. S. (2012). Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *Journal of Neuroscience*, *32*(26), 8988–8999. https://doi.org/10.1523/JNEUROSCI.0536-12.2012

Collins, A. M., & Loftus, E. F. (1975). A Spreading Activation Theory of Semantic Processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1016/B978-1-4832-1446-7.50016-9

Collins, A. M., & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. *Journal Of Verbal Learning And Verbal Behavior*, *8*, 240–247.

Collins, J. A., & Olson, I. R. (2014). Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*, *61*(1), 65–79. https://doi.org/10.1016/j.neuropsychologia.2014.06.005

Connell, L., & Lynott, D. (2014). Principles of Representation: Why You Can't Represent the Same Concept Twice. *Topics in Cognitive Science*, *6*(3), 390–406. https://doi.org/10.1111/tops.12097

Connolly, H. L., Young, A. W., & Lewis, G. J. (2019). Recognition of facial expression and identity in part reflects a common ability, independent of general intelligence and visual

short-term memory. *Cognition and Emotion*, *33*(6), 1119–1128. https://doi.org/10.1080/02699931.2018.1535425

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, *386*(6625), 608–611. https://doi.org/10.1038/386608a0

Coutanche, M. N., & Thompson-Schill, S. L. (2015). Creating concepts from converging features in human cortex. *Cerebral Cortex*, *25*(9), 2584–2593. https://doi.org/10.1093/cercor/bhu057

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity Behavioral and Brain Sciences. *Behavioral and Brain Sciences*, *4*, 87–185.

Cowan, N. (2010). The magical mystery four: How is WMC limited, and why? *Memory*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Craig, K. S., Berman, M. G., Jonides, J., & Lustig, C. (2013). Escaping the recent past: Which stimulus dimensions influence proactive interference? *Memory & Cognition*, *41*(5), 650–670. https://doi.org/10.3758/s13421-012-0287-0

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*(1). https://doi.org/10.1186/1741-7015-11-126

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*, 179–194.

Dalrymple, K. A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience*, *10*, 10–20. https://doi.org/10.1016/j.dcn.2014.07.003

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology*, *32*(4), 331–331. https://doi.org/10.1212/WNL.32.4.331

Daneman, M., & Carpenter, P. A. (2004). Individual Differences in Working Memory and Reading. *Journal of Gastrointestinal Surgery*, *8*(1), 120–126. https://doi.org/10.1016/j.gassur.2003.10.009

Davelaar, E. J., Haarmann, H. J., Goshen-Gottstein, Y., & Usher, M. (2006). Semantic similarity dissociates short- from long-term recency effects: Testing a neurocomputational model of list memory. *Memory and Cognition*, *34*(2), 323–334. https://doi.org/10.3758/BF03193410

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722. https://doi.org/10.1080/17470210902811249

De Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, *39*(4), 695–710. https://doi.org/10.1080/02664763.2011.610445

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and Testing Effects: A Deeply Critical, Lengthy, and At Times Discursive Review of the Literature. In *Psychology of Learning and Motivation—Advances in Research and Theory* (Vol. 53, Issue C). https://doi.org/10.1016/S0079-7421(10)53003-2

Deng, Z., Chandrasekaran, B., Wang, S., & Wong, P. C. M. (2016). Resting-state low-frequency fluctuations reflect individual differences in spoken language learning. *Cortex*, *76*, 63–78. https://doi.org/10.1016/j.cortex.2015.11.020

Dennett, H. W., Mckone, E., Tavashmi, R., Hall, A., Pidcock, M., & Edwards, M. (2011). *The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test , with norms , reliability , sex differences , dissociations from face memory , and expertise effects*. https://doi.org/10.3758/s13428-011-0160-2

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: An fMRI study. *Neuropsychologia*, *45*(14), 3157–3166. https://doi.org/10.1016/j.neuropsychologia.2007.07.003

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

D'Esposito, M., & Postle, B. R. (2012). The Cognitive Neuroscience of Working Memory. *The Cognitive Neuroscience of Working Memory, November 2014*, 1–408. https://doi.org/10.1093/acprof:oso/9780198570394.001.0001

Detre, G. J., Natarajan, A., Gershman, S. J., & Norman, K. A. (2013). Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia*, *51*(12), 2371–2388. https://doi.org/10.1016/j.neuropsychologia.2013.02.017

Devue, C., Wride, A., & Grimshaw, G. M. (2019). New insights on real-world human face recognition. *Journal of Experimental Psychology: General*, *148*(6), 994–1007. https://doi.org/10.1037/xge0000493

Dhamala, E., Jamison, K. W., Jaywant, A., Dennis, S., & Kuceyeski, A. (2021). Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults. *Human Brain Mapping*, *42*(10), 3102–3118. https://doi.org/10.1002/hbm.25420

Díez, E., Gómez-Ariza, C. J., Díez-Álamo, A. M., Alonso, M. A., & Fernandez, A. (2017). The processing of semantic relatedness in the brain: Evidence from associative and categorical false recognition effects following transcranial direct current stimulation of the left anterior temporal lobe. *Cortex*, *93*, 133–145. https://doi.org/10.1016/j.cortex.2017.05.004

Dimsdale-Zucker, H. R., & Ranganath, C. (2019). Representational Similarity Analyses: A Practical Guide for Functional MRI Applications. *Handbook of Behavioral Neuroscience*, *28*, 509–525. https://doi.org/10.1016/B978-0-12-812028-6.00027-6

Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within

human hippocampal subfields. *Nature Communications*. https://doi.org/10.1038/s41467-017-02752-1

DiStefano, C., Zhu, M., & Mîndrilã, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research, and Evaluation*, *14*(20). https://doi.org/10.7275/DA8T-4G52

Dixon, M. L., De La Vega, A., Mills, C., Andrews-Hanna, J., Spreng, R. N., Cole, M. W., & Christoff, K. (2018). Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. *Proceedings of the National Academy of Sciences*, *115*(7). https://doi.org/10.1073/pnas.1715766115

Donders, F. C. (1868). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

Drascher, M. L., & Kuhl, B. A. (2022). Long-term memory interference is resolved via repulsion and precision along diagnostic memory dimensions. *Psychonomic Bulletin & Review*, *29*(5), 1898–1912. https://doi.org/10.3758/s13423-022-02082-4

Druzgal, T. J., & D'Esposito, M. (2003). Dissecting Contributions of Prefrontal Cortex and Fusiform Face Area to Face Working Memory. *Journal of Cognitive Neuroscience*, *15*(6), 771–784. https://doi.org/10.1162/089892903322370708

Duchaine, B. C., & Nakayama, K. (2006). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, *16*(2), 166–173. https://doi.org/10.1016/j.conb.2006.03.003

Duchaine, B., & Nakayama, K. (2006a). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Duda, R. O., & Hart, P. E. (1974). *Pattern Classification and Scene Analysis* (Vol. 44). John Wiley& Sons. https://www.journals.uchicago.edu/doi/10.1086/620282

Duff, M. C., Covington, N. V., Hilverman, C., & Cohen, N. J. (2020). Semantic Memory and the Hippocampus: Revisiting, Reaffirming, and Extending the Reach of Their Critical Relationship. *Frontiers in Human Neuroscience*, *13*(January), 1–17. https://doi.org/10.3389/fnhum.2019.00471

Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), 5295–5299. https://doi.org/10.1073/pnas.1621147114

Eayrs, J. O., & Lavie, N. (2019). Individual differences in parietal and frontal cortex structure predict dissociable capacities for perception and cognitive control. *NeuroImage*, *202*, 116148. https://doi.org/10.1016/j.neuroimage.2019.116148

Ekman, M., Fiebach, C. J., Melzer, C., Tittgemeyer, M., & Derrfuss, J. (2016). Different Roles of Direct and Indirect Frontoparietal Pathways for Individual Working Memory Capacity. *Journal of Neuroscience*, *36*(10), 2894–2903. https://doi.org/10.1523/JNEUROSCI.1376-14.2016

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for kit of factorreferenced cognitive tests: 1976.* Educational Testing Service.

Elbich, D. B., & Scherf, K. S. (2017). Beyond the FFA: Brain-behavior correspondences in face recognition abilities. *NeuroImage*, *147*(June 2016), 409–422. https://doi.org/10.1016/j.neuroimage.2016.12.042

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, *3*(11), 1149–1152.

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. *The Journal of Neuroscience*, *33*(15), 6516–6523. https://doi.org/10.1523/JNEUROSCI.5732-12.2013

Engle, R. W. (2002). Working Memory Capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19-23.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual Differences in Working Memory Capacity and What They Tell Us About Controlled Attention, General Fluid Intelligence, and Functions of the Prefrontal Cortex. In *Models of working memory* (pp. 102–134). Cambridge University Press.

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Estes, Z., & Jones, L. L. (2009). Integrative priming occurs rapidly and uncontrollably during lexical processing. *Journal of Experimental Psychology: General*, *138*(1), 112–130. https://doi.org/10.1037/a0014677

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fang, X., Zhang, Y., Zhou, Y., Cheng, L., Li, J., Wang, Y., Friston, K. J., & Jiang, T. (2016). Resting-state coupling between core regions within the central-executive and salience networks contributes to working memory performance. *Frontiers in Behavioral Neuroscience*, *10*(FEB), 1–11. https://doi.org/10.3389/fnbeh.2016.00027

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*, 1–10. https://doi.org/10.1038/ncomms11066

Favila, S. E., Samide, R., Sweigart, S. C., & Kuhl, B. A. (2018). Parietal representations of stimulus features are amplified during memory retrieval and flexibly aligned with top-down goals. *Journal of Neuroscience*, *38*(36), 0564–18. https://doi.org/10.1523/JNEUROSCI.0564-18.2018

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences*, *108*(42), 17510–17515. https://doi.org/10.1073/pnas.1106439108

Ferreira, C. S., & Wimber, M. (2021). *The testing effect for visual materials depends on pre-existing knowledge*.

Finn, E. S. (2021). Is it time to put rest to rest? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2021.09.005

Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215*(April), 116828. https://doi.org/10.1016/j.neuroimage.2020.116828

Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*, *160*(March), 140–151. https://doi.org/10.1016/j.neuroimage.2017.03.064

Fischl, B. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, *14*(1), 11–22. https://doi.org/10.1093/cercor/bhg087

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, *97*(20), 11050–11055. https://doi.org/10.1073/pnas.200033797

Fisher, G. G., McArdle, J. J., McCammon, R. J., Sonnega, A., & Weir, D. R. (2014). *New Measures of Fluid Intelligence in the HRS*. University of Michigan. https://hrs.isr.umich.edu/sites/default/files/biblio/dr-027b.pdf

Fletcher, T. D. (2022). *QuantPsyc: Quantitative Psychology Tools* (1.6) [R]. https://CRAN.R-project.org/package=QuantPsyc

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., & Dolan, R. J. (1996). The Trouble with Cognitive Subtraction. *NeuroImage*, *4*(2), 97–104. https://doi.org/10.1006/nimg.1996.0033

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*(3), 346–355. https://doi.org/10.1002/mrm.1910350312

Furl, N., Garrido, L., Dolan, R., Driver, J., & Duchaine, B. (2011). *Fusiform gyrus face-selectivity reflects facial recognition ability*. *23*(7), 1723–1740. https://doi.org/10.1162/jocn.2010.21545.Fusiform

Gabrieli, J. D. E., Cohen, N. J., & Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and Cognition, 7*(2), 157–177. https://doi.org/10.1016/0278-2626(88)90027-9

Gallen, C. L., Hwang, K., Chen, A. J.-W., Jacobs, E. G., Lee, T. G., & D'Esposito, M. (2023). Influence of goals on modular brain network organization during working memory. *Frontiers in Behavioral Neuroscience*, *17*, 1128610. https://doi.org/10.3389/fnbeh.2023.1128610

Gao, S., Greene, A. S., Constable, R. T., & Scheinost, D. (2019). Combining multiple connectomes improves predictive modeling of phenotypic measures. *NeuroImage, 201*, 116038. https://doi.org/10.1016/j.neuroimage.2019.116038

Gathercole, S. E., & Alloway, T. P. (2006). Practitioner review: Short-term and working memory impairments in neurodevelopmental disorders: Diagnosis and remedial support. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 47*(1), 4–15. https://doi.org/10.1111/j.1469-7610.2005.01446.x

Gauthier, I. (2018). Domain-Specific and Domain-General Individual Differences in Visual Object Recognition. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721417737151

Geerligs, L., Rubinov, M., Tyler, L. K., Brayne, C., Bullmore, E. T., Calder, A. C., Cusack, R., Dalgleish, T., Duncan, J., Henson, R. N., Matthews, F. E., Marslen-Wilson, W. D., Rowe, J. B., Shafto, M. A., Campbell, K., Cheung, T., Davis, S., Geerligs, L., Kievit, R., … Henson, R. N. (2015). State and trait components of functional connectivity: Individual differences vary with mental state. *Journal of Neuroscience*, *35*(41), 13949–13961. https://doi.org/10.1523/JNEUROSCI.1324-15.2015

Gerlach, C., Klargaard, S. K., & Starrfelt, R. (2016). On the Relation between Face and Object Recognition in Developmental Prosopagnosia: No Dissociation but a Systematic Association. *PLOS ONE*, *11*(10), e0165561. https://doi.org/10.1371/journal.pone.0165561

Glahn, D. C., Kim, J., Cohen, M. S., Poutanen, V.-P., Therman, S., Bava, S., Van Erp, T. G. M., Manninen, M., Huttunen, M., & Lönnqvist, J. (2002). Maintenance and Manipulation in Spatial Working Memory: Dissociations in the Prefrontal Cortex. *NeuroImage*, *17*(1), 201–213. https://doi.org/10.1006/nimg.2002.1161

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*(1), 32–41. https://doi.org/10.1016/j.neuropsychologia.2006.04.015

Graham, K. S., Patterson, K., & Hodges, J. R. (1999). Episodic memory: New insights from the study of semantic dementia and Alzheimer's disease. *Current Opinion in Neurobiology*, *9*, 245–250.

Graham, K. S., Simons, J. S., Pratt, K. H., Patterson, K., & Hodges, J. R. (2000). Insights from semantic dementia on the relationship between episodic and semantic memorry. *Neuropsychologia*, *38*, 313–324. https://doi.org/10.1016/j.neuropsychologia.2007.06.021

Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, *9*(1), 2807. https://doi.org/10.1038/s41467-018-04920-3

Greicius, M. D., Krasnow, B., Boyett-Anderson, J. M., Eliez, S., Schatzberg, A. F., Reiss, A. L., & Menon, V. (2003). Regional analysis of hippocampal activation during memory encoding and retrieval: fMRI study. *Hippocampus*, *13*(1), 164–174. https://doi.org/10.1002/hipo.10064

Grenard, J. L., Ames, S. L., Wiers, Reinout, W., Thrush, C., Sussman, S., & Stacy, A. W. (2008). Working Memory Capacity Moderates the Predictive Effects of Drug-Related Associations on Substance Use. *Psychology of Addictive Behaviors*, *22*(3), 426–432. https://doi.org/10.1038/mp.2011.182.doi

Grill-Spector, K., & Malach, R. (2004). THE HUMAN VISUAL CORTEX. *Annual Review of Neuroscience*, *27*(1), 649–677. https://doi.org/10.1146/annurev.neuro.27.070203.144220

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun—An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*(4), 930–944. https://doi.org/10.3758/s13428-014-0529-0

Guo, J., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences*, *115*(28). https://doi.org/10.1073/pnas.1802246115

Hacker, C., & Biederman, I. (2020). The Proficiency for Distinguishing Faces is Independent of the Proficiency for Remembering Them. *PsyArXiv*, 1–20. https://doi.org/10.31234/osf.io/9bwct

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis* (7th ed.). Prentice Hall.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 801–812. https://doi.org/10.1037/a0023219

Hallenbeck, G. E., Sprague, T. C., Rahmati, M., Sreenivasan, K. K., & Curtis, C. E. (2021). Working memory representations in visual cortex mediate distraction effects. *Nature Communications*, *12*(1), 4714. https://doi.org/10.1038/s41467-021-24973-1

Hampson, M., Driesen, N., Roth, J. K., Gore, J. C., & Constable, R. T. (2010). Functional connectivity between task-positive and task-negative brain areas and its relation to working memory performance. *Magnetic Resonance Imaging*, *28*(8), 1051–1057. https://doi.org/10.1016/j.mri.2010.03.021

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832.Decoding

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/10.2307/2346830

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). *Human Neural Systems for Face Recognition and Social Communication*.

Haxby, J. V., Ungerleider, L. G., Horwitz, B., Maisog, J. M., Rapoport, S. I., & Grady, C. L. (1996). Face encoding and recognition in the human brain. *Proceedings of the National Academy of Sciences*, *93*(2), 922–927. https://doi.org/10.1073/pnas.93.2.922

Haxby, J. V, Hoffman, E. A., & Gobbini, M. I. (2002). *Human Neural Systems for Face Recognition and Social Communication*.

Hayes, S. M., Ryan, L., Schnyer, D. M., & Nadel, L. (2004). An fMRI study of episodic memory: Retrieval of object, spatial, and temporal information. *Behavioral Neuroscience*, *118*(5), 885–896. https://doi.org/10.1037/0735-7044.118.5.885

Henderson, D., Poppe, A. B., Barch, D. M., Carter, C. S., Gold, J. M., Ragland, J. D., Silverstein, S. M., Strauss, M. E., & MacDonald, A. W. (2012). Optimization of a Goal Maintenance

Task for Use in Clinical Applications. *Schizophrenia Bulletin*, *38*(1), 104–113. https://doi.org/10.1093/schbul/sbr172

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, *52*(1), 258–271. https://doi.org/10.1016/j.csda.2006.11.025

Hennig, C. (2023). *fpc: Flexible Procedures for Clustering* (2.2-10) [R]. https://CRAN.R-project.org/package=fpc

Higo, T., Mars, R. B., Boorman, E. D., Buch, E. R., & Rushworth, M. F. S. (2011). Distributed and causal influence of frontal operculum in task control. *Proceedings of the National Academy of Sciences*, *108*(10), 4230–4235. https://doi.org/10.1073/pnas.1013361108

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hornberger, M., & Piguet, O. (2012). Episodic memory in frontotemporal dementia: A critical review. *Brain*, *135*(3), 678–692. https://doi.org/10.1093/brain/aws011

Howard, M. C. (2016). A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? *International Journal of Human-Computer Interaction*, *32*(1), 51–62. https://doi.org/10.1080/10447318.2015.1087664

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, J. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision*, *10*(2), 1–27. https://doi.org/10.1167/10.2.24

Hubert, L. J., & Levin, J. R. (1976). A General Statistical Framework for Assessing Categorical Clustering in Free Recall. *Psychological Bulletin*, *83*(6), 1072–1080.

Hulbert, J. C., & Norman, K. A. (2015). Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cerebral Cortex*, *25*(10), 3994–4008. https://doi.org/10.1093/cercor/bhu284

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

Irish, M., Bunk, S., Tu, S., Kamminga, J., Hodges, J. R., Hornberger, M., & Piguet, O. (2016). Preservation of episodic memory in semantic dementia: The importance of regions beyond the medial temporal lobes. *Neuropsychologia*, *81*, 50–60. https://doi.org/10.1016/j.neuropsychologia.2015.12.005

Irish, M., & Vatansever, D. (2020). Rethinking the episodic-semantic distinction from a gradient perspective. *Current Opinion in Behavioral Sciences*, *32*, 43–49. https://doi.org/10.1016/j.cobeha.2020.01.016

Jackson, J. E. (2005). Oblimin Rotation. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470011815.b2a13060

Jafarpour, A., Buffalo, E. A., Knight, R. T., & Collins, A. G. E. (2022). Event segmentation reveals working memory forgetting rate. *iScience*, *25*(103902).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer US. https://doi.org/10.1007/978-1-0716-1418-1

Jarjat, G., Ward, G., Hot, P., Portrat, S., & Loaiza, V. M. (2020). Distinguishing the Impact of Age on Semantic and Nonsemantic Associations in Episodic Memory. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. https://doi.org/10.1093/geronb/gbaa010

Jernigan, T. L., Brown, S. A., & Dowling, G. J. (2018). The Adolescent Brain Cognitive Development Study. *Journal of Research on Adolescence : The Official Journal of the Society for Research on Adolescence*, *28*(1), 154–156. https://doi.org/10.1111/jora.12374

Jiang, R., Zuo, N., Ford, J. M., Qi, S., Zhi, D., Zhuo, C., Xu, Y., Fu, Z., Bustillo, J., Turner, J. A., Calhoun, V. D., & Sui, J. (2020). Task-induced brain connectivity promotes the detection of individual differences in brain-behavior relationships. *NeuroImage*, *207*, 116370. https://doi.org/10.1016/j.neuroimage.2019.116370

Johnson, M. K., McMahon, R. P., Robinson, B. M., Harvey, A. N., Hahn, B., Leonard, C. J., Luck, S. J., & Gold, J. M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology*, *27*(2), 220–229. https://doi.org/10.1037/a0032060

Johnson, P. O., & Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika*, *15*(4), 349–367. https://doi.org/10.1007/BF02288864

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, *1*, 57–93.

Jonas, J., Rossion, B., Brissart, H., Frismand, S., Jacques, C., Hossu, G., Colnat-Coulbois, S., Vespignani, H., Vignal, J. P., & Maillard, L. (2015). Beyond the core face-processing network: Intracerebral stimulation of a face-selective area in the right anterior fusiform

gyrus elicits transient prosopagnosia. *Cortex, 72*, 140–155. https://doi.org/10.1016/j.cortex.2015.05.026

Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proceedings of the National Academy of Sciences*, *115*(43), 11084–11089. https://doi.org/10.1073/pnas.1800006115

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE, 5*(1). https://doi.org/10.1371/journal.pone.0008622

Kadlec, J., Walsh, C., Sadé, U., Amir, A., Rissman, J., & Ramot, M. (2023). *Putting cognitive tasks on trial: A measure of reliability convergence* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.07.03.547563

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory and Cognition, 30*(6), 823–840. https://doi.org/10.3758/BF03195769

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200. https://doi.org/10.1007/BF02289233

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement, 20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415. https://doi.org/10.1007/BF02291817

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4–5), 528–558. https://doi.org/10.1080/09541440601056620

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B*, *361*, 2109–2128.

Kapur, N., & Brooks, D. J. (1999). Temporally-specific retrograde amnesia in two cases of discrete bilateral hippocampal pathology. *Hippocampus*, *9*(3), 247–254. https://doi.org/10.1002/(SICI)1098-1063(1999)9:3<247::AID-HIPO5>3.0.CO;2-W

Karlsson Wirebring, L., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *Journal of Neuroscience*, *35*(26), 9595–9602. https://doi.org/10.1523/JNEUROSCI.3550-14.2015

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kassambara, Alboukadel. (2021). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (0.7.0) [Computer software]. https://CRAN.R-project.org/package=rstatix

Keller, J. B., Hedden, T., Thompson, T. W., Anteraper, S. A., Gabrieli, J. D. E., & Whitfield-Gabrieli, S. (2015). Resting-state anticorrelations between medial and lateral prefrontal cortex: Association with working memory, aging, and individual differences. *Cortex*, *64*, 271–280. https://doi.org/10.1016/j.cortex.2014.12.001

Kim, G., Norman, K. A., & Turk-Browne, N. B. (2017). Neural differentiation of incorrectly predicted memories. *Journal of Neuroscience*, *37*(8), 2022–2031. https://doi.org/10.1523/JNEUROSCI.3272-16.2017

Kondo, H., Osaka, N., & Osaka, M. (2004). Cooperation of the anterior cingulate cortex and dorsolateral prefrontal cortex for attention shifting. *NeuroImage*, *23*(2), 670–679. https://doi.org/10.1016/j.neuroimage.2004.06.014

Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., Sun, N., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2019). Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cerebral Cortex*, *29*(6), 2533–2551. https://doi.org/10.1093/cercor/bhy123

Kong, R., Yang, Q., Gordon, E., Xue, A., Yan, X., Orban, C., Zuo, X.-N., Spreng, N., Ge, T., Holmes, A., Eickhoff, S., & Yeo, B. T. T. (2021). Individual-Specific Areal-Level Parcellations Improve Functional Connectivity Prediction of Behavior. *Cerebral Cortex*, *31*(10), 4477–4500. https://doi.org/10.1093/cercor/bhab101

Koolschijn, R. S., Emir, U. E., Pantelides, A. C., Nili, H., Behrens, T. E. J., & Barron, H. C. (2019). The hippocampus and neocortical inhibitory engrams protect against memory interference. *Neuron*, *101*, 528–541. https://doi.org/10.1101/366377

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Kornell, N., & Vaughn, K. E. (2016). *How Retrieval Attempts Affect Learning*. 183–215. https://doi.org/10.1016/bs.plm.2016.03.003

Kovács, G. (2020). Getting to know someone: Familiarity, person recognition, and identification in the human brain. *Journal of Cognitive Neuroscience*, *32*(12), 2205–2225. https://doi.org/10.1162/jocn_a_01627

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(51), 20600–20605. https://doi.org/10.1073/pnas.0705654104

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*(JUL), 1–13. https://doi.org/10.3389/fpsyg.2012.00245

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences*, *108*(14), 5903–5908. https://doi.org/10.1073/pnas.1016939108

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*(1–2), 83–97. https://doi.org/10.1002/nav.3800020109

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer. https://doi.org/10.1007/978-1-4614-6849-3

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Kyllonen, P., & Christal, R. (1990). Reasoning Ability Is (Little More Than) Working-Memory Capacity?! *Intelligence*, *433*, 389–433.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(NOV), 1–12. https://doi.org/10.3389/fpsyg.2013.00863

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240. https://doi.org/10.1037//0033-295x.104.2.211

Landi, S. M., Viswanathan, P., Serene, S., & Freiwald, W. A. (2021). A fast link between face perception and memory in the temporal pole. *Science, 373*(6554), 581–585. https://doi.org/10.1126/science.abi6671

Langenecker, S. A., Zubieta, J. K., Young, E. A., Akil, H., & Nielson, K. A. (2007). A task to manipulate attentional load, set-shifting, and inhibitory control: Convergent validity and test-retest reliability of the Parametric Go/No-Go Test. *Journal of Clinical and Experimental Neuropsychology, 29*(8), 842–853. https://doi.org/10.1080/13803390601147611

Langer, N., Von Bastian, C. C., Wirz, H., Oberauer, K., & Jäncke, L. (2013). The effects of working memory training on functional brain network efficiency. *Cortex, 49*(9), 2424–2438. https://doi.org/10.1016/j.cortex.2013.01.008

Lapointe, M. L. B., Blanchette, I., Duclos, M., Langlois, F., Provencher, M. D., & Tremblay, S. (2013). Attentional bias, distractibility and short-term memory in anxiety. *Anxiety, Stress and Coping, 26*(3), 293–313. https://doi.org/10.1080/10615806.2012.687722

Larocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *Journal of Neuroscience, 33*(13), 5466–5474. https://doi.org/10.1523/JNEUROSCI.4293-12.2013

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience, 16*(8), 997–999. https://doi.org/10.1038/nn.3452

Lenartowicz, A., Truong, H., Enriquez, K. D., Webster, J., Pochon, J.-B., Rissman, J., Bearden, C. E., Loo, S. K., & Bilder, R. M. (2021). Neurocognitive subprocesses of working memory performance. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(6), 1130–1152. https://doi.org/10.3758/s13415-021-00924-7

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.7.2) [Computer software]. https://CRAN.R-project.org/package=emmeans

Leonard, C. J., Kaiser, S. T., Robinson, B. M., Kappenman, E. S., Hahn, B., Gold, J. M., & Luck, S. J. (2013). Toward the Neural Mechanisms of Reduced Working Memory Capacity in Schizophrenia. *Cerebral Cortex*, *23*(7), 1582–1592. https://doi.org/10.1093/cercor/bhs148

Levakov, G., Sporns, O., & Avidan, G. (2022). Modular community structure of the face network supports face recognition. *Cerebral Cortex*, *32*(18), 3945–3958. https://doi.org/10.1093/cercor/bhab458

Levakov, G., Sporns, O., & Avidan, G. (2023). Fine-scale dynamics of functional connectivity in the face-processing network during movie watching. *Cell Reports*, *42*(6), 112585. https://doi.org/10.1016/j.celrep.2023.112585

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience, 24*(1), 61–79. https://doi.org/10.1162/jocn_a_00140

Li, D., Christ, S. E., & Cowan, N. (2014). Domain-general and domain-specific functional networks in working memory. *NeuroImage, 102*, 646–656. https://doi.org/10.1016/j.neuroimage.2014.08.028

Li, J., Kong, R., Liégeois, R., Orban, C., Tan, Y., Sun, N., Holmes, A. J., Sabuncu, M. R., Ge, T., & Yeo, B. T. T. (2019). Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage, 196*, 126–141. https://doi.org/10.1016/j.neuroimage.2019.04.016

Li, X., O'Sullivan, M. J., & Mattingley, J. B. (2022). Delay activity during visual working memory: A meta-analysis of 30 fMRI experiments. *NeuroImage, 255*, 119204. https://doi.org/10.1016/j.neuroimage.2022.119204

Li, Y. P., Cooper, S. R., & Braver, T. S. (2021). The Role of Neural Load Effects in Predicting Individual Differences in Working Memory Function. *NeuroImage*, *245*(October), 118656. https://doi.org/10.1016/j.neuroimage.2021.118656

Lifanov, J., Linde-Domingo, J., & Wimber, M. (2021). Feature-specific reaction times reveal a semanticisation of memories over time and with repeated remembering. *Nature Communications*, *12*(1), 1–10. https://doi.org/10.1038/s41467-021-23288-5

Linden, D. E. J., Bittner, R. A., Muckli, L., Waltz, J. A., Kriegeskorte, N., Goebel, R., Singer, W., & Munk, M. H. J. (2003). Cortical capacity constraints for visual working memory: Dissociation of fMRI load effects in a fronto-parietal network. *NeuroImage*, *20*(3), 1518–1530. https://doi.org/10.1016/j.neuroimage.2003.07.021

Liu, X. L., & Ranganath, C. (2021). Resurrected memories: Sleep-dependent memory consolidation saves memories from competition induced by retrieval practice. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-021-01953-6

Liu, X., Li, X., Song, Y., & Liu, J. (2021). Separate and Shared Neural Basis of Face Memory and Face Perception in Developmental Prosopagnosia. *Frontiers in Behavioral Neuroscience*, *15*, 668174. https://doi.org/10.3389/fnbeh.2021.668174

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. J. (2019). Human Replay Spontaneously Reorganizes Experience. *Cell*, 1–13. https://doi.org/10.1016/j.cell.2019.06.012

Loaiza, V. M., & Srokova, S. (2020). Semantic Relatedness Corrects the Age-Related Binding Deficit in Working Memory and Episodic Memory. *The Journals of Gerontology: Series B*, *75*(9), 1841–1849. https://doi.org/10.1093/geronb/gbz055

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Luppi, A. I., Gellersen, H. M., Liu, Z.-Q., Peattie, A. R. D., Manktelow, A. E., Adapa, R., Owen, A. M., Naci, L., Menon, D. K., Dimitriadis, S. I., & Stamatakis, E. A. (2023). *Converging on consistent functional connectomics* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.06.23.546329

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Machizawa, M. G., Driver, J., & Watanabe, T. (2020). Gray Matter Volume in Different Cortical Structures Dissociably Relates to Individual Differences in Capacity and Precision of Visual Working Memory. *Cerebral Cortex (New York, NY)*, *30*(9), 4759–4770. https://doi.org/10.1093/cercor/bhaa046

Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63*(1), 46–63. https://doi.org/10.1016/j.jml.2010.03.001

Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences*, *23*(6), 247–251. https://doi.org/10.1016/S0166-2236(00)01569-1

Magnussen, S., & Greenlee, M. W. (1992). Retention and Disruption of Motion Information in Visual Short-Term Memory. *Journal of Experimental Psychology : Learning , Memory , and Cognition*, *18*(1), 151–156.

Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, *62*(2–3), 81–92. https://doi.org/10.1007/s004260050043

Manns, J. R., Hopkins, R. O., & Squire, L. R. (2003). Semantic memory and the human hippocampus. *Neuron*, *38*(1), 127–133. https://doi.org/10.1016/S0896-6273(03)00146-6

Manoach, D. S., Schlaug, G., Siewert, B., Darby, D. G., Bly, B. M., Benfield, A., Edelman, R. R., & Warach, S. (1997). Prefrontal cortex fMRI signal changes are correlated with working memory load. *NeuroReport*, *8*(2), 545–549. https://doi.org/10.1097/00001756-199701200-00033

Mansour, S., Tian, Y., Yeo, B. T. T., Cropley, V., & Zalesky, A. (2021). High-resolution connectomic fingerprints: Mapping neural identity and behavior. *NeuroImage*, *229*, 117695. https://doi.org/10.1016/j.neuroimage.2020.117695

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Rose Donohue, M., Foran, W., Miller, R. L., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Chen, J., Newbold, D. J., Zheng, A., Seider, N. A., Van, A. N., Laumann, T. O.,

Thompson, W. K., … Dosenbach, N. (2020). Towards Reproducible Brain-Wide Association Studies Affiliations. *bioRxiv*, *11*, 15–18.

Markett, S., Reuter, M., Heeren, B., Lachmann, B., Weber, B., & Montag, C. (2018). Working memory capacity and the functional connectome—Insights from resting-state fMRI and voxelwise centrality mapping. *Brain Imaging and Behavior*, *12*(1), 238–246. https://doi.org/10.1007/s11682-017-9688-9

Martin, A. (2007). The Representation of Object Concepts in the Brain. *Annual Reviews in Psychology*, *58*, 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L. Y., & Barense, M. D. (2018). Integrative and distinctive coding of perceptual and conceptual object features in the ventral visual stream. *eLife*, *7*(e31873), 1–29. https://doi.org/10.7554/eLife.31873

Mayer, J. S., Bittner, R. A., Goebel, R., Nikolić, D., Linden, D. E. J., & Bledowski, C. (2007). Common neural substrates for visual working memory and attention. *NeuroImage*, *36*(2), 441–453. https://doi.org/10.1016/j.neuroimage.2007.03.007

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(1). https://doi.org/10.1186/s41235-018-0112-9

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*, *102*(3), 419–457.

McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt Expertise Test reveals domain-general and domain-specific sex effects in

object recognition. *Vision Research*, *69*, 10–22. https://doi.org/10.1016/j.visres.2012.07.014

McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R. B., Rivolta, D., Yovel, G., Davis, J. M., & O'Connor, K. B. (2011). Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: Evidence from the Cambridge Face Memory Test–Australian. *Cognitive Neuropsychology*, *28*(2), 109–146. https://doi.org/10.1080/02643294.2011.616880

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107. https://doi.org/10.1038/nn2024

Menon, V., & D'Esposito, M. (2022). The role of PFC networks in cognitive control and executive function. *Neuropsychopharmacology*, *47*(1), 90–103. https://doi.org/10.1038/s41386-021-01152-w

Miller, B. T., Vytlacil, J., Fegen, D., Pradhan, S., & D'Esposito, M. (2011). The Prefrontal Cortex Modulates Category Selectivity in Human Extrastriate Cortex. *Journal of Cognitive Neuroscience*, *23*(1), 1–10. https://doi.org/10.1162/jocn.2010.21516

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*(2), 159–179.

Minamoto, T., Tsubomi, H., & Osaka, N. (2017). Neural Mechanisms of Individual Differences in Working Memory Capacity: Observations From Functional Neuroimaging Studies. *Current Directions in Psychological Science*, *26*(4), 335–345. https://doi.org/10.1177/0963721417698800

Molnar, C. (2020). *Interpretable Machine Learning*.

Morgan, H. M., Klein, C., Boehm, S. G., Shapiro, K. L., & Linden, D. E. J. (2008). Working memory load for faces modulates P300, N170, and N250r. *Journal of Cognitive Neuroscience*, *20*(6), 989–1002. https://doi.org/10.1162/jocn.2008.20072

Moriya, J., & Sugiura, Y. (2012). High visual working memory capacity in trait social anxiety. *PLoS ONE*, *7*(4), 2–7. https://doi.org/10.1371/journal.pone.0034244

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G., & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory. *Journal of Anatomy*, *207*(1), 35–66. https://doi.org/10.1111/j.1469-7580.2005.00421.x

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*(3), 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing*, *5*(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4

Murphy, A. C., Bertolero, M. A., Papadopoulos, L., Lydon-Staley, D. M., & Bassett, D. S. (2020). Multimodal network dynamics underpinning working memory. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-15541-0

Myers, L., & Sirois, M. J. (2006). *Spearman Correlation Coefficients, Differences between*.

Nador, J. D., Zoia, M., Pachai, M. V., & Ramon, M. (2021). Psychophysical profiles in super-recognizers. *Scientific Reports*, *11*(1), 13184. https://doi.org/10.1038/s41598-021-92549-6

Natu, V. S., Jiang, F., Narvekar, A., Keshvari, S., Blanz, V., & O'Toole, A. J. (2010). Dissociable neural patterns of facial identity across changes in viewpoint. *Journal of Cognitive Neuroscience*, *22*(7), 1570–1582. https://doi.org/10.1162/jocn.2009.21312

Nelson, D. L., Evoy, C. L. M. C., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme and word fragment norms. *Behavior Research Methods, Instruments, & Computer*, *36*(3), 402–407.

Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9998–10003. https://doi.org/10.1073/pnas.1102433108

Nieto-Castañón, A. (2015). *Artifact Detection Tools (ART)* [Computer software]. https://www.nitrc.org/projects/artifact_detect/

Norman, G. R., & Streiner, D. L. (2014). *Biostatistics: The bare essentials* (Fourth edition). People's Medical Publishing House-USA.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, *74*(1), 18–22. https://doi.org/10.1037/0022-0663.74.1.18

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In *Variation in working memory* (pp. 49–75). Oxford University Press.

O'Craven, K. M., & Kanwisher, N. (2000). Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023. https://doi.org/10.1162/08989290051137549

Oh, D. W., Walker, M., & Freeman, J. B. (2021). Person knowledge shapes face identity perception. *Cognition*, *217*(April), 104889. https://doi.org/10.1016/j.cognition.2021.104889

Ooi, L. Q. R., Chen, J., Zhang, S., Kong, R., Tam, A., Li, J., Dhamala, E., Zhou, J. H., Holmes, A. J., & Yeo, B. T. T. (2022). Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI. *NeuroImage*, *263*, 119636. https://doi.org/10.1016/j.neuroimage.2022.119636

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. *Cognitive Science*, *38*(6), 1229–1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*(12), 505–510. https://doi.org/10.1016/S1364-6613(02)02005-3

Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., & Shibasaki, H. (2003). The neural basis of individual differences in working memory capacity: An fMRI study. *NeuroImage*, *18*(3), 789–797. https://doi.org/10.1016/S1053-8119(02)00032-0

Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: An fMRI study based on individual differences. *NeuroImage*, *21*(2), 623–631. https://doi.org/10.1016/j.neuroimage.2003.09.069

Palermo, R., O'Connor, K. B., Davis, J. M., Irons, J., & McKone, E. (2013). New Tests to Measure Individual Differences in Matching and Labelling Facial Expressions of Emotion, and Their Association with Ability to Recognise Vocal Emotions and Facial Identity. *PLoS ONE*, *8*(6), e68126. https://doi.org/10.1371/journal.pone.0068126

Pat, N., Wang, Y., Bartonicek, A., Candia, J., & Stringaris, A. (2022). Explainable Machine Learning Approach to Predict and Explain the Relationship between Task-based fMRI and Individual Differences in Cognition. *Cerebral Cortex*, 1–22. https://doi.org/10.1093/cercor/bhac235

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for semantically related and unrelated declarative information: The benefit of sleep, the cost of wake. *PLoS ONE*, *7*(3), 1–7. https://doi.org/10.1371/journal.pone.0033079

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825−2830.

Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *Journal of Neuroscience*, *32*(45), 15728–15736. https://doi.org/10.1523/JNEUROSCI.1953-12.2012

Pereira, F., Lou, B., Pritchett, B., Kanwisher, N., Botvinick, M., & Fedorenko, E. (2016). *Decoding of generic mental representations from functional MRI data using word embeddings*. 057216. https://doi.org/10.1101/057216

Perlstein, W. M., Carter, C. S., Noll, D. C., & Cohen, J. D. (2001). Relation of Prefrontal Cortex Dysfunction to Working Memory and Symptoms in Schizophrenia. *American Journal of Psychiatry, 158*, 1105–1113.

Perrodin, C., Kayser, C., Abel, T. J., Logothetis, N. K., & Petkov, C. I. (2015). Who is That? Brain Networks and Mechanisms for Identifying Individuals. *Trends in Cognitive Sciences*, *19*(12), 783–796. https://doi.org/10.1016/j.tics.2015.09.002

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*(6157), Article 6157. https://doi.org/10.1038/331585a0

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS Evidence for the Involvement of the Right Occipital Face Area in Early Face Processing. *Current Biology*, *17*(18), 1568–1573. https://doi.org/10.1016/j.cub.2007.07.063

Poirier, M., & Saint-Aubin, J. (1995). Memory for Related and Unrelated Words: Further Evidence on the Influence of Semantic Factors in Immediate Serial Recall. *The Quarterly Journal of Experimental Psychology Section A*, *48*(2), 384–404. https://doi.org/10.1080/14640749508401396

Poldrack, R. A. (2010). Subtraction and Beyond: The Logic of Experimental Designs for Neuroimaging. In *Foundational Issues in Human Brain Mapping*. The MIT Press. https://direct.mit.edu/books/edited-volume/3191/chapter/89558/Subtraction-and-Beyond-The-Logic-of-Experimental

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*(5756), 1963–1966. https://doi.org/10.1126/science.1117645

Popov, V., Zhang, Q., Koch, G. E., Calloway, R. C., & Coutanche, M. N. (2019). Semantic knowledge influences whether novel episodic associations are represented symmetrically or asymmetrically. *Memory and Cognition*, *47*(8), 1567–1581. https://doi.org/10.3758/s13421-019-00950-4

Postle, B. R. (2005). Delay-period activity in prefrontal cortex: One function is sensory gating. *Journal of Cognitive Neuroscience*, *17*(11), 1679–1690. https://doi.org/10.1162/089892905774589208

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005a). Portraits or people? Distinct representations of face identity in the human visual cortex. *Journal of Cognitive Neuroscience*, *17*(7), 1043–1057. https://doi.org/10.1162/0898929054475181

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005b). View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: An event-related fMRI study. *NeuroImage*, *24*(4), 1214–1224. https://doi.org/10.1016/j.neuroimage.2004.10.038

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 10.1016/j.neuroimage.2013.08.048. https://doi.org/10.1016/j.neuroimage.2013.08.048

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*(6002), 335. https://doi.org/10.1126/science.1191465

Quian Quiroga, R. (2020). No Pattern Separation in the Human Hippocampus. *Trends in Cognitive Sciences*, 1–14. https://doi.org/10.1016/j.tics.2020.09.012

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430. https://doi.org/10.1002/bs.3830120511

Rademaker, R. L., Bloem, I. M., De Weerd, P., & Sack, A. T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1650–1665. https://doi.org/10.1037/xhp0000110

Rafidi, N. S., Hulbert, J. C., Brooks, P. P., & Norman, K. A. (2018). Reductions in Retrieval Competition Predict the Benefit of Repeated Testing. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-29686-y

Ragland, J. D., Ranganath, C., Barch, D. M., Gold, J. M., Haley, B., MacDonald, A. W., Silverstein, S. M., Strauss, M. E., Yonelinas, A. P., & Carter, C. S. (2012). Relational and Item-Specific Encoding (RISE): Task Development and Psychometric Characteristics. *Schizophrenia Bulletin*, *38*(1), 114–124. https://doi.org/10.1093/schbul/sbr146

Ragland, J. D., Turetsky, B. I., Gur, R. C., Gunning-Dixon, F., Turner, T., Schroeder, L., Chan, R., & Gur, R. E. (2015). *Working Memory for Complex Figures: An fMRI Comparison of Letter and Fractal n-Back Tasks*.

Rajah, M. N., & McIntosh, A. R. (2005). Overlap in the functional neural systems involved in semantic and episodic memory retrieval. *Journal of Cognitive Neuroscience*, *17*(3), 470–482. https://doi.org/10.1162/0898929053279478

Ramot, M., Grossman, S., Friedman, D., & Malach, R. (2016). Covert neurofeedback without awareness shapes cortical network spontaneous connectivity. *Proceedings of the National*

Academy of Sciences of the United States of America, *113*(17), E2413-2420. https://doi.org/10.1073/pnas.1516857113

Ramot, M., Walsh, C., & Martin, A. (2019). Multifaceted integration—Memory for faces is subserved by widespread connections between visual , memory , auditory and social networks. *Journal of Neuroscience*.

Ranganath, C., DeGutis, J., & D'Esposito, M. (2004). Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Cognitive Brain Research*, *20*(1), 37–45. https://doi.org/10.1016/j.cogbrainres.2003.11.017

Ranganath, C., & D'Esposito, M. (2005). Directing the mind's eye: Prefrontal, inferior and medial temporal mechanisms for visual working memory. *Current Opinion in Neurobiology*, *15*(2), 175–182. https://doi.org/10.1016/j.conb.2005.03.017

Rasero, J., Sentis, A. I., Yeh, F.-C., & Verstynen, T. (2021). Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. *PLOS Computational Biology*, *17*(3), e1008347. https://doi.org/10.1371/journal.pcbi.1008347

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Pearson.

Rawson, K. A., Wissman, K. T., Vaughn, K. E., Rawson, K. A., Wissman, K. T., & Vaughn, K. E. (2015). Does Testing Impair Relational Processing? Failed Attempts to Replicate the Negative Testing Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Rawson, K. A., & Zamary, A. (2019). Why is free recall practice more effective than recognition practice for enhancing memory? Evaluating the relational processing hypothesis. *Journal*

of Memory and Language, *105*(July 2018), 141–152. https://doi.org/10.1016/j.jml.2019.01.002

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From Knowing to Remembering: The Semantic–Episodic Distinction. *Trends in Cognitive Sciences*, *23*(12), 1041–1057. https://doi.org/10.1016/j.tics.2019.09.008

Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research.* (2.3.6) [R]. Northwestern University. https://CRAN.R-project.org/package=psych.

Reverberi, C., Görgen, K., & Haynes, J. D. (2012). Compositionality of rule representations in human prefrontal cortex. *Cerebral Cortex*, *22*(6), 1237–1246. https://doi.org/10.1093/cercor/bhr200

Richler, J., Cheung, O., & Gauthier, I. (2012). Holistic Processing Predicts Face Recognition. *Psychological Science*, *22*(4), 464–471. https://doi.org/10.1177/0956797611401753.Holistic

Rickard, T. C., & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin and Review*, *25*(3), 847–869. https://doi.org/10.3758/s13423-017-1298-4

Riley, M. R., & Constantinidis, C. (2016). Role of Prefrontal Persistent Activity in Working Memory. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00181

Rissman, J., Gazzaley, A., & D'Esposito, M. (2008). Dynamic Adjustments in Prefrontal, Hippocampal, and Inferior Temporal Interactions with Increasing Visual Working Memory Load. *Cerebral Cortex*, *18*(7), 1618–1629. https://doi.org/10.1093/cercor/bhm195

Rissman, J., & Wagner, A. D. (2012). Distributed Representations in Memory: Insights from Functional Brain Imaging. *Annual Review of Psychology*. https://doi.org/10.1146/annurev-psych-120710-100344

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends in Cognitive Sciences*, 1–17. https://doi.org/10.1016/j.tics.2019.06.007

Rose, N. S. (2020). The Dynamic-Processing Model of Working Memory. *Current Directions in Psychological Science*, *29*(4), 378–387. https://doi.org/10.1177/0963721420922185

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience*, *8*(1), 107–113. https://doi.org/10.1038/nn1370

Rowland, C. A. (2014). The Effect of Testing Versus Restudy on Retention: A Meta-Analytic Review of the Testing Effect. *Psychological Bulletin*. https://doi.org/10.1037/a0037559

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

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology*, *21*(2), 137–147. https://doi.org/10.1080/14640746908400206

Ryan, J. J., & Lopez, S. J. (2001). Wechsler Adult Intelligence Scale-III. In W. I. Dorfman & M. Hersen (Eds.), *Understanding Psychological Assessment* (pp. 19–42). Springer US. https://doi.org/10.1007/978-1-4615-1185-4_2

Ryan, L., Cox, C., Hayes, S. M., & Nadel, L. (2008). Hippocampal activation during episodic and semantic memory retrieval: Comparing category production and category cued recall. *Neuropsychologia*, *46*(8), 2109–2121. https://doi.org/10.1016/j.neuropsychologia.2008.02.030

Rypma, B., Berger, J. S., & Esposito, M. D. (2002). The Influence of Working-Memory Demand and Subject Performance on Prefrontal Cortical Activity. *Journal of Cognitive Neuroscience*, *14*(5), 721–731.

Rypma, B., & D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: Effects of memory load and individual differences. *Proceedings of the National Academy of Sciences*, *96*, 6558–6563. https://doi.org/10.1038/sj.bdj.4809188

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*(4), 1088–1096. https://doi.org/10.1016/j.neuroimage.2005.12.062

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*. https://doi.org/10.1016/j.cobeha.2014.07.005

Schöttner, M., Bolton, T. A. W., Patel, J., Nahálka, A. T., Vieira, S., & Hagmann, P. (2023). Exploring the latent structure of behavior using the Human Connectome Project's data. *Scientific Reports*, *13*(1), 713. https://doi.org/10.1038/s41598-022-27101-1

Schwartz, L., & Yovel, G. (2018). Learning Faces as Concepts Rather Than Percepts Improves Face Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1733–1747.

Schwartz, L., & Yovel, G. (2019). Learning Faces as Concepts Rather Than Percepts Improves Face Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1733–1747.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*(1), 11–21. https://doi.org/10.1136/jnnp.20.1.11

Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Research*, *128*, 53–67. https://doi.org/10.1016/j.visres.2016.09.010

Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal cortex to episodic memory. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn.2017.6

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): A self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science*, *2*(6), 140343. https://doi.org/10.1098/rsos.140343

Shannon, B. J., Raichle, M. E., Snyder, A. Z., Fair, D. A., Mills, K. L., Zhang, D., Bache, K., Calhoun, V. D., Nigg, J. T., Nagel, B. J., Stevens, A. A., & Kiehl, K. A. (2011). Premotor functional connectivity predicts impulsivity in juvenile offenders. *Proceedings of the*

National Academy of Sciences of the United States of America, *108*(27), 11241–11245. https://doi.org/10.1073/pnas.1108241108

Sherry, D. F., & Schacter, D. L. (1987). The Evolution of Multiple Memory Systems. *Psychological Review*, *94*(4), 439–454. https://doi.org/10.1037/0033-295X.94.4.439

Silverman, J. (2022). *RcppHungarian: Solves Minimum Cost Bipartite Matching Problems* (0.2) [R]. https://CRAN.R-project.org/package=RcppHungarian

Simmons, S., & Estes, Z. (2006). Using latent semantic analysis to estimate similarity. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2169–2173. https://doi.org/10.1.1.534.1407

Sinclair, A. H., & Barense, M. D. (2019). Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation. *Trends in Neurosciences*, 8–13.

Sliwinska, M. W., Brown, L. R., Earl, M., O'Gorman, D., Pollicina, G., Burton, A. M., & Pitcher, D. (2022). Face learning via short real-world social interations induces changes in face-selective brain areas and hippocampus. *Perception*, *51*(8), 1689–1699. https://doi.org/10.1177/03010066221098728

Snowden, J. S., Griffiths, H. L., & Neary, D. (1996). Semantic-episodic memory interactions in semantic dementia: Implications for retrograde memory function. *Cognitive Neuropsychology*, *13*(8), 1101–1139. https://doi.org/10.1080/026432996381674

Solomon, S. H., & Schapiro, A. C. (2020). Semantic Search as Pattern Completion across a Concept. *Trends in Cognitive Sciences*, *24*(2), 95–98. https://doi.org/10.1016/j.tics.2019.12.003

Solomon, S. H., & Thompson-Schill, S. L. (2017). Finding features, figuratively. *Brain and Language*, *174*, 61–71. https://doi.org/10.1016/j.bandl.2017.07.002

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89. https://doi.org/10.1016/j.tics.2013.12.001

Sreenivasan, K. K., & D'Esposito, M. (2019). The what, where and how of delay activity. *Nature Reviews Neuroscience*, 10–14. https://doi.org/10.1038/s41583-019-0176-7

Sripada, C., Rutherford, S., Angstadt, M., Thompson, W. K., Luciana, M., Weigard, A., Hyde, L. H., & Heitzeg, M. (2020). Prediction of neurocognition in youth from resting state fMRI. *Molecular Psychiatry*, *25*(12), 3413–3421. https://doi.org/10.1038/s41380-019-0481-6

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. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00954

Stark, S. M., Kirwan, C. B., & Stark, C. E. L. (2019). Mnemonic Similarity Task: A Tool for Assessing Hippocampal Integrity. *Trends in Cognitive Sciences*, *23*(11), 938–951. https://doi.org/10.1016/j.tics.2019.08.003

Steiger, J. H., & Lind, J. (1980). *Statistically based tests for the number of common factors.* annual spring meeting of the Psychometric Society, Iowa City.

Sternberg, S. (1969). The Discovery of Processing Stages: Extensions of Donders' Method. *Acta Psychologica*, *30*, 267–315.

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

Storm, B. C., Friedman, M. C., Murayama, K., & Bjork, R. A. (2014). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(1), 115–124. https://doi.org/10.1037/a0034252

Sugiura, M., Sassa, Y., Watanabe, J., Akitsuki, Y., Maeda, Y., Matsue, Y., Fukuda, H., & Kawashima, R. (2006). Cortical mechanisms of person representation: Recognition of famous and personally familiar names. *NeuroImage*, *31*(2), 853–860. https://doi.org/10.1016/j.neuroimage.2006.01.002

Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. (2020). Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. *Biological Psychiatry*, *88*(11), 818–828. https://doi.org/10.1016/j.biopsych.2020.02.016

Tetereva, A., Li, J., Deng, J. D., Stringaris, A., & Pat, N. (2022). Capturing brain-cognition relationship: Integrating task-based fMRI across tasks markedly boosts prediction and test-retest reliability. *NeuroImage*, *263*, 119588. https://doi.org/10.1016/j.neuroimage.2022.119588

Thompson, J. M., Hamilton, C. J., Gray, J. M., Quinn, J. G., Mackin, P., Young, A. H., & Ferrier, I. N. (2006). Executive and visuospatial sketchpad resources in euthymic bipolar disorder: Implications for visuospatial working memory architecture. *Memory, 14*(4), 437–451. https://doi.org/10.1080/09658210500464293

Thurstone, L. L. (1938). *Primary Mental Abiliites*. Psychometric monographs.

Thurstone, L. L. (1947). *Multiple factor analysis*. University of Chicago Press.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex ´. *Nature*, *166*(2003), 751–754.

Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective and Behavioral Neuroscience*, *5*(2), 144–155. https://doi.org/10.3758/CABN.5.2.144

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, *17*(6), 520–525. https://doi.org/10.1093/bioinformatics/17.6.520

Tsukiura, T., Mano, Y., Sekiguchi, A., Yomogida, Y., Hoshi, K., Kambara, T., Takeuchi, H., Sugiura, M., & Kawashima, R. (2010). Dissociable roles of the anterior temporal regions in successful encoding of memory for person identity information. *Journal of Cognitive Neuroscience*, *22*(10), 2226–2237. https://doi.org/10.1162/jocn.2009.21349

Tsukiura, T., Mochizuki-Kawai, H., & Fujii, T. (2006). Dissociable roles of the bilateral anterior temporal lobe in face-name associations: An event-related fMRI study. *NeuroImage*, *30*(2), 617–626. https://doi.org/10.1016/j.neuroimage.2005.09.043

Tsukiura, T., Suzuki, C., Shigemune, Y., & Mochizuki-Kawai, H. (2008). Differential contributions of the anterior temporal and medial temporal lobe to the retrieval of memory for person identity information. *Human Brain Mapping*, *29*(12), 1343–1354. https://doi.org/10.1002/hbm.20469

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus*, *8*(3), 198–204. https://doi.org/10.1002/(SICI)1098-1063(1998)8:3<198::AID-HIPO2>3.0.CO;2-G

Turk-Browne, N. B., Norman-Haignere, S. V, & McCarthy, G. (2010). Face-Specific Resting Functional Connectivity between the Fusiform Gyrus and Posterior Superior Temporal Sulcus. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00176

Turner, B. O., Mumford, J. A., Poldrack, R. A., & Ashby, F. G. (2012). Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage*, *62*(3), 1429–1438. https://doi.org/10.1016/j.neuroimage.2012.05.057

Unsworth, N., & Engle, R. W. (2007a). On the Division of Short-Term and Working Memory: An Examination of Simple and Complex Span and Their Relation to Higher Order Abilities. *Psychological Bulletin*, *133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., & Engle, R. W. (2007b). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working Memory and Fluid Intelligence: Capacity, Attention Control, and Secondary Memory Retrieval. *Journal of Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1007/s11065-015-9294-9.Functional

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). Working Memory Delay Activity Predicts Individual Differences in Cognitive Abilities. *Journal of Cognitive Neuroscience*, *27*(5), 853–865. https://doi.org/10.1162/jocn

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. https://doi.org/10.3758/BF03192720

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage, 80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Van Gulick, A. E., McGugin, R. W., & Gauthier, I. (2016). Measuring nonvisual knowledge about object categories: The Semantic Vanderbilt Expertise Test. *Behavior Research Methods*, *48*(3), 1178–1196. https://doi.org/10.3758/s13428-015-0637-5

van Kesteren, M. T. R., Rignanese, P., Gianferrara, P. G., Krabbendam, L., & Meeter, M. (2020). Congruency and reactivation aid memory integration through reinstatement of prior knowledge. *Scientific Reports*, *10*(4776), 1–13. https://doi.org/10.1038/s41598-020-61737-1

Vaughn, K. E., & Rawson, K. A. (2014). Effects of criterion level on associative memory: Evidence for associative asymmetry. *Journal of Memory and Language, 75*, 14–26. https://doi.org/10.1016/j.jml.2014.04.004

Velenosi, L. A., Wu, Y.-H., Schmidt, T. T., & Blankenburg, F. (2020). Intraparietal sulcus maintains working memory representations of somatosensory categories in an adaptive, context-dependent manner. *NeuroImage*, *221*, 117146. https://doi.org/10.1016/j.neuroimage.2020.117146

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, *141*, 217–227. https://doi.org/10.1016/j.visres.2016.12.014

Verosky, S. C., Todorov, A., & Turk-Browne, N. B. (2013). Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neuropsychologia*, *51*(11), 2100–2108. https://doi.org/10.1016/j.neuropsychologia.2013.07.006

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

Volfart, A., Jonas, J., Maillard, L., Coulbois-Colnat, S., & Rossion, B. (2020). Neurophysiological evidence for crossmodal (face-name) person-identity representation in the human left ventral temporal cortex. *PLOS Biology*, *18*(4), e3000659.

Votruba, K. L., & Langenecker, S. A. (2013). Factor Structure, construct validity, and age and education-based normative data for the Parametric Go/No-Go Test. *Journal of Clinical and Experimental Neuropsychology*, *35*(2), 132–146.

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169–177. https://doi.org/10.1177/0956797611420575

Wang, Y., Collins, J. A., Koski, J., Nugiel, T., Metoki, A., & Olson, I. R. (2017). Dynamic neural architecture for social knowledge retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(16), E3305–E3314. https://doi.org/10.1073/pnas.1621234114

Waters, G. S., & Caplan, D. (1996). The Measurement of Verbal Working Memory Capacity and Its Relation to Reading Comprehension. *The Quarterly Journal of Experimental Psychology Section A, 49*(1), 51–79. https://doi.org/10.1080/713755607

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers*, *35*(4), 550–564. https://doi.org/10.3758/BF03195534

Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Watson, C. G. (2020). *brainGraph: Graph Theory Analysis of Brain MRI Data* (3.0.0) [R]. https://CRAN.R-project.org/package=brainGraph

Wechsler, D. (1945). A Standardized Memory Scale for Clinical Use. *Journal of Psychology: Interdisciplinary and Applied*, *19*(1), 87–95. https://doi.org/10.1080/00223980.1945.9917223

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*(JUL), 1–22. https://doi.org/10.3389/fpsyg.2013.00433

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: The example of face recognition. *Cognitive Neuropsychology*, *29*(5–6), 360–392. https://doi.org/10.1080/02643294.2012.753433

Wilmer, J. B., Germine, L. T., & Nakayama, K. (2014). Face recognition: A model specific ability. *Frontiers in Human Neuroscience*, *8*(October), 1–5. https://doi.org/10.3389/fnhum.2014.00769

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*(4), 582–589. https://doi.org/10.1038/nn.3973

Wing, E. A., Burles, F., Ryan, J. D., & Gilboa, A. (2022). The structure of prior knowledge enhances memory in experts by reducing interference. *Proceedings of the National Academy of Sciences*, *119*(26), e2204172119. https://doi.org/10.1073/pnas.2204172119

Wing, E. A., Geib, B. R., Wang, W. C., Monge, Z., Davis, X. S. W., & Cabeza, R. (2020). Cortical overlap and cortical-hippocampal interactions predict subsequent true and false memory. *Journal of Neuroscience*, *40*(9), 1920–1930. https://doi.org/10.1523/JNEUROSCI.1766-19.2020

Xiao, Y., Lin, Y., Ma, J., Qian, J., Ke, Z., Li, L., Yi, Y., Zhang, J., Cam-CAN, & Dai, Z. (2021). Predicting visual working memory with multimodal magnetic resonance imaging. *Human Brain Mapping*, *42*(5), 1446–1462. https://doi.org/10.1002/hbm.25305

Xu, X., & Biederman, I. (2010). Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. *Journal of Vision*, *10*(14), 1–13. https://doi.org/10.1167/10.14.1

Xu, Y. (2020). Revisit once more the sensory storage account of visual working memory. *Visual Cognition*, *28*(5–8), 433–446. https://doi.org/10.1080/13506285.2020.1818659

Yang, H., Susilo, T., & Duchaine, B. (2016). The Anterior Temporal Face Area Contains Invariant Representations of Face Identity That Can Persist Despite the Loss of Right FFA and OFA. *Cerebral Cortex*, *26*(3), 1096–1107. https://doi.org/10.1093/cercor/bhu289

Ye, Z., Shi, L., Li, A., Chen, C., & Xue, G. (2020). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *eLife*, *9*, 1–51. https://doi.org/10.7554/eLife.57023

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Yovel, G., & Schwartz, L. (2016). The Roles of Perceptual and Conceptual Information in Face Recognition. *Journal of Experimental Psychology: General*, *145*(11), 1493–1511.

Yu, Q., & Shim, W. M. (2017). Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *NeuroImage*, *157*, 97–107. https://doi.org/10.1016/j.neuroimage.2017.05.055

Zanto, T. P., & Gazzaley, A. (2009). Neural Suppression of Irrelevant Information Underlies Optimal Working Memory Performance. *The Journal of Neuroscience*, *29*(10), 3059–3066. https://doi.org/10.1523/JNEUROSCI.4621-08.2009

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience*, *14*(5), 656–661. https://doi.org/10.1038/nn.2773

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2018). The Impact of Inserting an Additional Mental Process. *Computational Brain & Behavior*, *1*(1), 22–35. https://doi.org/10.1007/s42113-018-0002-8

Zhen, Z., Fang, H., & Liu, J. (2013). The Hierarchical Brain Network for Face Recognition. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059886

Zhu, J., Li, Y., Fang, Q., Shen, Y., Qian, Y., Cai, H., & Yu, Y. (2021). Dynamic functional connectome predicts individual working memory performance across diagnostic categories. *NeuroImage: Clinical*, *30*, 102593. https://doi.org/10.1016/j.nicl.2021.102593

Zhu, Q., Zhang, J., Luo, Y. L. L., Dilks, D. D., & Liu, J. (2018). *Resting-State Neural Activity across Face-Selective Cortical Regions Is Behaviorally Relevant*. https://doi.org/10.1523/JNEUROSCI.0873-11.2011

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x