

How prior knowledge scaffolds memory - a “kind of naturalistic” approach

Jiawen Huang

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
under the Executive Committee  
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2025

© 2025

Jiawen Huang

All Rights Reserved

## **Abstract**

How prior knowledge scaffolds memory - a “kind of naturalistic” approach

Jiawen Huang

In our daily life, we constantly use our structured knowledge built from repeated experiences, often referred to as schemas, to provide a scaffold for perceiving, understanding, and remembering what happens around us. This is a difficult process to study, because these kinds of prior knowledge we use in daily life are typically complex and could take a long time to develop. In this dissertation, I present findings from a few experiments that take a “kind of naturalistic” approach, which allows me to capture some complexity of such knowledge, yet is still feasible in a lab setting. In Chapter 1-3, I present behavioral and neuroimaging findings from a board game paradigm. In Chapter 1, I show that the development of knowledge in the board game increased predictive eye-movements during sequence encoding, which in turn improves memory. In Chapter 2, I show that there are two potentially distinct processes when people remember something schema-consistent: probability given the context and prediction accuracy. I show that both contribute to better memory, but through different mechanisms. In Chapter 3, I use functional magnetic resonance imaging (fMRI) to show that different default mode network (DMN) regions respond to these two processes. In addition, I show differences in how the brain makes memory- vs. schema-based predictions. In Chapter 4, I use fMRI to look at people using the method of loci (MoL) to encode a list of words. I found that using MoL creates conjunctive representations in DMN that are more than the sum of its parts. Overall, the dissertation highlights the importance of prediction in the relationship between schemas and memory, and the importance of DMN in this process.

# Table of Contents

Table of Contents .....	i
List of Figures .....	iv
Acknowledgments.....	vi
Introduction.....	1
Chapter 1: Schema-based predictive eye movements support sequential memory encoding .....	13
1.1 Introduction.....	14
1.2 Methods.....	17
1.3 Results.....	28
1.3.1 Memory and gameplay improvement .....	28
1.3.2 Modeling schema-related memory improvement .....	30
1.3.3 Eye movements became more predictive over training.....	35
1.3.4 Better performance in better players is mediated by more schema-based predictions.....	37
1.3.5 Model-free measures of prediction confidence and surprise .....	38
1.4 Discussion.....	40
1.4.1 Schema-related memory effects and how they develop .....	40
1.4.2 The development of predictions and their influences on memory.....	42

Chapter 2: Accurate predictions facilitate robust memory encoding independently from stimulus probability .....	46
2.1 Introduction.....	47
2.2 Method .....	51
2.3 Results.....	60
2.3.1 Manipulation check.....	60
2.3.2 Move probability and prediction accuracy both improve subsequent memory.....	61
2.3.3 Eye movements reveal multiple distinct retrieval strategies.....	65
2.3.4 Prediction accuracy, but not probability, reduces schema use at retrieval .....	68
2.4 Discussion.....	70
2.4.1 Accurate predictions facilitate memory .....	70
2.4.2 Accurate predictions led to reduced reliance on schema at retrieval.....	73
2.4.3 Methodological implications .....	75
Chapter 3: Distinct neural representation of different types of predictions and prediction errors .....	77
3.1 Introduction.....	77
3.2 Methods.....	80
3.3 Results.....	89
3.4 Discussion.....	99

Chapter 4: Binding items to contexts through conjunctive neural representations with the Method of Loci .....	105
4.1 Introduction.....	106
4.2 Methods.....	111
4.3 Results.....	126
4.3.1 Widespread representations of locus and item by themselves and during encoding.....	129
4.3.2 Encoding residuals track idiosyncratic semantic combinations of loci and items .....	131
4.3.3 Widespread and robust conjunctive representation in the brain .....	135
4.3.4 Relationship between conjunctive representation in the DMN and training and behavior.....	136
4.4 Discussion.....	141
4.4.1 A new approach to study conjunctive representation and MoL .....	142
4.4.2 Conjunctive representations in the DMN .....	144
4.4.3 Relating conjunctive coding to creativity and concept combination .....	146
Conclusion .....	149
References.....	157
Appendix A: Chapter 2 Supplement.....	187
Appendix B: Chapter 4 Supplement .....	196

## List of Figures

Figure 1.1 A schematic of the experimental method .....	17
Figure 1.2 Participants' performance in memory and gameplay .....	29
Figure 1.3 The effect of schema consistency on memory and its development across training sessions .....	33
Figure 1.4 Using eye movements to reveal encoding strategies .....	35
Figure 1.5 The relationship between prediction, playing strengths, and recall accuracy .	36
Figure 1.6 Model-free measure of prediction and their relationship with memory .....	39
Figure 2.1 Experimental design. ....	53
Figure 2.2 Real-time generation of the moves.....	58
Figure 2.3 Impact of move probability and trial-wise predictions on recall performance	64
Figure 2.4 Modeling and predicting eye movements at retrieval .....	67
Figure 3.1 Overview of the task structure.....	82
Figure 3.2 Memory for the sequences and the relationship with move probability .....	92
Figure 3.3 Measurement of prediction accuracy and its correlate with brain activity during encoding .....	93
Figure 3.4 ROI analysis of coefficient of move probability and prediction accuracy in predicting univariate activity in the brain .....	94
Figure 3.5 Analysis of retrieval activity .....	96
Figure 3.6 Behavioral and brain results of the prediction task (top) .....	98
Figure 4.1 Illustration of the paradigm and behavioral performance .....	128
Figure 4.2 Brain regions representing loci and items alone and during encoding.....	130
Figure 4.3 Encoding residuals track semantic similarity across stories.....	134

Figure 4.4 Conjunctive representation in the brain.....	137
Figure 4.5 Measuring the novelty of generated stories and linking this measure to brain activity.....	141
Figure A.1 Residual of pupil size as a function of time.....	192
Figure A.2 Distribution of perplexity in the boards shown to the participants.....	193
Figure A.3 Illustration of the number of “landmarks” measure .....	194
Figure B.1 Results of the ROI analyses done in hippocampal ROIs .....	196

## Acknowledgments

I came to New York City without knowing anyone here almost exactly 5 years ago during peak COVID, to be stuck in my apartment after already being isolated in London for 6 months. Little did I know at the time that this would be the best 5 years of my life so far. I have learned how to do science. I have completed projects that I am proud of. I have learned a lot about myself. I have traveled, become (somewhat) fluent in Spanish, and gotten (kind of) good at dancing salsa. I have become a better person. None of it would have been possible without the people I am going to thank.

First of all, big big big thank you to Chris. I have been excited to write the acknowledgement since my first year after working with you for a few months. You are such a smart, creative, patient, humble, and kind person, everything one can ask for in a mentor. Talking to you in our weekly meetings has been one of the things I look forward to the most in grad school, in good and bad days – when I found something exciting with my analyses, I was always happy to share them with you; when I got stuck with an analysis, you always helped me find out what was wrong; when I was not very productive in some weeks, you made me feel better and we got to talk about random science ideas or random video games we had been playing; when I was going through a hard time, you allowed me to take a break and gave me space. From you, I have not only learned how to do good science, but also how to work with others, and to be a kind person. I wish I could stay in your lab forever.

I also want to thank my past mentors David Shanks and Jeremy Skipper. David got me into memory, and Jeremy got me into using naturalistic stimuli and prediction. The dissertation would not have the shape it has now without your influence in my academic journey.

Many thanks to my amazing committee members, you have been amazing mentors and inspired so much of my work. To Serra Favila, for teaching me to use eye-tracking and to write good experiment codes, for inspiring me to do research with real-time eye-tracking and eye-tracking + fMRI, for always providing such useful insights and helping me answer questions in lab meetings. To Mariam Aly, for all the feedback in my work in lab meetings and for an amazing seminar class that inspired my studies into different kinds of prediction. To Lila Davachi and Daphna Shohamy, for the feedback on my work over the years, as well as for letting me hang out with the cool people in your lab. I am so exciting to hear your thoughts on my work in this dissertation.

Starting grad school during a pandemic was a lonely process, but I was lucky to have a cool cohort of people to struggle through it together. Thank you Daniel, Victoria, and Wen. I like to believe that we manifested things going back in person by changing our group chat into “In Person Only 2021 Ph.Ds”. One of my favorite things about the Psychology Department is the amazing community. Thank you to my friends that I have spent so much time with: Caroline, John Andrew, Chey, Claudia, and Arlene, for all the food, happy hours, hiking, board game nights, weekend getaways, deep conversations, and support; and to the wonderful people in the Alyssano group and beyond: Narjes, Craig, Zall, Taylor, Matt, Manasi, Hannah, Mike, Sam, Halle, Alex, Wangjing, Yifan, Moncia, Vasiki, and many, many more. My science journey would not have been the same without you.

Thank you to my friends outside of Psychology department. Thank you to my BFF, Yifang, for being a constant in my life. It gives me courage to explore the world and going out of my comfort zone knowing that you will always be there. To Rio, I’m so glad you moved to NYC this year. To Tim, for being my Spanish buddy. Thank you to my amazing salsa community in

NYC. To my salsa teachers, Nancy and Serena, for helping me grow as a dancer. To my friends, Matilda, Shirley, Busola, Aaron, Felipe, José, Kate, Lauren, Lyse, Sean. You were here to see me grow as a dancer and as a person and you made NYC feel like home to me.

Lastly, I want to thank my family. To my grandparents, aunts and uncles that have given me so much love and made me who I am. And to my parents. I'm continuously amazed by how much you were able to achieve with the resources you had. I can't be happier for where I am right now, and I'm forever grateful that how you helped me get here.

## Introduction

When I see people dance Salsa, I am able to recognize the moves being carried out, predict the next part of music and what might happen next given the current position of the dancers, and remember (and even try to carry out later) some of the moves. None of these was possible a year ago when I did not know anything about Salsa. Prior knowledge like the dance knowledge I acquired this past year, characterized by an associative network structure, basis on multiple episodes, lack of unit detail, and adaptability, are referred to as schemas in the literature (Ghosh & Gilboa, 2014). They play a huge role in how we remember events in our life. They allow us to remember our daily experiences in great detail, such as events in a movie (Chen et al., 2017), which is in sharp contrast to people's ability to remember a simple list of unrelated words, when people have much more limited capacity (Murdock, 1974).

In the 1970s, a lot of research looked at the behavioral mechanisms by which schemas influence event perception and memory, discovering mechanisms through which schemas help memory at encoding and retrieval (J. R. Anderson, 1981; R. C. Anderson & Pichert, 1978). More recently, neuroscience research including animal models and human neuroimaging started looking at the neural mechanisms of schemas and schema-facilitated memory. There are two main types of paradigms that were commonly used - one where participants used their existing schemas of the complex, dynamic world (Baldassano et al., 2018; Bower et al., 1979; Graesser & Nakamura, 1982); the other where participants learned simple novel schemas such as speed of little characters, or location-item pairings (Brod, 2021; van Buuren et al., 2014). Both paradigms have their strengths and shortcomings - existing schemas of the world (like my dance schema) are more naturalistic and closer to people's experiences, but typically take a long time to

develop, and are difficult to model. On the other hand, the simple schemas are easier to learn, and can be easily modeled, yet the extent to which they make use of the same cognitive and neural mechanisms as natural schemas, (since, for example, there are a very small number of potential next stimuli) is unclear.

In this dissertation, I took advantage of recent research with naturalistic paradigms, increased computational power, and novel neuroimaging analyses methods, and used a “kind of naturalistic” approach that sits in the middle of the two types of paradigms and tries to capture the advantages of both. I allow people to develop a relatively complex schema that is still modellable, and look at how the development of these schemas facilitate memory for relatively simple stimuli that could be remembered without a schema. I begin by looking at the behavioral and neural mechanisms by which the knowledge of a board game influences how people perceive, predict, and remember moves (or move sequences) in the game (Chapter 1-3). Then, I pivot to how spatial scaffolds like a memory palace facilitate memory for word lists (Chapter 4). Below, I review some past literature that inspired this dissertation.

### **Behavioral work on how schema and expertise facilitate memory**

One of the pioneering studies in schema is (Bartlett, 1932), which presented stories from a different culture to the participants. They found that participants’ recalls of unfamiliar details in the story were changed to details consistent with their knowledge. This sparked interest in research in schemas, which can be defined as adaptable knowledge structures reflecting generalized information abstracted from multiple episodic experiences (Ghosh & Gilboa, 2014). These early studies typically use common, existing schemas that people have developed over the course of a lifetime, such as the process of eating at a restaurant, doing laundry, and look at how

they influence memory (Alba & Hasher, 1983; R. C. Anderson et al., 1983; R. C. Anderson & Pichert, 1978; Bower et al., 1979; Lampinen et al., 2000).

Somewhat in parallel, another line of research looked at the effect of expertise, such as in chess, on different aspects of cognition, including memory. The pioneering work started with Chase & Simon, (1973) which looked at expertise in chess. They found that chess experts were able to use their knowledge to divide a valid chess board into meaningful chunks, enabling them to reconstruct the boards easily. However, for boards with random arrangement, experts did not show much advantage over novices. Looking at expertise in a specific field allows researchers to investigate how memory works when relevant schemas are absent, which is difficult to do with the common schemas that most people have, like in previous research. However, a challenge of studying expertise is the amount of time specific skills typically take to develop, and it is difficult to see how the development of a schema influences memory. Another issue is that these specific skills are usually highly complex, making them difficult to model. To address some of these limitations in previous research, in Chapter 1, I took advantage of a board game paradigm developed by van Opheusden et al. (2023), which is novel for participants, learnable in a short scale of time, and has an ideal complexity that allows for large state space yet is still modellable. I looked at how the development of expertise in the game changed how people remember move sequences.

Schemas and expertise can act as a scaffold for memories through multiple mechanisms that occur at different stages of remembering. During encoding, schemas allow us to create meaningful representations of the stimuli by integrating it with the schema (Bransford & Johnson, 1972; Chase & Simon, 1973; Shin et al., 2021). During retrieval, schemas allow us to

come up with possible outcomes in a given context, which could provide a cue, making the recall task a task of recognition (R. C. Anderson & Pichert, 1978; Watkins & Gardiner, 1979).

An example of how people can strategically use a schema to create meaningful representations of stimuli is an ancient mnemonic technique called the Method of Loci (MoL), or memory palace. People good at this technique are memory “experts”, and can compete in championships where they are able to remember a long list of up to over 100 words in a short period of time. The technique involves building and consolidating a spatial layout (memory palace) in the mind, with ordered locations (loci) in the memory palace that serve as a schematic scaffold. During encoding, each item to be remembered is combined with each of the loci in order by forming a meaningful connection between the two, such as imagining an event involving that item occurring at that locus. During retrieval, people mentally retrace their steps through the memory palace in order, using each locus as a cue to recall its associated item. In some sense, MoL works because it allows people to remember word lists like events in real life, by linking event details (words) to their schema (locus in the memory palace). It can therefore provide a testbed for how schemas are used to facilitate memory in a “kind of naturalistic” setting. Chapter 4 of this dissertation looks at the neural mechanisms of this process, how they develop over time, and how they are linked to behavior.

### **The relationship between schema, prediction, and memory**

In cognitive science, the idea of the brain as a “prediction machine” has been proposed (Bar, 2009; Clark, 2013; Friston, 2010). For example, in language research, it has been proposed that when we have a conversation with someone, we constantly try to predict and covertly produce what our conversation partner is saying (Pickering & Gambi, 2018). It makes intuitive sense to link schema to prediction, because predictions depend on some kind of past experience

(Cheung & Bar, 2012; Cowan et al., 2021). There is also evidence that experts automatically make predictions in the field they have expertise in, such as in the context of basketball (Didierjean & Marmèche, 2005; Gorman et al., 2011, 2012). Yet, the process by which the development of schema influences prediction is unclear due to the typical complexity of expert skills mentioned in the previous section. Another challenge is the methodological difficulty in measuring prediction, which is often considered an implicit process and integral part of perception (Pickering & Gambi, 2018; Skipper, 2015), and therefore explicitly asking for a prediction might disrupt or alter these automatic processes. As a result, despite past research linking expertise and schema to prediction, to our knowledge, no studies have shown when prediction comes online and how it develops during schema development. In Chapter 1, I measure participants' gaze while watching a sequence of moves from the game, because gaze has been found to be able to distinguish level of expertise in participants with machine learning algorithms (Y. Liu et al., 2009). Specifically, I look at predictive eye-movements where participants look at probable next moves while watching a sequence, and study how it develops over the course of the development of the schema.

A popular view in the field of cognitive neuroscience about the relationship between prediction and memory is that prediction errors are beneficial for episodic memory. Previous work with animal models showed the crucial role of reward prediction error in learning, which can be explained elegantly by the firing of dopamine neurons (Schultz, 1998). Human research has provided support for this view, showing improved learning (Den Ouden et al., 2012) and memory when there is a prediction error (Bein et al., 2021; Marvin & Shohamy, 2016; Quent et al., 2022; Rouhani et al., 2018). Since schemas allow people to make automatic predictions about the upcoming information (Gorman et al., 2012), and since prediction errors could be beneficial

for memory, one possible mechanism previously unexplored is that developing schemas and expertise improve memory because they enable predictions of the upcoming stimuli given the current context. In Chapter 1, I explore this hypothesis.

Despite the longstanding history of the idea of prediction, this term is a very broad term that can refer to a lot of different processes across studies. Consider the following scenarios: 1. A social scientist predicts societal changes (Grossmann et al., 2024). 2. Someone re-watches a movie multiple times when they can perfectly anticipate what is going to happen (C. S. Lee et al., 2021). 3. A rat presses a lever, an action that it had learnt to be associated with a reward (Schultz, 1998). 4. While remembering a list of words, one of the words in the middle is in red, when all the other words are in black. While some kind of predictive process can be said to be happening in each scenario, the underlying processes are likely very different. In a recent opinion piece, (Ortiz-Tudela, Nicholls, et al., 2023) proposed that predictions can vary in five different dimensions - flow of information, mnemonic origin, specificity, complexity of reactivated information, and temporal precision. With these distinctions in mind, and considering the different ways memory can be tested (e.g., recognition (Bein et al., 2021); forced choice questions (Quent et al., 2022); free recall (Lew & Howe, 2017)), it is perhaps not surprising that some previous studies have failed to find benefits of prediction error on memory (Höltje & Mecklinger, 2022; Ortiz-Tudela et al., 2021; Poskanzer et al., 2025; Van Kesteren et al., 2013).

This idea that there is subtlety in what is considered a prediction is an important theme in this current dissertation. Due to the challenges in measuring prediction mentioned above, when studying the impact of prediction and prediction error on memory, prediction is often a process that is implied rather than measured. For example, in Bein et al., (2021), participants learned some sort of association, and were shown violation of these learned associations, which the

experimenters defined as prediction errors, implying that participants were making a prediction and then experiencing a violation. In Sherman & Turk-Browne (2020), some image categories were predictive of the next image categories while some were not, and the authors found predictive hippocampal representation in these trials, and inferred that participants made predictions. Another example is that research on schema often uses prediction error to describe when something that is inconsistent with the schema is shown (e.g., when a pot is found on the kitchen floor instead of kitchen counter, (Quent et al., 2022)). An important question is whether these predictions are made ahead of time or if contextual violations are assessed after seeing an outcome. In Chapter 2, in studying the effect of prediction and prediction error on memory, I made a contrast between *making a prediction and comparing what is predicted with the outcome* and *simply responding to something that has high or low probability in a context without making a prediction beforehand*. In chapter 3, which uses the same game paradigm as Chapter 1 and 2, I look at one of the five dimensions in the framework proposed by Ortiz-Tudela et al., (2023) - mnemonic origin - comparing prediction based on general schema of the game vs. episodic memory retrieval.

### **Neural mechanisms of schema-facilitated memory and prediction**

Considering the importance of schema in cognition and memory, numerous neuroscience research has looked at the neural mechanisms of schema (van Buuren et al., 2014; Van Kesteren et al., 2013). This line of research highlights an important role of the medial prefrontal cortex (mPFC) in representing schema and schema-facilitated memory (Bein & Niv, 2025; Gilboa & Marlatte, 2017). Lesion studies with patients with damage to the ventromedial PFC showed deficit in schema reinstatement (Ghosh et al., 2014). Van Kesteren et al., (2013) showed that higher activation of the mPFC during encoding predicts correct memory for schema-consistent

image pairs. The mPFC is a part of the Default Mode Network (DMN), which was found to be active when participants were laying still in the scanner (Seeley et al., 2007), and argued to be involved in internal processing (Menon, 2023). However, this view has been challenged by recent neuroimaging studies, where the DMN shows great involvement in remembering and anticipating temporally-extended naturalistic stimuli like movies (Chen et al., 2017; C. S. Lee et al., 2021) and narratives (Baldassano et al., 2018). In Baldassano et al., (2018), participants were shown movie clips and audio narratives of people going through the script of a “restaurant” (enter restaurant - seated at table - ordering food - food arrives) or “airport” (enter airport - airport security - boarding - on plane). Although the lower level visual and auditory features of the stimuli were very different, the DMN regions represented the stimuli from the same script more similarly. In summary, past research with controlled, experimental paradigms and naturalistic stimuli show the involvement of the DMN, specifically the mPFC, in schema-related processing. Chapter 3 looked at how the DMN is involved in the encoding of move sequences in the board game which sits in the middle of the experimental paradigms and naturalistic stimuli. Chapter 4 looked at how the DMN, and specifically the mPFC, is involved in combining schema (location in a memory palace) with a novel item in a creative, conjunctive, and schema-consistent way.

The other key region of interest of the current research is the medial temporal lobe (MTL). In the SLIMM model (schema-linked interactions between medial prefrontal and medial temporal regions) proposed by van Kesteren et al., (2012), they argue that mPFC detects the congruency between schema and the upcoming input, allowing for rapid neocortical learning and suppressing the MTL. On the other hand, MTL captures novel experiences that cannot be linked to a prior schema, and creates a new memory instance in the cortex, consistent with countless

studies highlighting MTL as a memory system (Scoville & Milner, 1957; Squire & Zola-Morgan, 1991). However, a new perspective has been recently proposed regarding the role of MTL as a more general process that is related to processing relational information, and thus involved in not only memory, but also perception (Bonnen et al., 2021; Ruiz et al., 2020), attention (Aly & Turk-Browne, 2016, 2017), decision making (Bakkour et al., 2019; Biderman et al., 2020), and statistical learning (Schapiro et al., 2012). Some studies have also shown the involvement of the MTL in making predictions about the future (Brown et al., 2016; Brunec & Momennejad, 2022; Sherman & Turk-Browne, 2020). A common criticism of this new perspective is to argue that these predictions are based on episodic memory, and are therefore hippocampally-dependent because they rely on retrieval of a past episode rather than a prediction about a novel episode. In Chapter 3, by asking participants to make predictions about novel and previously seen boards (and therefore allowing them to make predictions based on schema and episodic memory), I test whether the hippocampus is involved in schema-based prediction, episodic memory-based prediction, or both.

### **Overview of the current research**

In Chapter 1, I introduced a new paradigm to study schema and memory with the board game developed by van Opheusden et al., (2023). The game, called four-in-a-row, is an extension of tic-tac-toe, but played on a 4x9 board and the winning condition is to connect four pieces in a row (horizontally, vertically, or diagonally). The game is novel for participants, learnable in a short scale of time, and has an ideal complexity that allows for a large state space yet is still modellable. In the first three chapters, schemas are operationalized as people's knowledge of the strategies of the game and probable moves in different contexts. We first had

participants remember sequences of moves from the game, without any information about the nature of the stimuli, allowing us to measure a baseline of memory pattern without a schema. They then learned the rules of the game, and played games themselves, before again performing a memory task, and this process of building expertise and testing memory was repeated across 6 sessions for each participant. We found that participants' memory improved over time, and they became better at remembering moves that had high probability according to a gameplay model (schema-consistent). With eye-tracking, we showed that participants spontaneously started making predictions, looking at empty squares on the board that had a high probability, and that this predictive behavior was associated with better memory. We found evidence for a model where schema facilitates memory through enabling prediction of the future, and showed the process by which this develops.

In Chapter 2, I continue using this paradigm to look at the influence of prediction on memory. Chapter 1 showed that the paradigm allows us to measure participants' prediction with eye-tracking, and we developed a gameplay model that can output the probability of moves on a given board. By using real-time eye-tracking, we use participants' eye-movement (while looking at the board before a move shows up) to present participants with moves that varied in how much they were accurately predicted and how probable they were under our gameplay model. We were given a unique opportunity to study whether prediction accuracy and probability, two highly correlated properties of a stimulus, which had been treated synonymously (low probability = prediction error) in the literature, are indeed the same process. We showed that prediction accuracy and probability are two separate processes that both positively contribute to subsequent memory, but through different mechanisms.

In Chapter 3, I used fMRI to look at the neural mechanisms of schema, prediction, and memory using the four-in-a-row paradigm. Participants encoded and recalled sequences of moves from the game, and then later in the experiment, were asked to make predictions about the next move on a given board, which can be either from a previously seen sequence or a completely new board. We showed that, consistent with Chapter 2, encoding and retrieving probable vs. predicted moves are associated with distinct neural representations. Additionally, predictions based on schema and episodic memory also showed different activation patterns, especially in the retrosplenial cortex.

In Chapter 4, I train a group of novice participants to use the Method of Loci, and look at the representation of the loci in their memory palace, words by themselves before they were combined with the loci, and how they were combined when the loci were used to remember the words. In this context, schemas are operationalized as the sequence of locations in their memory palace. I show representation of locus and item during encoding the words with Method of Loci. More importantly, I show that, in DMN regions, the locus and item patterns by themselves only account for a small portion of the retrieval representation, it was the conjunctive representation formed during encoding that was mostly strongly reinstated during retrieval. In mPFC, this conjunctive representation also increased over the course of training, and tracked the amount of novel elements participants added to the locus-item pair in their story.

Together, these “kind of naturalistic” paradigms generate important insights about how schemas provide a scaffold for memory, both by enabling complex predictions and by forming conjunctive representation. I also show the distinct mechanisms in responding to and making different types of prediction.



# **Chapter 1: Schema-based predictive eye movements support sequential memory encoding**

*A version of this chapter was published as:*

*Huang, J., Velarde, I., Ma, W. J., & Baldassano, C. (2023). Schema-based predictive eye movements support sequential memory encoding. *Elife*, 12, e82599.*

## 1.1 Introduction

A key benefit of having a memory system is that it enables us to use the past to make predictions about the future, which is adaptive for survival (Cowan et al., 2021). Prediction can be defined as a top-down process in which people generate expectations about what they will experience next, based on their previous experiences and the current context. A particularly important source of prediction are schemas, which are adaptable knowledge structures reflecting generalized information abstracted from multiple episodic experiences (Ghosh & Gilboa, 2014). For example, we may have schemas about the kinds of objects that tend to occur at a beach, the social norms for ordering food at different kinds of restaurants, or the typical stages of a chess game.

Previous research has shown that having a detailed and robust schema yields improvements in memory (Alba & Hasher, 1983). This memory improvement has been attributed in part to processes during recall, since schemas provide cues that can be used to retrieve episodic details that would otherwise be forgotten (R. C. Anderson & Pichert, 1978; Watkins & Gardiner, 1979). Additionally, schemas could play a role during memory encoding, by helping people represent information more meaningfully (Bransford & Johnson, 1972; Chase & Simon, 1973).

One question that has been relatively unexplored in this literature is whether schemas improve episodic memory by enabling sophisticated predictions during encoding. In the past 20 years, there has been a growing recognition of the importance of prediction on how we perceive, understand, and interact with the world (Bar, 2009; Clark, 2013; Friston, 2010; Rao & Ballard, 1999). There is evidence that domain experts (i.e., people with a strong schema in a specific field) automatically engage these predictive processes; for example, basketball experts shown a

clip from a game tended to remember the final positions of the players as being ahead of their actual positions (Gorman et al., 2012). These predictive processes are only possible with a pre-existing schema and might partially explain the large literature showing selective memory improvement for schema-consistent (i.e., predictable) information (J. R. Anderson, 1981) and highly unexpected information (i.e., prediction errors, Bonasia et al., 2018; Quent et al., 2022)).

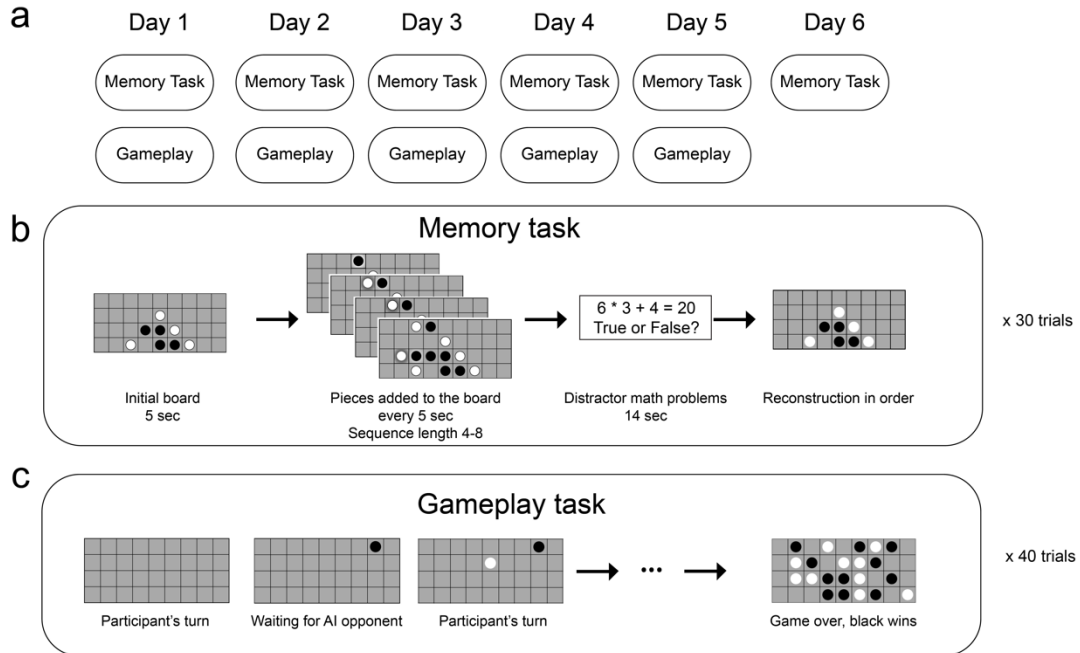
Previous prediction research has typically used two types of approaches. The first approach uses people's pre-existing real-world knowledge, such as the regularity of language structure (Goldstein et al., 2022; Shain et al., 2020) or event sequences in a familiar setting (Baldassano et al., 2018). Due to their complexity, however, it is difficult to build a ground-truth model for these schemas. These types of knowledge are also learned slowly, making it difficult to study their development in a lab setting. The second paradigm, commonly used in memory research, teaches participants simple and novel sequences of discrete stimuli like abstract shapes and pictures (e.g., Schapiro et al., 2012; Sherman & Turk-Browne, 2020). Due to their simplicity, it is possible for participants to learn these sequences in a short period of time and they are easy to model. However, the predictive processes examined in these studies might not engage the same mechanisms as when predictions are based on more complex schemas that generalize to new stimuli.

In this study, we investigate prediction as a potential mechanism for schema-related memory improvement, in a domain that avoids the issues with overly simplistic or poorly defined schemas. Specifically, we have participants remember sequences from a simple board game recently developed (van Opheusden et al., 2023) called 4-in-a-row. In this two-player game, a generalization of tic-tac-toe, players compete to be the first to connect 4 pieces in a row on a 4x9 board. The schema of the game encompasses not just this rule (which is easily learned),

but an understanding of what kinds of move sequences are typical, an emergent property of the game rules that requires playing experience to learn. The game has a complexity far exceeding typical tasks in previous schema and prediction research, yet it is possible to capture near-optimal play with a linear model. The novelty and simplicity of the game ensure that participants start the experiment without a schema but can acquire the schema over several hours of practice, allowing us to tractably study how schema development is related to changes in prediction and memory longitudinally in a lab setting. The fact that possible moves correspond to different spatial locations allows us to use eye movements as an indication of people's predictive processes, as in previous paradigms with spatial actions (e.g., Tal et al., 2021).

Based on findings in previous schema research, we hypothesized that people's memory for move sequences would improve over training sessions, alongside improvements in schema, operationalized as improved gameplay ability. Although previous research has sometimes found novelty-driven memory improvements for schema-inconsistent information (Frank & Kafkas, 2021), studies of expert memory for complex memoranda such as chess boards have shown an advantage for schema-consistent stimuli (e.g. board positions from actual chess games) (Gobet & Waters, 2003). Thus, we hypothesized that the memory improvement resulting from the development of schema in 4-in-a-row should similarly be specific to moves that are schema-consistent. People's schema quality should also be related to their memory performance, such that people with stronger gameplay ability will have better sequence memory. If prediction is indeed a potential mechanism for schema-related improvement, we would expect people's eye movements to become more anticipatory as they gain more experience playing the game and can rely more on internal predictive models. The extent of predictive eye movements should also mediate the relationship between schema quality and memory performance, providing a

mechanism through which schematic knowledge can impact episodic memory. Here, we find support for all these hypotheses.



**Figure 1.1 A schematic of the experimental method. a:** task structure across 6 (non-consecutive) days. **b:** memory task. In each of 30 trials, participants saw an initial board for 5 sec, and then a move was added to the board every 5 sec. After viewing a sequence of 4-8 moves and completing a distractor task, participants were shown the initial board and asked to reconstruct the sequence by placing the pieces on the board. **c:** gameplay task. Participants played 40 games against an adaptive-difficulty AI agent.

## 1.2 Methods

The study used a longitudinal design, in which participants completed 6 sessions over a period of about one to two weeks (Figure 1.1a). The mean interval between sessions was 2.15 days. In sessions 1 to 5, participants completed 2 (practice) + 30 (formal) trials of the memory task, followed by playing 40 games against an AI opponent. The gameplay task was designed to both develop participants' schema and measure their playing strength in each session. On day 6, participants completed 30 trials of the memory task only. Both the memory task and the gameplay were built on Psiturk (Gureckis et al., 2016) and hosted on Heroku

(<https://www.heroku.com/>).

### ***Participants***

The first set of 19 participants (12 female, 6 male, and 2 non-binary) completed the task online on their home computers (and were instructed to maximize their browser windows during the task). The online study participants had a mean age of 20.95 years ( $SD = 2.84$ ). Participants were paid \$70 upon completion of the study. A second set of 16 participants (9 female, 6 male, and 1 declined to answer) performed an identical task in the lab while eye-tracking data was collected. The eye-tracking study participants had a mean age of 21.81 years ( $SD = 3.21$ ). Participants were paid \$100, plus up to \$20 performance-based bonus. For both versions of the study, we recruited participants via online ads and personal contacts. All participants were over 18 years of age with normal or corrected-to-normal vision, and gave informed consent for the study. The experimental protocol was approved by the Institutional Review Board of Columbia University (AAAS0252). One participant from the online study did not complete the last session of the study, and we have included their data for the first five sessions. All the other participants completed all six sessions of the study. Two in-person participants experienced technical issues in one of their sessions, resulting in the loss of data from 1 game for one participant and 15 games for the other participant. The same technical issues during the eye-tracking study resulted in a small number of trials being shown more than once to several participants; we included only data from the first presentation of each trial in the dataset.

### ***Experimental Design***

The memory task (Figure 1.1b) required participants to watch a sequence of moves and then recall the moves from memory. In each trial, participants saw an initial board for 5 sec. Then one move was added to the board every 5 sec. After all moves in the sequence had been

added, participants completed 14 seconds of simple distractor math problems asking them to judge whether an equation is true or false. They had 6 seconds to respond to each question. After the distractor, they were shown the initial board and instructed to reconstruct the sequence in the right locations and in the right order. To account for motor mistakes, they could undo the most recent move they placed.

The first two memory trials on each day were always practice trials that had 4-move sequences, and participants were given feedback on whether a move they just placed was correct or not during retrieval. Participants needed to get all moves correct in the practice trials to proceed to the formal study. The practice trials made sure participants understood the task and followed the instruction. No one was excluded from the study for failing to follow the instructions. Participants then completed 30 trials of the main memory task where they did not receive feedback during the retrieval.

The sequences of moves were generated using an AI agent from van Opheusden et al. (2021) with an relatively weak Elo rating similar to participants' average playing strengths during their first session. We sampled from 180 unique games, each longer than 16 moves and shorter than 36 moves (i.e., the game did not end with a draw). For each session, 30 game segments with lengths ranging from 4 to 8 moves were extracted from unique games. To ensure that meaningful schematic predictions were always possible, the first move of the sequence was always after the fifth move of the game. For example, if a game was 30 moves long and the sequence length was 4, the beginning of the sequence could be anywhere between the 5th move and the 26th move of the game. Of the 30 sequences in each session, 10 sequences (2 of each length) ended with one player winning (i.e., the last move created a four-in-a-row). This single set of 30 sequences for this session was shown to all participants in a randomized order (with a

different set for each session). The whole memory task takes about 50 minutes.

At the beginning of day 1, the participants were told that the stimuli were “circles appearing on a grid”. After the memory tasks, they were told that the stimuli were actually drawn from a game that they were about to play. They were then shown the rules of the game and played 40 games against an AI agent (Figure 1.1c), which takes about 40 minutes. We used a staircasing procedure, such that the agent became stronger if the participant won and weaker if the participant lost. We used Elo rating to measure participants’ playing strength (Elo, 1978). The ratings were computed using BayesElo (Hunter, 2004) to measure participants’ playing strength in each session based on their performance against the AI agents. After they finished playing the games, they were asked whether they had guessed the stimuli in the memory tasks were from a game, and if so, whether they guessed the rules of the game. We did not include a control group that completed six memory sessions without being told the rules of the game or playing against the AI agent. Ensuring that participants never develop a game schema would be difficult, since even without explicit instruction they could still learn the schema through implicit statistical learning or by guessing the (simple) rules of the game. Due to these challenges, we instead used the first-session performance of each participant as our no-schema control.

From Day 2 onwards, they were reminded of the rules of the 4-in-a-row game at the beginning of the memory task. On Day 6, after completing the memory task, participants completed a questionnaire asking them what strategies they used for the memory and whether their strategies changed over the course of the training.

### ***Gameplay model***

The original model in van Opheusden et al., (2021) used a tree search model. However, obtaining accurate move probabilities from this model would require extensive sampling.

Instead, we used a feature-based myopic model of gameplay. We first defined features relevant for gameplay, which represent the relationship between potential next move and the current board state:

1. Distance of the move from the horizontal center
2. Distance of the move from the vertical center
3. How many 4-in-a-rows the move forms
4. How many 3-in-a-rows the move forms. There are 3 sub-categories based on the type of 3-in-a-row formed: connected, disconnected, horizontally connected that were not blocked on either side (force a win after opponent's move, so it might have a higher value)
5. How many 2-in-a-rows the move forms. There are 2 sub-categories based on the type of 2-in-a-row formed: connected, disconnected
6. How many opponent's 3-in-a-rows the move blocks
7. How many opponent's 2-in-a-rows the move blocks. There are 2 sub-categories based on the type of 2-in-a-row blocked: horizontally connected 2-in-a-row that was not blocked in either direction (if not blocked, opponent can force a win, so it might have a higher value), or other situations.

We then used a very strong AI agent to generate 800 games. This agent has an Elo of 365, a level similar to the best players in our study, so its move decisions can be considered an approximation of the optimal strategy. To fit the model, for each board state in the 800 games, we represented every possible move  $x_i$  as a feature vector  $F$  based on the features described above. Each move was assigned a value  $V$  that is a weighted combination of the features it forms, with weights defined by a vector  $w$ :

$$V(x_i) = \mathbf{w} \cdot \mathbf{F}(x_i)$$

We then applied a softmax function to get the probability distribution over possible moves,  $p(x_i) = e^{V(x_i)} / \sum_i e^{V(x_i)}$ . We used Pytorch (Paszke et al., 2019) backpropagation to learn the feature weights  $w$  that minimized the cross-entropy loss of the moves actually made by the agent. We trained eight models using the data from 100 of the 800 games for each model, and averaged the weight vectors across models to obtain a final weight vector  $w$ . Our analyses test the hypothesis that, through repeated gameplay, the schemas of individual players should converge to the move probabilities from this optimal model, and therefore these optimal move probabilities should come to impact subsequent memory (and memory errors). An alternative possibility could be that individuals develop schemas that are qualitatively distinct from the optimal schema but still effective in amateur gameplay. We did not find empirical evidence for this possibility in our dataset; during the gameplay sessions, participants tended to play moves with relatively high probability under the optimal model (average move probability of 0.213 across participants, compared to a random-move baseline of 0.040) and we observed a strong correlation between Elo and this move probability measure ( $r = .40$ ,  $p < .001$ ). Future work (with more extensive gameplay sessions) could attempt to model unique feature weights for each specific individual

We evaluated each move in the stimulus set with this model to determine the probability of each move under a near-optimal strategy. Note that the stimuli were generated by a non-optimal AI agent, and therefore move quality varies over a wide range; the moves participants observe are often not the optimal move for a given board configuration. We also used this model to measure the probability of the move participants recalled during retrieval when they made a mistake.

## *Eye-tracking*

For the in-lab subset of participants, eye-tracking data was collected during the memory task on each day. The design was otherwise identical to the online experiment. Participants were seated 100 centimeters in front of a monitor, and placed their heads in a chin rest 45 centimeters away from the eye-tracker. They were instructed to remain as still as possible while the eye-tracker was running, and were told that they could take breaks during the experiment in between trials. Before beginning the experiment and when the participants returned from their breaks, the eye-tracker calibration and subsequent validation was done using a nine-point grid. We recorded binocular eye movement using EyeLink 1000 plus at 1000 Hz recording frequency. Light levels remained constant for the duration of the 50-min memory portion of the study. The stimuli were displayed on a 24-inch LED monitor, with a resolution of 1920 by 1080 pixels and a refresh rate of 60 Hz. The outputted EDF files were converted to asc files and parsed with PyGaze (Dalmaijer et al., 2014)

Fixation maps were created for each 5-second period during which an initial board was shown or a move was shown. To handle uncertainty in assigning gaze to squares, we performed a soft assignment to board locations based on distance. For a fixation at position  $x_F$  with duration  $t_F$  the square with center coordinate  $x_i$  was assigned a fixation weight of

$$t_F \cdot \frac{e^{-\|x_F - x_i\|_2 / 25}}{\sum_j e^{-\|x_F - x_j\|_2 / 25}}$$

Here, distance is in the unit of pixels. Since the length of the square is 136 pixels, the smoothing temperature of 25 is approximately 1/5 the length of a square, and therefore is only relevant for fixations close to square boundaries. The weights for all fixations during the 5-

second window were summed to obtain a final map of fixation weights for all board squares.

### ***Eye movement regression model.***

We modeled fixation maps as a linear combination of 6 potential strategies that could be used during memory encoding. For each board, the regressors, described below, are length-36 vectors that correspond to the 36 tiles in the game.

1. *The most recent move*: has a value of 1 in the square that corresponds to the most recent move, and 0s elsewhere (all 0s at the initial board). This regressor captures the extent to which participants are looking at the move that just appeared on the board.
2. *Pieces related to the most recent move*: has a value of 1 in occupied squares that are related to the most recent move (within 3 tiles from the most recent move in any direction), and 0s elsewhere. This regressor captures the behavior of looking at how a move relates to previously-placed pieces (e.g., to detect whether it adds onto a line of same-color pieces).
3. *Anticipation of the upcoming move (prediction)*: for the unoccupied squares, is equal to the probabilities of that square being the location of the next move, as calculated by the near-optimal gameplay model. For occupied squares, has a constant value equal to the mean value of the unoccupied squares. This regressor measures how well participants' eye movements predict likely positions for the next move.
4. *Piece relevance to the next move*: for each occupied square, is the sum of the next-move probabilities for all empty squares that are related to it (within three squares). For unoccupied squares, its value is equal to the mean value of the occupied squares. This regressor reflects the tendency to focus on squares that are likely to be related to the next

move.

5. *Observed previous moves*: has a value of 1 for the moves seen earlier in the sequence (excluding the most recent move), 0s elsewhere. This reflects a strategy of reviewing previous moves during memory encoding.

6. *Occupied tiles*: has values of 1s on occupied tiles and 0s on unoccupied tiles. This measures the preference for looking at occupied squares versus empty squares.

Each regressor and the eye movement fixation maps were z-scored within each board.

We concatenated each of these across all the boards that each participant saw during a session and ran a multiple linear regression (with the fixation maps as the outcome variable, and the six regressors as the predictor variables). The coefficient for each regressor reflects the extent to which this participant used this eye movement strategy when encoding boards in this session. We also ran this regression separately for each individual board (for each move shown during encoding) to obtain trial-level estimates of eye movement strategy.

### ***Regression Models***

For each model, we started with the most complex frequentist model, including a subject slope for all of the predictors. In case they did not converge, we used simpler models with either fewer random slopes or just a random intercept. If this happens, we also checked whether the effect hold with a more complex model using a Bayesian model. The reported effects with more simpler frequentist models have been shown to hold with a more complex Bayesian model. If a different result was obtained, we report the results of the more complex Bayesian model. For frequentist models, we use lme4 package (Bates et al., 2015). For the Bayesian models, we used the default settings of rstanarm package (Goodrich et al., 2022).

### ***Mediation analysis***

The total effect was calculated by running a linear regression predicting recall accuracy from Elo. Next, we calculated the effect of Elo on prediction coefficient, and the effect of Elo and prediction coefficient on recall accuracy. The significance of the mediation was computed with the package *mediation* (Tingley et al., 2014) that used a bootstrapping procedure. Standardized indirect effects were computed for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles.

### ***Estimating trial-level prediction confidence and surprise***

We constructed two trial-level measures of fixation statistics which allowed us to describe eye movement strategies at a finer scale and in a model-free way (without assuming that participants were making optimal predictions according to our schema model). The first is prediction confidence, which measured the extent to which a participant spent time focused on specific empty squares. High values of confidence indicate that a participant spent a large fraction of the trial looking at only a small number of empty squares, indicating a strong prediction about the upcoming move. We compute this as the expected information gain between a uniform distribution over all empty squares and the fixation distribution. Given the fixation time  $T(x_i)$  for each square  $x_i$ , we define  $P(\text{empty})$  as the fraction of the 5-second window spent fixating on empty squares, and  $P(x_i) = T(x_i)/P(\text{empty})$  as the normalized fixation distribution over empty squares. The information gain from a fixation is 0 for fixations on occupied squares, and for fixations on empty squares reflects the entropy difference between a uniform distribution and the fixation distribution. Therefore, we define:

$$\text{Prediction confidence} = P(\text{empty}) \cdot (\log(N_{\text{empty}}) - \sum_i^N P(x_i) \log P(x_i))$$

The second is prediction surprise, indicating the extent to which a participant failed to look at the location where the next move was going to appear. High values of surprise indicate that the move appeared in a location that the participant spent very little time looking. This is defined as the negative log of the percentage of time participants looked at the correct upcoming move position ( $x_{next}$ )

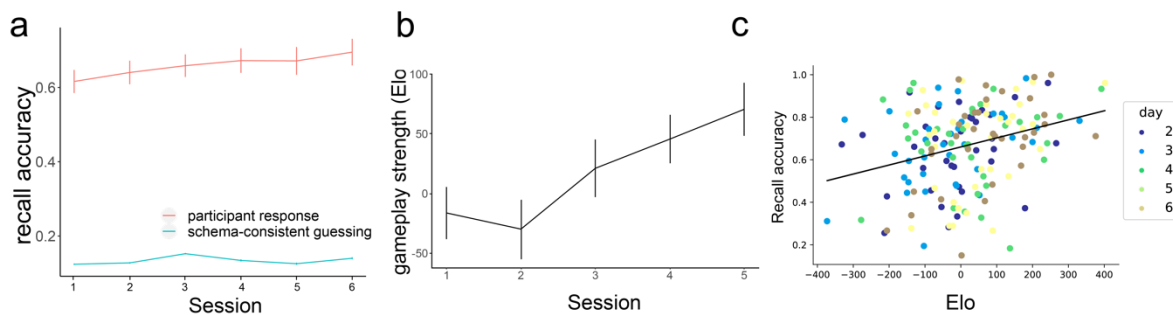
$$\text{Prediction surprise} = -\log(P(x_{next}))$$

## 1.3 Results

### 1.3.1 Memory and gameplay improvement

Over 6 sessions (each separated by 2.15 days on average), participants performed two separate tasks (Figure 1.1). In the memory task, participants were presented with a novel gameplay sequence and asked to remember it. After a distractor task, they were shown the first board of the sequence and they were given unlimited amount of time to reconstruct the rest of the sequence from memory. In the first session, participants were not told that these stimuli came from a game, and were only instructed to remember circles appearing on a grid. The first session therefore provided a no-schema baseline, since participants could not use a game model to make predictions about upcoming moves. In a post-task questionnaire, we confirmed that in session 1 most participants did not suspect that the stimuli were from a game or guess the rules of the game (14 out of 19 in the online study, and 13 out of 16 in the in-person study). In a separate game play task (occurring after the memory task on all but the last day), they were provided with the rules of the game, and played the 4-in-a-row game against an AI opponent, staircased to match the skill level of the player. Participants' recall accuracy (combined across the online and in-lab data), calculated as the percent of moves they recalled at the right location in the right order, improved across training sessions, improving from 61.6% ( $SD = 18.1\%$ ) in session 1 to 69.5% ( $SD = 20.6\%$ ) in session 6 (Figure 1.2a). At the same time, participants' playing strength, measured with an Elo rating (Elo, 1978), increased (Figure 1.2b). Elo ratings are computed based on how often people win against opponents of varying skill levels, and we use Elo as a measure of schema quality – the better a player is, the better knowledge they have about the move probabilities during near-optimal gameplay. We fit mixed-effect models with a fixed effect of session and a random slope for participant to predict recall accuracy and Elo. The fixed effect of session was significant in both

models ((for memory,  $\beta = 0.015$ ,  $t = 3.215$ ,  $p = .003$ ; for Elo,  $\beta = 24.92$ ,  $t = 5.102$ ,  $p < .001$ ), demonstrating improvement in both recall accuracy and gameplay over time. We also found that Elo rating in a session was correlated with memory performance in the following session, Pearson  $r = .298$ ,  $p < .001$  (Figure 1.2c), showing that having a better schema for the game was associated with better episodic memory for move sequences. To understand whether this relationship was present within individual participants, we fit a linear mixed effect model to predict memory performance from Elo with per-participant intercepts and slopes as random effects. We found that the relationship between Elo and memory accuracy was not significant in this model ( $\beta = 0.012$ ,  $t = 1.196$ ,  $p = .242$ ), suggesting that this effect was primarily driven by individual differences (people with better schema tend to have better memory) rather than across-session improvements in Elo.



**Figure 1.2. Participants' performance in memory and gameplay.** **a.** On average, Participants become better at remembering sequences over the course of the training (red line). In each session, memory accuracy is much higher than the performance that would be achieved if people were simply guessing according to the optimal gameplay model. **b.** On average, participants become better at playing the game across sessions. Error bars represent the standard error of the mean. **c.** There is a positive correlation between people's playing strength and their recall accuracy (each dot corresponds to one session of one participant).

Out of the 30 sequences shown to participants in each session, ten of them ended with one player successfully getting four pieces in a row (more details in Method). After learning the rules

of the game, this should be a salient event for our participants and could lead to changes in memory performance. Indeed, we found that in sessions 2-6, memory for sequences that ended in a win state was significantly better ( $t = 6.67, p < .001$ ). We did not observe this pattern in the first session (before participants were taught the game rules), and actually found a marginally significant effect in the opposite direction, with worse memory for winning sequences ( $t = -2.01, p = .052$ ). This result provides additional evidence that schema-related features of a sequence might play an important role in memory performance.

### **1.3.2 Modeling schema-related memory improvement**

To investigate how schema-consistency is related to memory for individual moves and how this relationship evolves, we trained a model of the schema for the game on moves played by a near-optimal AI agent. The model identifies the features each move forms (such as 3-in-a-row, a line of three pieces of the same color) and assigns a value to each feature based on the training data (described in more detail in the Method section). Given a current board position, the model outputs a probability distribution over the next move that would likely be played by a near-optimal player (Figure 1.3a). Using this model, we can measure the extent to which each move shown to participants is likely (schema-consistent) vs unlikely (schema-inconsistent), and how this is linked to recall accuracy. Here, schema-consistency is used as an objective, subject-independent measure of how good a move is. Each subject will exhibit different degrees of alignment to this “ground-truth” schema.

Separately for each session, we ran a mixed-effects logistic regression with subsequent memory (right or wrong) as the outcome variable, with the probability of the move as the fixed effect, and with a participant random intercept. As shown in Figure 1.3b, there was initially no relationship between the probability of a move and the probability that the move will be

remembered in session 1 ( $\beta = -0.031$ ,  $z = -1.813$ ,  $p = .07$ ), but in sessions 2 through 6, schema-consistent moves were more likely to be remembered (all  $p < .001$ ). This effect emerged over the first four sessions of learning (Figure 1.3c), and then dropped slightly in session 5 and 6.

To see whether this increase in the relationship between move probability and recall accuracy over time is significant, we aggregated the data from all the sessions and ran a mixed-effects logistic regression with subsequent memory as the outcome variable, with the probability of the move, the session the move was in, and their interactions as fixed effects, and with a participant random slope of session and move probability. We found a main effect of session on memory ( $\beta = 0.164$ ,  $z = 5.763$ ,  $p < .001$ ) but no main effect of move probability ( $\beta = -0.0046$ ,  $z = -0.249$ ,  $p = .804$ ). However, there was a significant interaction between move probability and session on memory ( $\beta = 0.035$ ,  $z = 7.561$ ,  $p < .001$ ), indicating that participants became better at remembering schema-consistent moves over the time.

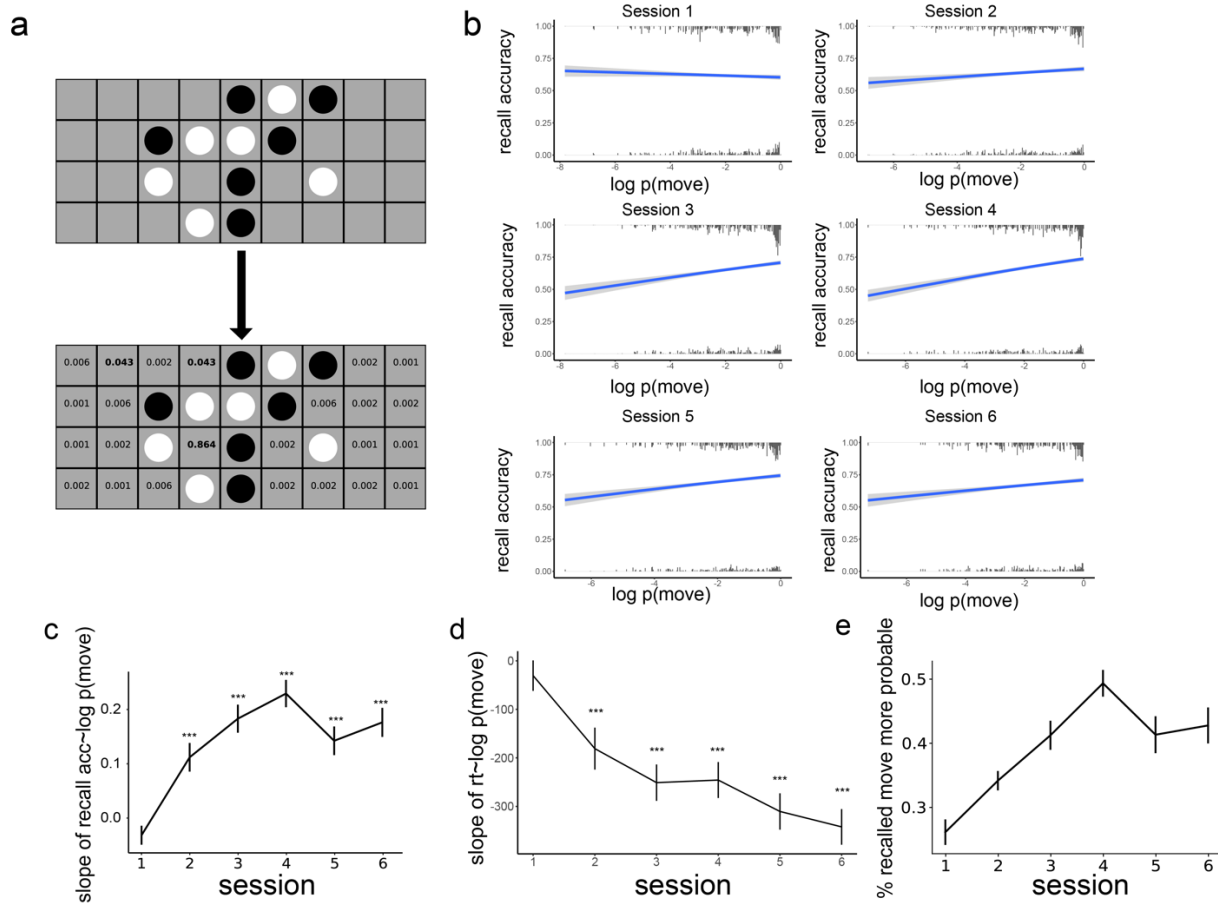
In addition to recall accuracy, we looked at reaction time for placing moves at recall. We only looked at moves that were correctly recalled in each session and removed outliers with reaction times longer than 30 sec (0.6% of all the correct moves were removed this way). Similar to accuracy, we found that reaction time during retrieval was initially not related to the probability of the move ( $\beta = -30.5$ ,  $t = 0.975$ ,  $p = .329$ ) but consistently faster for more schema-consistent moves in session 2 through 6 (all  $p < .001$ , figure 1.3d). Running a mixed-effects model as for the accuracy (using linear rather than logistic regression), we found a main effect of session on reaction time ( $\beta = -156.3$ ,  $t = -2.39$ ,  $p = .017$ ), such that participants were faster in later training sessions. There was no main effect of move probability ( $\beta = -46.1$ ,  $t = -0.440$ ,  $p = .660$ ) but a significant interaction between move probability and session on reaction time ( $\beta = -$

71.06,  $t = -2.593$ ,  $p = .010$ ), with faster responses when remembering more schema-consistent moves.

If participants are using a schema as part of their recall process, we would expect a bias toward schema-consistent moves when participants make mistakes during sequence reconstruction. To test whether this is the case, we looked at the mistakes participants made in each sequence reconstruction, measuring the fraction of the time that their answer was more schema-consistent than the correct answer, indicating a bias toward schema-consistent moves during recall. We only considered the first mistake, since after that mistake, the encoding and retrieval boards are different, and therefore the likely moves are no longer comparable. As can be seen in figure 1.3e, the proportion of schema-consistent mistakes increased over the first 4 training sessions. After that, there is a drop similar to the drop in accuracy in session 5 and 6. We ran a mixed-effects model with a participant random slope, the proportion of schema-consistent mistakes as the outcome variable, and session as a fixed-effect predictor, and found a significant effect of session ( $\beta = 0.032$ ,  $t = 5.413$   $p < .001$ ).

Although an overall increase across sessions was observed in both schema-consistent mistakes and the relationship between recall accuracy and move probability, these effects were significantly weaker at session 5 (for schema-consistent mistakes,  $t = -2.51$ ,  $p = .02$ ; for the relationship between recall accuracy and move probability,  $t = -3.24$ ,  $p = .001$ ), and numerically weaker at session 6 (for schema consistent mistakes,  $t = -1.70$ ,  $p = 0.099$ ; for the relationship between recall accuracy and move probability,  $t = -1.93$ ,  $p = .053$ ). One potential explanation for this drop is that people developed new strategies in later sessions to remember schema-inconsistent moves. As can be seen in figure 1.3b, the modeled recall accuracy for highly schema-inconsistent moves (low  $x$  values) increased from session 4 to 5 while the recall

accuracy for schema-consistent moves remained relatively stable. This is consistent with the highest overall accuracy occurred in session 6 (Figure 1.2a), suggesting that this decrease in slope does not come at a cost for memory accuracy. Note that reaction time did not show this weakening relationship with move probability in session 5 and 6, as its slope continues to decrease.

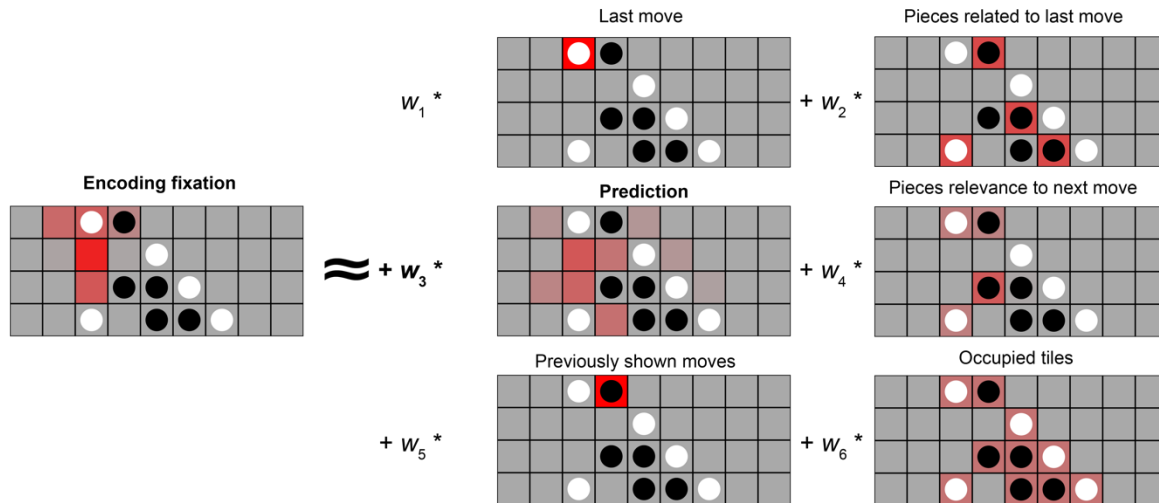


**Figure 1.3. The effect of schema consistency on memory and its development across training sessions.** **a:** an example of the model's evaluation of a board where the next player is black. Based on features that would be generated by each possible next move (e.g., creating 3-in-a-row), the model generates a probability distribution over potential next moves. We apply this model to the stimuli in the memory task to estimate a probability for each move. **b:** the effect of schema consistency of a move on the recall accuracy for the move. The x-axis is the log probability of a move, which reflects schema consistency. The y-axis is the probability that a participant will remember the move. In session 1, there is no relationship between probability of moves and subsequent memory. In session 2-6, people are more likely to remember schema-consistent moves. The histogram at the top and bottom of the figure is the frequency of moves

with a certain log probabilities that are remembered and forgotten, respectively. **c:** relationship between move probability and recall accuracy over the 6 sessions. (\*\*\*) denotes  $p < .001$ ). **d:** relationship between move probability and reaction time at recall for correctly remembered moves. (\*\*\*) denotes  $p < .001$ ). **e:** the proportion of times the first mistake in a sequence is a more probable move than the actual move that was observed. Error bars represent standard error of the mean.

The results demonstrate that the memory benefit from gameplay training is driven in part by enhanced memory for schema-consistent moves. A possible alternative explanation of this effect is that, rather than using generalized schematic knowledge, participants are in fact using episodic memories of specific move sequences that they have seen in past gameplay sessions. To test this possibility, we examined all the boards that participants saw during gameplay and the moves played on these boards, and found repetition of exact move sequences to be very rare; out of over 1000 moves across 6 sessions there were on average 1.0 repeats ( $SD = 1.67$ ) that participants experienced during gameplay and were later part of a memory sequence. Thus, mere repetition of schema-consistent moves could not be driving the observed differences in memory. Another potential explanation could be a change in participants' strategy. In the beginning, participants can only use their episodic memory, but over time participants could start to use a purely schema-based strategy in which they simply place moves as if they are playing the game from this board position. We simulated this strategy by drawing moves probabilistically from the gameplay (schema) model for each initial board that participants saw during the memory task in each session. We found that purely schema-based guessing can only achieve an average accuracy of 13.39% across the 6 sessions (Figure 1.2a), which is much lower than participants' average recall accuracy of 65.91%. This low accuracy is due to the fact that, for most of the sequences, there are multiple plausible moves that could be played and it is therefore difficult to guess the

sequence without episodic knowledge. Therefore, the improvement in performance after learning the schema is unlikely due to mere guessing of the boards.



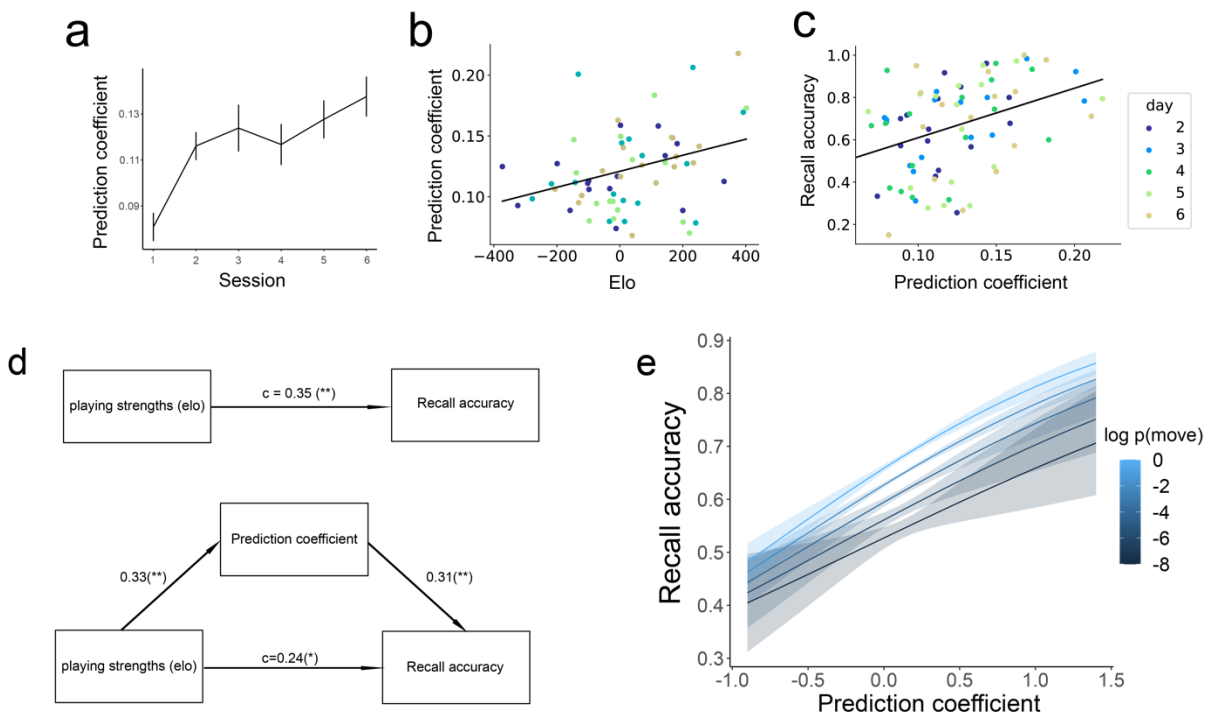
**Figure 1.4. Using eye movements to reveal encoding strategies.** left: a participant’s fixation heatmaps over a 5 sec encoding period. right: the 6 regressors that we consider as potential encoding strategies. 1: Participants could look at the last (most recent) move, which is what they need to remember. 2: Participants could look at occupied tiles that might be relevant to the most recent move, to try to see what features the move forms. 3: Participants could be anticipating the upcoming move, meaning they will look at the empty squares on the board that are likely to be the next move. 4: Participants could also look at current pieces that are relevant for predicting the next move, i.e., pieces related to empty squares that are likely to be the next moves. 5: Participants could be looking at moves that previously appeared, in order to rehearse the observed move sequence. 6: Participants might have an overall tendency to look at occupied or unoccupied tiles.

### 1.3.3 Eye movements became more predictive over training

A potential driver of memory improvements could be a change in encoding strategy. In particular, schemas allow people to make online predictions, which allows additional encoding time if moves are successfully predicted or generates prediction errors otherwise, both of which have been shown to be beneficial for memory (e.g., Quent et al., 2021). We used eye-tracking data to understand how participants’ encoding strategies were related to schemas and later memory.

We modeled fixations as a linear combination of six different possible strategies (Figure 1.4), including a “prediction” strategy (regressor 3) in which fixation durations are related to the model-derived probabilities for the next move. The coefficient for each fitted regressor reflects the extent to which the strategy is used during encoding.

To look at whether the eye movements become more predictive over time, we first ran a linear mixed-effects model with prediction coefficient as the outcome variable, session as the predictor variable, and a participant random slope, to see if eye movements became more predictive over time. There was a significant fixed effect of session,  $\beta = 0.009$ ,  $t = 4.261$ ,  $p < .001$ , providing evidence that people’s eye movements became more predictive over training (Figure 1.5a).



**Figure 1.5 The relationship between prediction, playing strengths, and recall accuracy.** **a:** the extent to which eye movements on empty squares align with an optimal gameplay model (prediction coefficient; see Figure 1.4) increases over training sessions. **b:** correlation between Elo (playing strength) and prediction coefficient. **c:** correlation between prediction coefficient and recall accuracy. **d:** mediation analysis: the effect of playing strength on recall accuracy is

mediated by the amount of prediction participants made during encoding. The values on the arrows represent regression coefficients for standardized measures of Elo, prediction coefficient, and recall accuracy (\* denotes  $p < .05$ , \*\* denotes  $p < .01$ ). **e**: move-level analysis on the effect of schema consistency on memory, predicting whether a move will be remembered based on the probability of the move (brighter colors are more schema-consistent) and prediction coefficient in previous move. Error bars represent standard error of the mean.

### 1.3.4 Better performance in better players is mediated by more schema-based predictions

We next sought to test whether this increase in predictive eye movements could serve as a mechanism through which expertise improves memory performance. We found that, on a session-by-session basis, more sophisticated gameplay (higher Elo) was associated with more prediction in the next memory session (Pearson  $r = .33$ ,  $p = .002$ ), demonstrating that better players predict more during encoding (Figure 1.5b). Memory sessions in which a player exhibited high levels of predictive eye movements also showed better reconstruction accuracy (Pearson  $r = .40$ ,  $p < .001$ ), despite there being no explicit demand to generate predictions during encoding (Figure 1.5c). A mediation analysis was performed to assess the mediating role of the prediction coefficient on the link between Elo and recall accuracy (Figure 1.5d). The total effect of Elo on recall accuracy was significant ( $\beta = 0.35$ ,  $t = 3.26$ ,  $p = .002$ ). The effect of Elo on prediction coefficient was also significant, ( $\beta = 0.33$ ,  $t = 3.13$ ,  $p = .002$ ). The bootstrapped indirect effect was 0.10, 95%  $CI = [0.03, 0.19]$ ,  $p = .006$ , suggesting that improved memory for better players is partially mediated by improved predictions during encoding. Significant mediation was not found for any of the other regressors (all  $p > 0.3$ , except previously shown moves,  $p = .056$ ), suggesting that better prediction is uniquely important for driving the memory benefit from better gameplay. With the inclusion of prediction coefficient, the impact of Elo on recall accuracy was reduced but remained significant ( $\beta = 0.24$ ,  $t = 2.26$ ,  $p = .03$ ), indicating that expertise also improves memory accuracy through additional mechanisms (at encoding and/or retrieval). We again tested whether these effects were present at the individual-participant level by running a linear mixed effect model with participant

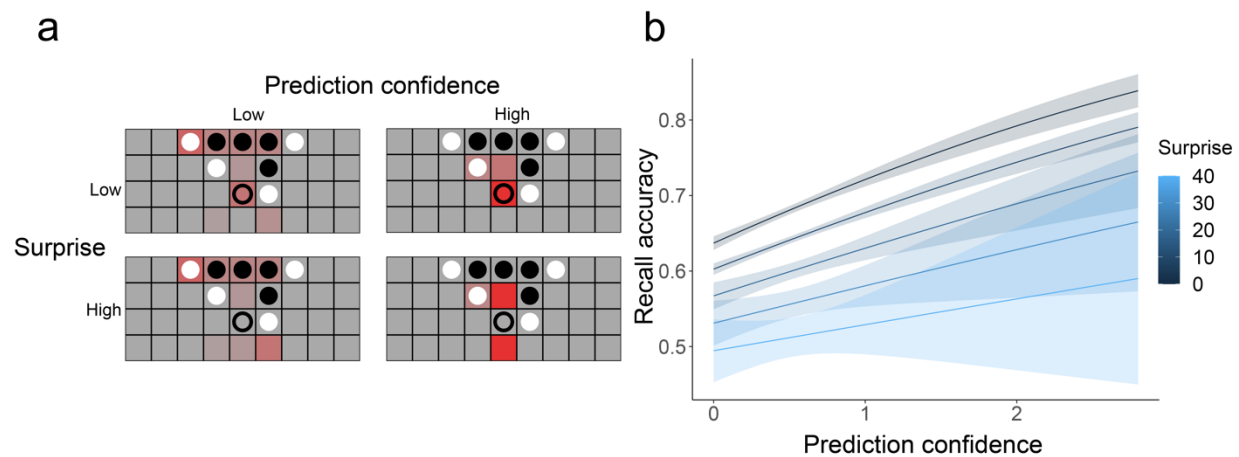
random intercepts and slopes. Due to model convergence issues, we employed a Bayesian version of this model to regularize the coefficient estimates (see Methods). We found that the 95% credibility interval for the fixed effect of Elo on prediction coefficient overlapped with 0 ( $\beta = 0.007$ , 95% CI = [-0.003, 0.016]). The 95% credibility interval for the fixed effect of prediction coefficient on recall accuracy also overlapped with 0 ( $\beta = 0.164$ , 95% CI = [-0.54, 0.87]). Together with previous findings, these results suggest that the observed mediation effect is primarily driven by individual differences rather than within-subject development over time.

The above analysis demonstrated that making more schema-consistent predictions is correlated with better memory at the level of experimental sessions. We next constructed a logistic regression model to predict subsequent memory at the scale of individual moves, as a function of the move's probability and the prediction coefficient during the five-second window before the move appeared with a subject random slope for prediction coefficient (figure 5e). There was a main effect of log probability of the move ( $\beta = 0.083$ ,  $z = 5.882$ ,  $p < .001$ ), consistent with our session-level results demonstrating a memory benefit for more likely moves. We also found a significant effect of prediction coefficient ( $\beta = 0.508$ ,  $z = 3.187$ ,  $p = .001$ ). There is no interaction between the move probability and prediction coefficient ( $\beta = 0.024$ ,  $z = 0.395$ ,  $p = .693$ ). This result shows that making schema-consistent predictions before a move appears is correlated with better memory for the move, independent of the schema-consistency of the move.

### **1.3.5 Model-free measures of prediction confidence and surprise**

In addition to the prediction coefficient, which measured the extent to which eye movements are consistent with next-move predictions from our gameplay model, we considered two model-free aspects of eye movements on empty squares (Figure 1.6a). The first is prediction *confidence*, measuring the extent to which a participant was attending to a specific empty square

(vs. looking evenly at many empty squares or at occupied squares). We use this as a measure of whether participants have a specific expectation about the next move position. The second measure is *surprise*, which is the negative log of the proportion of fixation time spent looking at the correct next move position. A high level of surprise indicates that a participant spent very little time looking at the square where the next move ended up appearing, which indicates low prediction accuracy. We then conducted a move-level logistic regression similar to the one described above, to predict whether a move will be remembered or not based on both prediction confidence and surprise while looking at the previous board (Figure 1.6b), with subject random slope for both prediction confidence and surprise. We found main effects of both factors, indicating that a) making confident predictions before a move is related to better memory for that move ( $\beta = 0.375, z = 4.372, p < .001$ ) and b) unexpected (surprising) moves are more poorly remembered ( $\beta = -0.015, z = -2.448, p = .014$ ).



**Figure 1.6. Model-free measure of prediction and their relationship with memory.** a: example fixation heatmaps (in red) of high and low prediction confidence and surprise. Confidence measures the extent to which fixations were focused on a small number of unoccupied squares, while surprise measures how well these fixations aligned with the actual next move (indicated with an empty circle). b: recall accuracy was best when prediction confidence was high and surprise was low. Error bars represent standard error of the mean.

## 1.4 Discussion

The main goal of the current study was to look at how schema-based predictive processes support sequential memory encoding. Participants learned a novel schema (likely move sequences and board configurations in 4-in-a-row) with much higher complexity than most artificial schemas learned in the lab. Developing a schema for moves in the game requires a significant amount of playing experience (reflected in the slow improvement in Elo ratings we observed across sessions), but can still be accomplished within a single experiment, unlike games like chess. We observed schema-related effects on sequence memory, and found that these effects emerge over time as the schema develops. We also investigated the role predictions during encoding played in the relationship between schema and memory, finding that making schema-consistent predictions, making confident predictions, and making accurate predictions were all related to successful recall at the level of individual moves.

### 1.4.1 Schema-related memory effects and how they develop

Our behavioral findings showed that previously-studied schema effects from previous studies extend to this more complex experimental paradigm, and demonstrated how these effects developed over time. First, memory gradually improved as people's schemas gradually improved. People with a better schema had better memory, consistent with the literature on how expertise influences memory (Chase & Simon, 1973; Gobet & Waters, 2003). Second, the memory benefit over the training session was driven by improved recall for schema-consistent information, consistent with schema literature (e.g., Anderson, 1981). Previous work has also shown that in some situations, schema-*inconsistent* information is better remembered (Frank & Kafkas, 2021), which we did not observe in this study. All the sequences we used were generated by a gameplay agent, so highly schema-inconsistent moves were rare, and none of our stimuli

exhibited the extreme kind of “contextual novelty” (Ranganath & Rainer, 2003; Stark et al., 2018) that often drive these effects; for example, we never showed a move that broke the rules of the game or played an expected sound. Future work could test whether highly schema-inconsistent items exhibit memory benefits equal to or greater than highly schema-consistent items. Third, participants’ mistakes became more schema-consistent across sessions. These errors could reflect false memories for schema-consistent sequences that never occurred, similar to false memories generated by naturalistic videos or stories conforming to schematic scripts (Lampinen et al., 2000; Neuschatz et al., 2002). Alternatively, participants may be guessing schema-consistent moves when they failed to encode or retrieve the sequence; in future work, confidence responses could be used to distinguish between these possibilities. Note that some of the overall accuracy increase in the memory task across sessions could be attributable to a practice effect, but general practice-related improvement does not explain the schema-related effects that we observed for individual trials.

Interestingly, the relationship between recall accuracy and memory is attenuated in the last 2 sessions, driven by improved memory for schema-inconsistent moves. Expertise effects that partially extend to non-schematic situations have been previously observed; although chess experts’ memory advantage was primarily driven by very high accuracy for schema-consistent (legitimate) boards in Gobet & Waters (2003), experts also exhibited memory for random boards that was better than that of weaker players. One possibility is that experts may be able to devote fewer attentional resources to schematic moves, freeing up additional resources to handle more unusual moves. Participants in our study may have also developed an effortful strategy to remember schema-inconsistent moves (such as switching to a shape-based encoding for these moves), consistent with the longer reaction times observed for low-probability moves in later

sessions.

There are other potential mechanisms for the schema-related memory improvement that were not explored in the current study. For example, the ways in which moves are represented in the brain may shift with schema learning, moving from a purely spatial system (e.g., remembering that a move occurred at row 3, column 2) to a more feature-based representation (e.g., remembered a move as “blocking opponent’s 3-in-a-row”), which makes remembering easier and more robust. Future neuroimaging work can use techniques such as representational similarity analysis (Kriegeskorte et al., 2008) to model the representational geometry of remembered board positions and track how it changes over sessions.

#### **1.4.2 The development of predictions and their influences on memory**

The eye-tracking study showed that participants spontaneously exhibited predictive eye movements consistent with the game schema, without any explicit prediction demands. This is consistent with earlier work showing automatic predictive representations in basketball experts (Didierjean & Marmèche, 2005). We also found that schema-based predictions were associated with improved memory (at the level of participants and the level of individual moves), and that these encoding-time predictions provide a mechanism for the impact of schema quality on memory performance. There are several reasons why within-participant increases in Elo across sessions were not predictive of improvements in prediction and memory. First, our estimate of Elo during each memory task was based on gameplay performance on the prior day, which may not be a precise measure of the participant’s current schema. Second, the biggest increase in prediction occurred between sessions one and two, but the session one memory-task data cannot be included in this prediction model since participants have not yet played the game (and we therefore cannot estimate their Elo). Third, we may simply be underpowered to detect within-

participant effects, since we only collected data from six sessions for each subject. Despite the lack of within-subject effect, the session-level analysis provided the novel insight that the memory advantage for more-expert players may be driven by superior predictive processes during encoding.

Despite the numerous studies looking at how item novelty impacts memory (A. Greve et al., 2017; Quent et al., 2021; van Kesteren et al., 2012), relatively few studies have looked at the impact of temporal schematic prediction on memory for upcoming items. In a recent study, Sherman and Turk-Browne (2020) showed that making predictions does not benefit memory for the upcoming item and impairs the encoding of the current item, which is inconsistent with our findings. One possibility is that their study (in which participants implicitly learned arbitrary statistical dependencies between image categories presented in a continuous stream) engages a predictive process that relies on the hippocampus (Schapiro et al., 2017), creating interference between prediction and encoding processes. On the other hand, the explicit schematic predictions in our study that are based on rules and progression towards goal states might rely on a different network such as the medial prefrontal cortex (Bonasia et al., 2018; Robin & Moscovitch, 2017; van Kesteren et al., 2012).

Our eye movement data suggest that making high-confidence predictions is related to better memory, even when that prediction is inaccurate. This is in line with the generation effect (Potts & Shanks, 2014; Slamecka & Graf, 1978), which showed that people better remember unfamiliar study materials (meaning of foreign words) if they make a guess before seeing the correct answer. Potts and Shanks (2014) demonstrated a benefit even for minimal predictions, when predictions are based on random guessing and are wrong almost all the time. In our study, although participants never made explicit predictions, high confidence fixations (even to the

wrong square) could benefit memory through a similar predictive process.

Memory is also best when predictions are most accurate (the observed move is the least surprising, coming exactly where participants anticipated it to be). Surprise or atypicality has sometimes been shown to improve memory, since it attracts attention and can result in stronger encoding (Frank & Kafkas, 2021; Neuschatz et al., 2002; Quent et al., 2021). On the other hand, unexpected events can sometimes be difficult to retrieve at recall, since they are less connected to the schema (Bower et al., 1979; Frankenstein et al., 2020). It is possible that the degree of surprise makes a difference; as discussed above, we never presented invalid moves or unexpected categories of stimuli, which could have strongly driven attention. The type of task used might also matter, as it has been previously shown that schema-inconsistent object-location pairs are better recognized, but not better recalled when given only location as a cue (Lew & Howe, 2017).

Previous research has found a relationship between pupil dilation and surprise (Antony et al., 2021; Lavín et al., 2014; Preuschoff et al., 2011), but our study was not well-suited to measure subtle changes in pupil diameter. Studies of pupil dilation generally require tight control over the experimental stimuli, using either purely non-visual stimuli (Preuschoff et al., 2011), or by presenting luminance-controlled stimuli at fixation (Lavín et al., 2014). In our study the luminance at fixation is highly variable, since eye movements are not controlled and squares can be unoccupied (grey) or occupied by white or black pieces. Future work could attempt to study pupil dilation in this paradigm using models to control for local and global luminance (Antony et al., 2021) or by modifying the stimuli to ensure that all board spaces are isoluminant.

To conclude, we showed that people spontaneously engage in more sophisticated predictive processes as their schemas develop, which are beneficial for memory. The current

study adds to the literature showing the adaptive values of making predictions for perception and action, and extends its benefit into memory domain. It also provides a novel mechanism for the benefit of schemas on memory. Our paradigm uses a complex temporal schema that is much more complicated than most artificial schemas (e.g., Tompary et al., 2020; van Buuren et al., 2014) but does not take years to develop, like a chess schema (e.g., Chase & Simon, 1973; Gobet & Waters, 2003). This makes 4-in-a-row an exciting testbed for future studies of the cognitive and neural mechanisms of schema development and its impact on episodic memory.

## **Chapter 2: Accurate predictions facilitate robust memory encoding independently from stimulus probability**

*A version of this chapter was published as:*

*Huang, J., Furness, E., Liu, Y., Kenmoe, M. J., Elias, R., Zeng, H. T., & Baldassano, C.  
(2025). Accurate Predictions Facilitate Robust Memory Encoding Independently From Stimulus  
Probability. Open Mind, 9, 940-958.*

## 2.1 Introduction

Our perception and memories of the world are scaffolded by our prior experiences. We can use schemas, the structured knowledge built from these prior experiences, to make predictions about upcoming events (Clark, 2013; Friston, 2010; Huang et al., 2023). For example, we can predict what characters in a story might do next or whether an athlete is about to score in a game. These predictions have important implications for how we remember events, but the ways in which accurate or inaccurate predictions impact subsequent memory are still controversial. While numerous research studies have shown that prediction errors lead to better episodic memory (Antony et al., 2023; Bein et al., 2021; Jang et al., 2019; Quent et al., 2022; Rouhani et al., 2018; Wahlheim et al., 2022), in part through enhanced encoding (Frank & Kafkas, 2021; Neuschatz et al., 2002), previous work found that prediction accuracy was associated with better memory (Huang et al., 2023), and learning new information is generally easier when it is congruent with prior knowledge (Bein et al., 2015; Brod & Shing, 2019; Buuren et al., 2014; Quent et al., 2022). In most memory paradigms, however, it is difficult to determine whether making an accurate prediction has a causal effect on memory at all, since the accuracy of predictions (made before the stimulus appears) is almost exactly confounded with the schema-consistency of the stimulus (the probability of the stimulus occurring in the current context). Probable stimuli are more likely to be predicted, and improbable stimuli tend to elicit prediction errors (Quent et al., 2021; Schliephake et al., 2021).

Because these two concepts are so closely related, prior work has largely conflated cognitive processes related to prediction with those related to probability, but in fact these may engage quite different mechanisms occurring at different points in time. Before a stimulus is presented, we can generate predictions about this stimulus based on what we have recently

observed. These predictions will often be sparse and incomplete; for realistic events, there is generally an enormous space of possible outcomes and only limited time and cognitive resources available to make predictions. After experiencing the stimulus, we can assess the accuracy of our prediction, and (whether we were right or wrong) we can try to make sense of the outcome by using our schematic knowledge to link it to our prior observations. For example, even a chess Grandmaster will sometimes commit a blunder in a game, especially under time pressure, failing to predict an opponent's move but immediately recognizing the move as sensible after observing it. An engaging narrative will often include events that we did not predict (Baldassano, 2023), but that in retrospect can in fact be meaningfully integrated into our current event model, such as when a character is revealed to be a villain and we can recognize in hindsight that this is consistent with previously-unexplained events. This dissociation between pre-stimulus predictions and the schema-consistency of the observed stimulus in fact plays a key role in some cognitive theories of humor (Raskin, 1984), which propose that jokes intentionally cause listeners to fail to predict a schema-consistent punchline.

In most lab-based paradigms, however, there are a very small number of possible outcomes, and participants can predict all the outcomes that “make sense” (have high probability). When memory differences are observed between predicted and unpredicted stimuli, it is therefore unclear whether these differences are driven by the match between the stimulus and the pre-stimulus predictions per se (prediction accuracy) or by post-presentation evaluations of the match between the stimulus and the schema (stimulus probability). The current study aimed at investigating the specific impact of accurate prediction on memory, separate from the probability of the stimulus, and the potential mechanisms behind prediction- and probability-related effects. These mechanisms could involve processes during encoding (Bransford &

Johnson, 1972) that improve memory precision (Bellana et al., 2021), or reconstruction processes that improve retrieval of specific kinds of information (R. C. Anderson & Pichert, 1978).

Schematic knowledge can serve as a probabilistic prior, biasing responses to be more schema-consistent (Alba & Hasher, 1983; Bartlett, 1932; Bransford & Johnson, 1972; Cheng et al., 2016; Graesser & Nakamura, 1982; Hemmer & Steyvers, 2009; Huttenlocher et al., 1991; Ramey et al., 2022), or providing retrieval cues that can allow access to weak episodic memories (Qureshi et al., 2014; Watkins & Gardiner, 1979). Recent work in visual scene perception has shown that patterns of visual attention for a repeated image are driven differentially by episodic memory versus schematic knowledge (Ramey et al., 2022), suggesting that eye movements in response to a memory cue could index the degree to which a schema-based retrieval strategy is being used. If prediction accuracy and stimulus probability influence memory differently, different strategies might be used when people encode and recall moves that are probable compared to moves that are predicted.

In this study, we used a paradigm recently developed (Huang et al., 2023), in conjunction with real-time eye-tracking, to manipulate prediction accuracy separately from stimulus probability and to determine when a schema-based strategy was used during retrieval. We used a game called 4-in-a-row, an extension of tic-tac-toe, where two players compete to connect four pieces in a row (in either horizontal, vertical, or diagonal direction) on a 4x9 board (van Opheusden et al., 2023). In a previous study, it was found that participants spontaneously engaged in predictive eye movements when trying to encode game sequences shown to them, and that making predictions consistent with the gameplay model was associated with improved subsequent memory. However, move predictions were much more likely to be accurate when the move was probable (according to a model of likely moves during effective gameplay); we

therefore could not determine whether generating an accurate prediction had a specific consequence for memory, separate from the probability of the observed move. In the current study, we adjusted the presented moves in real time, controlling both the probability of the move and whether the specific predictions a participant was making for this move was accurate or not (by analyzing eye-movement data in real time). We therefore independently manipulated prediction accuracy and stimulus probability by showing moves that people were predicting vs. not predicting, as well as moves that were probable vs. improbable. We hypothesized that both prediction accuracy and stimulus probability separately contribute to better memory, as reflected in higher accuracy, higher confidence, and faster reaction times.

We additionally developed a method for using eye-movements to detect the use of schematic knowledge at retrieval, allowing us to study the mechanisms by which stimulus probability and prediction accuracy influence recall. We hypothesized that probable moves would be encoded through a lower-precision gist-like representation (Bellana et al., 2021), requiring more reliance on schematic knowledge at retrieval. Current theories, however, provide conflicting hypotheses about the effect of prediction accuracy on recall strategy. Stimuli that do not generate a prediction error might be less salient, leading to a less robust episodic memory trace requiring more schema-based reconstruction at retrieval. Alternatively, generating an accurate representation of a stimulus before it appears could enhance the depth and quality of encoding due to additional encoding time (Naim et al., 2020) or by eliciting a positive emotional response (A. Y. Lee & Sternthal, 1999).

We conducted two experiments in which participants learned the rules of the 4-in-a-row game and then attempted to remember moves that were generated by combining participants' in-the-moment eye-movement with probable move locations. Moves separately varied in their

probability (their likelihood to occur during a 4-in-a-row game) and the degree to which they were predicted by the participant (based on their eye movements). Across both experiments, we found that prediction accuracy and move probability independently contributed to better memory, but through different mechanisms: accurately-predicted moves were more precisely encoded and could be recalled without relying on the game schema, while probable moves were remembered through a schema-based recall process.

## **2.2 Method**

### ***Participants***

For Exp. 1, we recruited participants through the Columbia University RecruitMe platform. They were paid \$30 for the completion of the experiment with up to \$10 bonus based on performance in both gameplay and memory task. The Experiment took about 2 hours in total (1 hour gameplay + 1 hour memory task). For Exp. 2, we used the Columbia SONA platform to recruit students seeking research participation credit as part of their introductory psychology classes. We reduced the number of games played by the participants to ensure that the whole experiment took less than the departmental limit of 1.5 hours for SONA participants. All participants were over 18 years of age and gave informed consent for the experiment. The experimental protocol was approved by the Institutional Review Board in a R01 University (AAAS0252).

There were 37 participants for the Exp. 1 (29 female, 9 male), all with normal or corrected-to-normal vision. These participants had a mean age of 26.02 (SD=7.49). The racial makeup of the group consisted of 6 who identified as mixed race, 14 as Asian, 4 as Black or African American, and 13 as White, and 1 who declined to answer. The sample size for Exp. 1 was similar to the previous study with this paradigm, and was intended as a proof-of-concept that

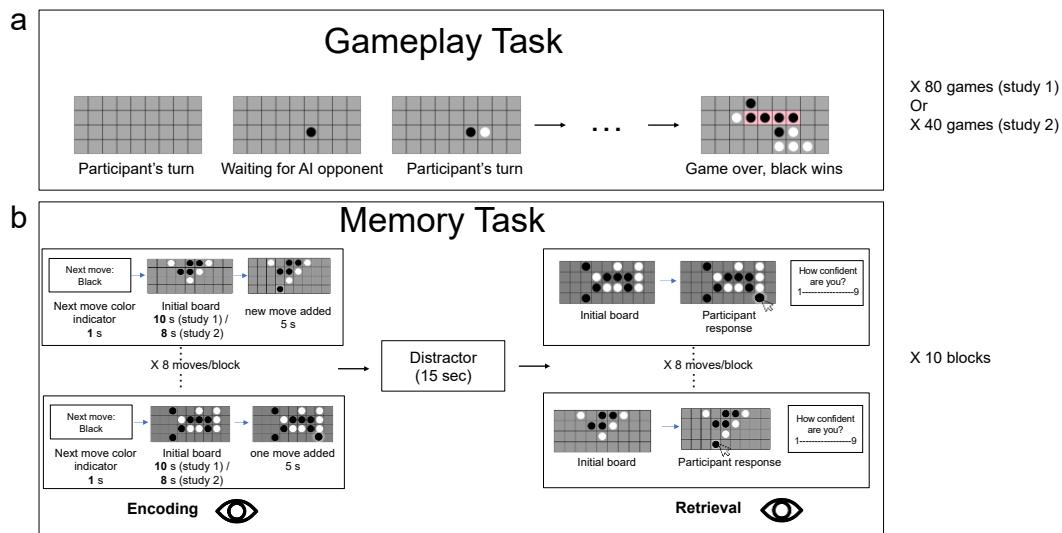
our paradigm could successfully manipulate prediction accuracy independently from move probability. Exp. 2 was pre-registered as a higher-powered replication with 80-100 participants, which was our estimate of how many participants we could collect from our University subject pool over the course of a full semester. In Exp. 2, a total of 120 participants signed up, and 105 participants completed both the gameplay and memory portions of the experiment. We excluded 9 participants due to failure to follow instructions ( $N = 3$ ) or extreme difficulty with eye-tracking ( $N = 6$ ), often due to participants wearing glasses; note that due to the rules for Columbia's SONA program, we were not allowed to apply pre-screening criteria for participants to select only participants who did not require glasses to correct their vision. Data from one participant were lost due to technical errors while running the experiment. The final sample consisted of 95 participants: 35 male, 52 female, and 8 whose demographic information was missing. These participants had a mean age of 20.61 ( $SD=3.82$ ). The racial makeup of this group consisted of 10 who identified as mixed race, 24 as Asian, 7 as Black or African American, 37 as White, and 6 as other.

### ***Design & Procedure***

The study was a two-part experiment. Participants first completed an online task where they learned the rule of the 4-in-a-row game and played 40 games to get familiar with it. The gameplay task was built on Psiturk (Gureckis et al., 2016) and hosted on Heroku (<https://www.heroku.com/>). In the second part, they came into the lab and complete the eye-tracking portion. Two experiments were conducted, and Exp. 2 design and analyses were pre-registered in AsPredicted (<https://aspredicted.org/jv8qv.pdf>).

The procedure of the experiments can be seen in Figure 2.1. Participants played 80 (Exp. 1) or 40 (Exp. 2) games against an AI opponent taken from (van Opheusden et al., 2023). The

gameplay was completed 1-7 days before the day of their scheduled in-person session. The game is an extension of tic-tac-toe, in which two players take turns to add a move on a 4x9 board. The first player to connect four pieces of theirs in a row (either horizontal, vertical, or diagonal) on the board wins. Participants were told how to win the game, and then played games against an adjustable AI opponent. When the participant won, the AI became stronger, and vice versa. Participants and AIs took turns moving first. They were told the current level of the opponent at every game. They were told to achieve a target level that represents a strong player and (in Exp. 1 only) that their bonus depended on the level they reached.



**Figure 2.1: Experimental design. a.** Online four-in-a-row gameplay task. Prior to coming to the lab, participants played 80 (Exp. 1) or 40 games (Exp. 2) against AI opponents. The participant and the AI took turns placing pieces on a 4x9 grid, with the goal of connecting four pieces in a row (in any direction) to win the game. **b.** In-lab memory encoding and retrieval task. In each block, participants first went through 8 encoding trials. Each trial indicated which player would make the next move for 1 second, a trial-unique initial game board was shown for 10 (Exp. 1) or 8 (Exp. 2) seconds, and then the next move was added to the board and shown for 5 seconds. Participants' task was to remember the single move associated with each board. After seeing all the eight boards and the individual moves associated with each board, participants completed a 15-second distractor task, in which participants judged whether a single digit number was odd or even, participants were shown the initial boards again (in a random order), and recalled the move shown for that board. After they selected the move, the board remained on the screen until participants responded to a prompt underneath the board requiring them to rate their

confidence in their response. Participants' eye movements were tracked throughout both encoding and retrieval.

For the eye-tracking portion of the experiment, participants came in to the lab and signed the consent form. They were given the overall instruction that: "To give you an overview of the task - in this experiment you will remember and recall moves appearing on 4-in-a-row boards. This task is about 60 minutes and can be quite difficult, so you should try to use your knowledge about the game to help you. The instructions for how to complete the tasks will be on the screen." Participants were seated 100 centimeters in front of a monitor and placed their heads in a chin rest 45 centimeters away from the eye-tracker. They were instructed to remain as still as possible while the eye-tracker was running and were told that they could take breaks during the experiment in between blocks. Before beginning the experiment and when the participants returned from their breaks, the eye-tracker calibration and subsequent validation were done using a nine-point grid. We recorded right eye movement using EyeLink 1000 plus at 1000 Hz recording frequency. Light levels remained constant for the duration of the 60 min memory portion of the experiments. The stimuli were displayed on a 24-inch LED monitor, with a resolution of 1920 by 1080 pixels and a refresh rate of 60 Hz. The outputted EDF files were converted to asc files and parsed with PyGaze (Dalmaijer et al., 2014). The experiment was programmed with PsychoPy.

Each of the ten blocks in the memory task consisted of eight trials. For each trial, participants first saw text for one second indicating which color player would be playing this move. This is followed by an initial board for 10 (Exp. 1) or 8 (Exp. 2) seconds, taken from the middle of a game between two AI players as in (Huang et al., 2023), and then the move was added to the board and shown for 5 sec. The number of pieces on the board ranged from 4 to 31, with a median of 13 and an average of 13.52 pieces ( $SD = 5.97$ ). Participants were instructed to

watch and remember single moves placed on different boards. In Exp. 2 they were additionally explicitly instructed to try to guess the location of the move and move their eyes to that location before it appeared. This instruction was designed to ensure that participants' fixations on empty squares indeed reflected their predictions. We additionally changed how moves were generated in Exp. 2 to encourage participants to make more predictions with their knowledge of the game (more details in "real time generation of the move" section in Methods). After all eight moves were shown, there was a 15-second distractor task in which random one-digit numbers flashed at the center of the screen every 1 sec and participants were instructed to a button every time an even number appeared. The distractor task was designed to be relatively short because the task of remembering eight moves is highly challenging even at short retention intervals. Participants were then shown the eight initial boards again, one by one, in random order, and recalled the move that was placed on each board by clicking on the corresponding location. After each recall, they also rated their confidence about the move on a scale of 1 to 9 (with 9 being most confident) by pressing the keyboard.

A persistent technical issue with the eye-tracker sometimes caused the experiment to freeze at random points in the middle of the experiment during both experiments. In case this happened, we continued the experiment from the next block and did not include data from that block in that participant. As a result, 9 participants in Exp. 1 and 13 participants in Exp. 2 had one or two out of the ten blocks missing from their data. The data in the intact blocks of these participants were used for the analysis.

### ***Fixation smoothing and gameplay model***

We used the same approach for processing fixations and generating probabilities from the gameplay model in our previous study (Huang et al., 2023). Fixation maps were created for each

10- or 8-second period before a move was shown. To handle uncertainty in assigning gaze to squares, we performed a soft assignment to board locations based on distance. For a fixation at position  $x_F$  with duration  $t_F$  the square with center coordinate  $x_i$  was assigned a fixation weight of

$$t_F \cdot \frac{e^{-\|x_F - x_i\|_2 / 25}}{\sum_j e^{-\|x_F - x_j\|_2 / 25}}$$

Here, distance is in the unit of pixels. The length of the square is 180 pixels so the smoothing on the scale of 25 pixels is only relevant for fixations close to square boundaries. The weights for all fixations during the 10- or 8-second window were summed to obtain a final map of fixation weights for all board squares.

The gameplay model is a linear myopic model based on features manually selected for gameplay, trained on games played by strong AI with PyTorch (Paszke et al., 2019). The model outputs a probability distribution of likely next moves. A full description and validation of this model can be found in our previous study (Huang et al., 2023).

### *Real time generation of the move*

To generate a distribution of a participant's predictions (**prediction distribution**, Figure 2.2a, bottom right) for a board, we obtained a fixation heatmap during the initial board period (10 or 8 sec depending on experiment). Then, the fixations on the empty squares were extracted and normalized such that the fixations sum to one. Similarly, we generated a **move probability distribution** (figure 2.2a, top left) using the gameplay model.

To produce an unpredicted / improbable condition, one or both of these distributions could be inverted, such that the probability of a move  $P(m)$  became:

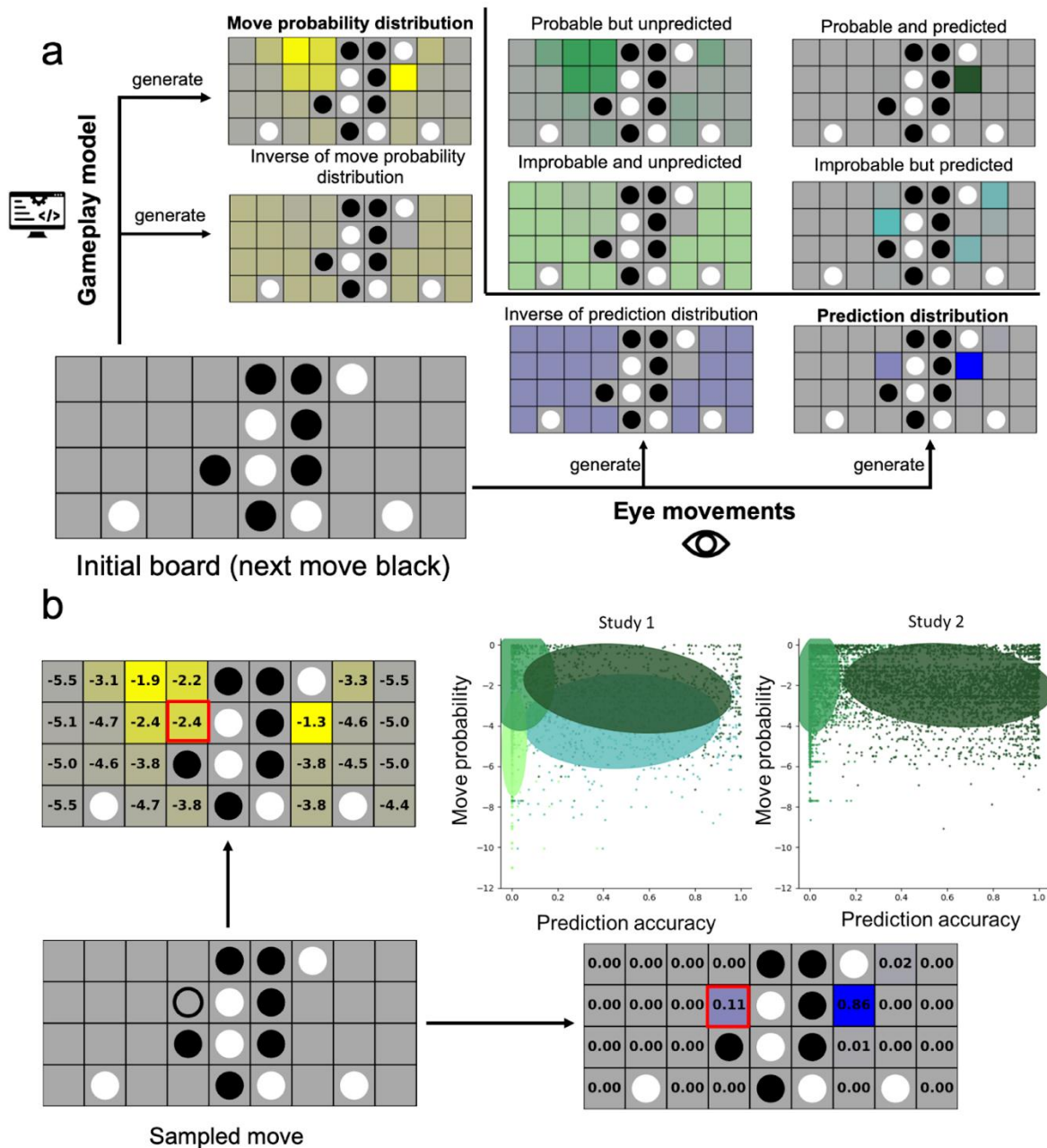
$$\frac{Max(P(m)) - P(m)}{\sum_{i=0}^n (Max(P(i)) - P(i))}$$

where  $n$  is the number of empty squares on the board. This makes the most probable/predicted moves the least likely to occur, while the rest of the moves (originally with near-zero likelihood to occur) became roughly equally likely.

The final distribution for the next move was calculated by multiplying the two distributions together and normalized such that the distributions sum to one.

$$P(m) = \frac{P_{probability}(m) \times P_{prediction}(m)}{\sum_{i=0}^n (P_{probability}(i) \times P_{prediction}(i))}$$

Distributions corresponding to four conditions were generated: probable and predicted, probable but unpredicted, improbable but predicted, and improbable and unpredicted. The move shown to the participants was sampled according to its probability in one of these combined distributions. In Exp. 1, each block contained two moves drawn from each of the four distributions. One potential downside of this design is that this results in many improbable moves, which may discourage participants from using their general knowledge about the game. In Exp. 2, we instead sampled moves only from the two probable conditions, with half of the moves predicted and half unpredicted. All of the distribution calculations were completed at the end of the initial board presentation period of each individual trial, allowing the presented move to be selected based on the computed distribution.



**Figure 2.2: Real-time generation of the moves.** **a:** After the initial board was shown, two distributions were produced: one based on gameplay model that computes the probability of the move to be played by a strong player (yellow), and one based on the participant's predictive fixations during the initial board period (blue). Darker colors indicate higher values in the squares. The move probability and prediction distributions or their inverses were then combined to generate four possible distributions from which the presented move was drawn: improbable and unpredicted (used in Exp. 1 only), improbable but predicted (Exp. 1 only), probable but unpredicted (both experiments), and probable and predicted (both experiments). A move was sampled from one of the distributions. **b:** Distribution of move probability and

prediction accuracy in each of the four or two conditions in Exp. 1 and 2. After a move was sampled, it was evaluated with the move probability distribution and the prediction distribution. This generates a move probability and prediction accuracy for each move (in this example, move probability is -2.4 and prediction accuracy is 0.11). Each point in the scatterplot is a move, and the color of the points correspond to the distribution from which the move was drawn. The ellipses represent the  $3\sigma$  confidence ellipses for each condition, with colors corresponding to the conditions in (a), showing where most of the points in each condition are located. The correlation between prediction accuracy and move probability is close to zero.

### ***Key measures***

For our main analyses, we ignored the binary condition labels under which moves were generated and instead measured prediction accuracy and move probability values for each specific move (Figure 2.2b). Move probability is the log probability according to the gameplay model of the move that was shown. Prediction accuracy is the percent of fixations at empty squares that were focused on the move that was shown. Both move probability and prediction accuracy were z-scored within each Experiment, across all participants and all moves.

We also derived a model-based measure of schema-based eye movements at retrieval (Figure 2.4a). The fixations during the retrieval period were extracted and converted to heatmaps as described above. For each move, fixations on empty squares were extracted and normalized to sum to 1. We then ran a linear regression with the extracted fixation heatmap as the outcome variable, and two regressors: one consisting of the probability distribution over possible next moves according to the gameplay model, and another that had a value of 0 everywhere except at the square corresponding to the correct move, where its value was 1. The resulting coefficients for these predictors were  $w_{moveProb}$  and  $w_{correctMove}$ .

### ***Statistical models***

We first z-scored all the variables (other than RT and confidence, which are more intuitive in their raw form), such that the betas reported in the paper reflect the effect size in terms of how many SD changes in outcome variable is related to changes in one SD of the

predictor variable. All statistical models were fit in R with the lme4 Package (Bates et al., 2015). We started with the most complex model with random subject effects for all regressors. In cases where the models did not converge, we simplified the models to reduce the number of random effect terms, until only random intercepts remained in the model.

For the mediation analysis, the significance of the mediation was computed with the package *mediation* (Tingley et al., 2014) that uses a bootstrapping procedure. Standardized indirect effects were computed for each of the 10,000 bootstrapped samples, and the 95% CI was computed by determining the indirect effects at the 2.5th and 97.5th percentiles.

### ***Pre-registration and data sharing***

The experimental design and potential analysis were pre-registered in AsPredicted (<https://aspredicted.org/jv8qv.pdf>) before we started data collection for Exp. 2. The results below present a simplified version of the pre-registered analysis plan; see the supplementary material for full results from the planned analyses in the pre-registration. We also conducted additional exploratory analyses looking at how retrieval eye movements are related to the probability of the selected move, to better understand what this measure means in terms of retrieval strategies. All the data and the code can be found online.

## **2.3 Results**

### **2.3.1 Manipulation check**

In our previous study, it was found that experienced participants spent more time fixating on empty squares where a move was likely to appear in typical play (based on a game model); since the actual moves shown were drawn from this same distribution, this led to a substantial correlation between prediction accuracy and move probability ( $r = .228, p < .001$ ) (Huang et al.,

2023). We also found this relationship in the current experiments; computing the mean correlation between fixation on empty squares of the initial board and the probability of the moves for each subject, a one-sample t-test across subjects showed that these correlations were significantly above 0. (Exp. 1: Mean correlation = 0.273,  $t(36) = 25.92$ ,  $p < 0.001$ , Cohen's  $d = 4.20$ ; Exp. 2: Mean correlation = 0.354,  $t(94) = 43.91$ ,  $p < 0.001$ , Cohen's  $d = 4.51$ ), meaning people tend to make predictions on probable moves. Despite the correlation, our real-time procedure for selecting moves allowed us to separately control the extent to which moves were likely and were predicted. Figure 2.2b shows the distribution of move probability and prediction accuracy for all the moves in the experiment, demonstrating that our conditions successfully sampled from different portions of this space and decorrelated prediction accuracy from move probability in both Exp. 1 ( $r = .074$ ) and Exp. 2 ( $r = -.057$ ). While these correlations are still significantly different from 0 due to the very large number of moves in the experiments ( $p < .001$ ), they are much smaller than in our previous study (and have different signs in each experiment), allowing us to more rigorously estimate the independent impact of move probability and prediction accuracy on memory. Because moves are sampled probabilistically, they vary within and across our four conditions in terms of both their probability to occur and the degree to which the participant predicted them on this trial. For the rest of the paper, we focus on move probability and prediction accuracy for each move as continuous-valued predictors rather than treating each condition as a discrete category.

### **2.3.2 Move probability and prediction accuracy both improve subsequent memory**

Although the task is difficult, requiring participants to remember 8 moves associated with 8 unique boards, overall performance was reasonably high with a mean accuracy of 0.49 (SD = 0.19) in Exp. 1, and 0.55 (SD = 0.15) in Exp. 2. This is much higher than the level of

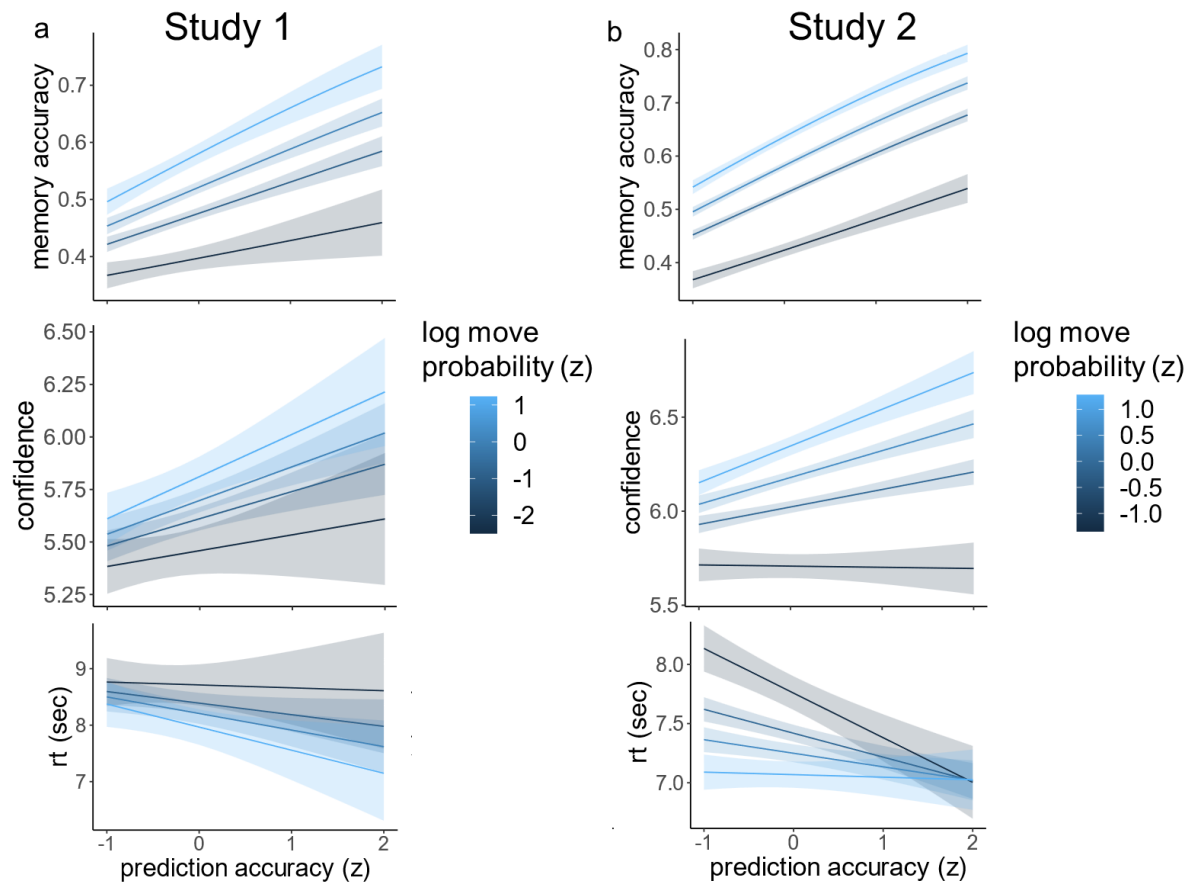
performance that would be achieved if participants were guessing randomly (Mean accuracy 4.95%). We also simulated the mean accuracy if participants used the gameplay model to guess the next move, instead of their episodic memory, which gave us accuracy still lower than the accuracy of the participants (Exp. 1: 25%; Exp. 2: 31%), suggesting that they indeed used episodic memory to complete the task. The mean confidence (out of 9) is 5.65 (SD = 1.24) in Exp. 1 and 6.11 (SD = 1.19) in Exp. 2. The mean reaction time is 8.3 sec (SD = 3.33) in Exp. 1 and 7.34 (SD = 2.35) in Exp. 2.

We first conducted a mixed-effects logistic regression to predict memory accuracy from move probability, prediction accuracy, and their interactions with random subject-specific slopes for move probability and prediction accuracy (Figure 2.3, top). In both experiments, memory was better for moves that were probable (Exp. 1:  $\beta = 0.233$ ,  $z = 4.64$ ,  $p < .001$ ; Exp. 2:  $\beta = 0.379$ ,  $z = 10.07$ ,  $p < .001$ ) and for moves that were correctly predicted (Exp. 1:  $\beta = 0.289$ ,  $z = 5.19$ ,  $p < .001$ ; Exp. 2:  $\beta = 0.334$ ,  $z = 11.47$ ,  $p < .001$ ). The interactions were not significant in either Exp. ( $p > .30$ ).

We next looked at the effects of both measures and their interactions on memory confidence and reaction time (Figure 2.3, middle and bottom). Due to convergence issues, these analyses were performed with linear mixed-effects models predicting confidence and reaction times from move probability and prediction accuracy, with random intercepts only. In both experiments, participants were more confident about their answers when the move probability was higher (Exp. 1:  $\beta = 0.09$ ,  $t(2879.5) = 2.08$ ,  $p = .04$ ; Exp. 2:  $\beta = 0.258$ ,  $t(7164.5) = 7.90$ ,  $p < .001$ ) and if the prediction accuracy was higher (Exp. 1:  $\beta = 0.159$ ,  $t(2881.2) = 2.90$ ,  $p = .004$ ; Exp. 2:  $\beta = 0.111$ ,  $t(7168.0) = 4.04$ ,  $p < .001$ ). Both move probability and prediction accuracy led to faster reaction times in Exp. 2 (move probability:  $\beta = -0.268$ ,  $t(7167.8) = -$

3.60,  $p < .001$ ; prediction accuracy:  $\beta = -0.170$ ,  $t(7172.1) = -2.71$ ,  $p = .007$ ). Similar results were found in Exp. 1 but the effects were not statistically significant (move probability:  $\beta = -0.200$ ,  $t(2881.0) = -1.35$ ,  $p = .18$ ; prediction accuracy:  $\beta = -0.362$ ,  $t(2883.7) = -1.93$ ,  $p = .054$ ). None of the interactions between move probability and prediction accuracy were significant ( $p > .31$ ). These findings provide strong evidence that both move probability and prediction accuracy separately and causally contribute to stronger memories that are recalled faster and with more confidence.

Since prediction errors could benefit memory through enhanced encoding when remembering something schema-inconsistent (Quent et al., 2022), the relationship between prediction accuracy and memory accuracy could be non-linear. To test this hypothesis in our data, we conducted a mixed effect logistic regression predicting memory accuracy from move probability and prediction accuracy with linear and quadratic terms for both effects, and a random subject intercept. While the linear effects for both measures remained robust in both experiments (all  $p < .001$ ), quadratic effects were largely absent. The square of prediction accuracy showed no effect on memory ( $p = .38$  in Exp. 1 and  $p = .92$  in Exp. 2). Move probability squared did show a significant effect in Exp. 2 ( $\beta = .05$ ,  $z = 1.99$ ,  $p = .047$ ) and trended towards significance in Exp. 1 ( $\beta = .03$ ,  $z = 1.71$ ,  $p = .087$ ), but this effect was weak compared to the linear effect; critically, memory still improved monotonically with move probability (with an attenuated slope for low-probability moves).



**Figure 2.3. Impact of move probability and trial-wise predictions on recall performance.**

Regression analyses were used to model memory accuracy (top), memory confidence (middle), and reaction time (bottom) as a function of both move probability and prediction accuracy in Exp. 1 (a) and Exp. 2 (b). In both experiments, memory was significantly more accurate and confident for more likely moves and when predictions were accurate. These two factors also improved reaction times, though this effect was significant only in Exp. 2.

### 2.3.3 Eye movements reveal multiple distinct retrieval strategies

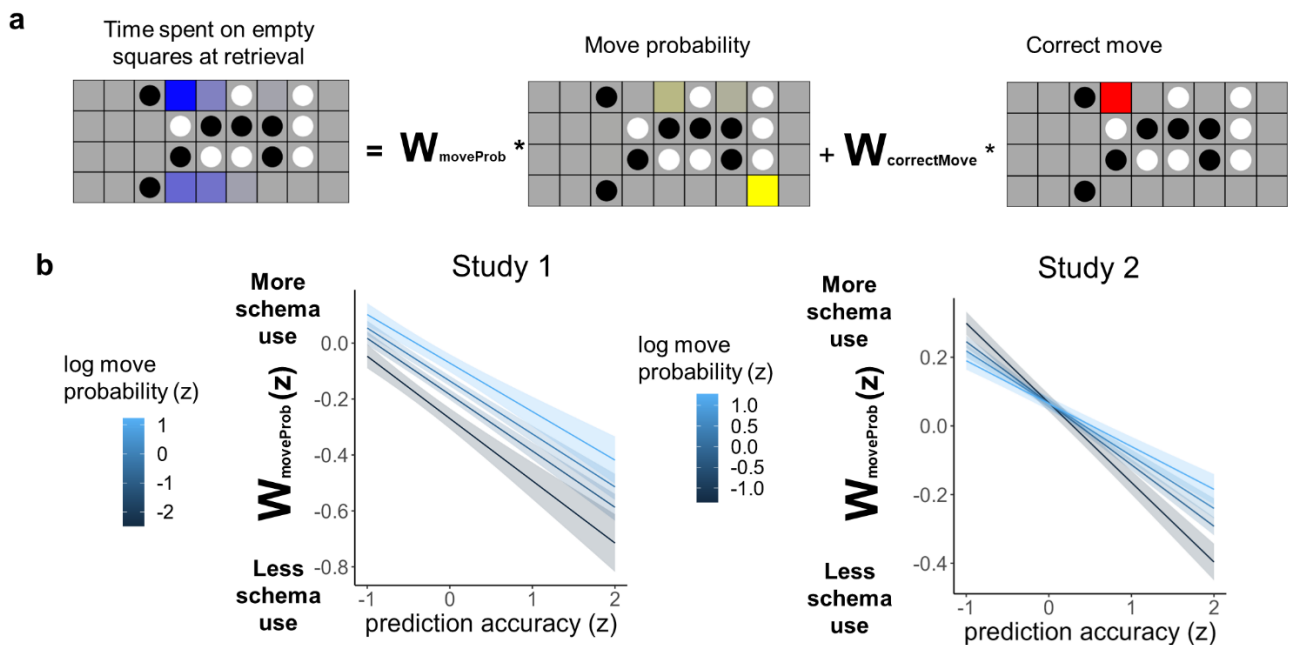
We next sought to understand the strategies that participants are using at retrieval by modeling their eye movements. It's possible that participants could use their schematic knowledge about likely moves for a particular recall board, either by generating plausible moves for the given board then attempting to recognize which move was previously seen (Watkins & Gardiner, 1979) or by simply biasing their guesses toward probable moves. Alternatively, if participants formed a precise episodic memory for a move on a board, they could directly retrieve this memory when the board is shown, without having to make use of schematic knowledge. To use eye movements to detect when participants were using a schema-based strategy at retrieval, we ran a linear regression to predict retrieval fixations on empty squares as a combination of two regressors: (1) the move probability of each empty square (including the correct next move) and (2) the correct next move (with a value of 0 in all other squares) (Figure 2.4a). The coefficient of the move probability regressor ( $w_{moveProb}$ ) represents how often a participant was fixating on likely moves during retrieval for this trial, indicating schema use. Note that, because the correct move is included as a nuisance regressor, simply making fixations on the correct move will not increase the value of  $w_{moveProb}$ .

Since we would expect participants to use their schema at retrieval more when they are more uncertain, we validated that  $w_{moveProb}$  reflects schema use by checking whether it was associated with worse memory. We conducted three mixed effect regressions predicting memory, reaction time, and confidence from  $w_{moveProb}$ . For memory accuracy, we conducted a mixed effect logistic regression, with random slope of  $w_{moveProb}$ . For reaction time, we conducted a linear mixed effect regression, with subject random intercept. For confidence, we

conducted a linear mixed effect regression, with subject random slope of  $w_{moveProb}$ . Overall, we confirmed that larger  $w_{moveProb}$  was associated with worse memory: more schematic eye movements predicted lower memory accuracy (Exp. 1:  $\beta = -0.67$ ,  $z = -7.14$ ,  $p < .001$ ; Exp. 2:  $\beta = -1.18$ ,  $z = -17.98$ ,  $p < .001$ ), higher reaction times both when considering all moves (Exp. 1:  $\beta = 0.47$ ,  $t(2896.2) = 2.76$ ,  $p = .006$ ; Exp. 2:  $\beta = 0.36$ ,  $t(7180.0) = 5.86$ ,  $p < .001$ ) or when restricting to correct moves only (Exp. 1:  $\beta = 0.76$ ,  $t(1419.9) = 4.87$ ,  $p < .001$ ; Exp. 2:  $\beta = 0.59$ ,  $t(3945.0) = 6.09$ ,  $p < .001$ ), and lower confidence for all moves (Exp. 1:  $\beta = -0.40$ ,  $t(26.0) = -5.11$ ,  $p < .001$ ; Exp. 2:  $\beta = -0.38$ ,  $t(90.9) = -9.91$ ,  $p < .001$ ) or for correct moves only (Exp. 1:  $\beta = -0.17$ ,  $t(1408.1) = -2.62$ ,  $p = .009$ ; Exp. 2:  $\beta = -.11$ ,  $t(3937.8) = -2.62$ ,  $p = .009$ ).

We further explored how the degree of schematic eye movements at retrieval ( $w_{moveProb}$ ) was related to participants' responses. One possibility is that schematic eye movements at retrieval simply indicate that participants are preparing to make a schema-consistent guess on that trial. If that is the case, we should see that participants select the most probable moves at recall when  $w_{moveProb}$  is high. Alternatively, schematic eye movements at retrieval could indicate that participants were using their schema to generate potential moves in hopes of being able to recognize the correct move from episodic memory, in which case participants will be able to report the correct move even if it was not the most probable move for this board position. We tested this prediction by measuring the relationship between schematic eye movements at retrieval and the probability of the **selected** move, separately for correct and incorrect trials. We conducted a linear mixed effect regression, predicting probability of selected move from  $w_{moveProb}$ , with a random subject intercept. When a move was incorrectly recalled, schematic eye movements strongly predicted the probability of the selected move (Exp. 1:  $\beta =$

0.58,  $t(1487.3) = 22.79$ ,  $p < .001$ , Exp. 2:  $\beta = 0.28$ ,  $t(3246.3) = 21.16$ ,  $p < .001$ ), suggesting that on these trials schematic eye movements were used to guess a likely move in the absence of episodic memory. When a move was correctly recalled, however, schematic eye movements were not related to the probability of the selected move (Exp. 1:  $\beta = .05$ ,  $t(1407.9) = 1.40$ ,  $p = 0.162$ , Exp. 2:  $\beta = -.003$ ,  $t(3934.0) = -0.17$ ,  $p = .865$ ). This provides evidence that schemas can be used strategically to retrieve episodic memories, rather than simply biasing guessing toward probable outcomes.



**Figure 2.4. Modeling and predicting eye movements at retrieval.** **a:** We predicted the distribution of fixations on empty squares when boards were presented at retrieval, using two regressors: the likelihood of each move according to our gameplay model, and the location of the correct move that was shown during encoding. The coefficient of the move probability regressor measures the extent to which schema-based eye movements were present at retrieval. **b:** In both experiments, when we evoked a prediction error during encoding, we observed more schema-driven eye movements at retrieval. In Exp. 1, moves that were more probable also showed more schematic eye movements; in Exp. 2, this relationship only held for moves that were accurately predicted.

### 2.3.4 Prediction accuracy, but not probability, reduces schema use at retrieval

We next examined whether participants' retrieval strategy (as measured through eye movements) differed for probable versus improbable moves and for predicted versus unpredicted moves. We conducted a linear mixed-effects regression with the schematic eye movement at retrieval ( $W_{moveProb}$ ) as the outcome variable and prediction accuracy (from eye movement before the move shows up), the probability of the move (according to the gameplay model), and their interaction as predictors, with a random subject intercept. In Exp. 1, both move probability and prediction accuracy significantly impacted schematic eye movement, but in the opposite directions (Figure 2.4b). Higher move probability was associated with more schematic eye movement ( $\beta = 0.05$ ,  $t(2890.0) = 3.30$ ,  $p = .001$ ), whereas low prediction accuracy was associated with more schematic eye movement ( $\beta = -1.94$ ,  $t(2897.2) = -9.75$ ,  $p < .001$ ). The interaction between move probability and prediction accuracy was not significant ( $\beta = 0.01$ ,  $t(2893.3) = 0.657$ ,  $p = .511$ ). In Exp. 2, move probability had no main effect on schematic eye movement ( $\beta = -.003$ ,  $t(7222.9) = -0.02$ ,  $p = .811$ ), while low prediction accuracy was associated with more schematic eye movement ( $\beta = -0.18$ ,  $t(7237.4) = -15.40$ ,  $p < .001$ ). There was a significant interaction between the two measures ( $\beta = 0.04$ ,  $t(7250.0) = 2.99$ ,  $p = .003$ ), such that more probable moves led to more schematic eye movement only if the prediction was accurate. Note that the direction of these effects reveals an interesting dissociation between prediction and probability: the moves which relied least on schematic retrieval were those that were simultaneously predicted and also low probability, a combination that would be difficult to observe without our independent manipulation of these two factors. For moves with large prediction errors, the effect of move probability was inconsistent across experiments, with  $W_{moveProb}$  increasing for higher probability moves only in Exp. 1.

Since we found that prediction accuracy decreased  $w_{moveProb}$  and that lower  $w_{moveProb}$  predicted more accurate responses, we tested whether reduced schematic eye movements mediated the memory benefits of accurate prediction. In both experiments, this mediation was significant (Exp. 1: bootstrapped indirect effect = 0.023, 95% CI = [0.0176, 0.03],  $p < .001$ ; Exp. 2: bootstrapped indirect effect = 0.0341, 95% CI = [0.028, 0.04],  $p < .001$ ). This suggests that making accurate predictions during encoding is associated with the creation of precise episodic memories, allowing participants to rely less on schema at retrieval. On the other hand, move probability did not significantly reduce schematic eye movements in either Exp. (and in fact increased them in Exp. 1), providing evidence that better memory for probable moves does not rely on this same mechanism of facilitated episodic encoding.

One potential explanation of our prediction-related memory effect is that, when a move is shown at a participant's gaze position, it increases the total encoding time for the move and therefore improves memory simply through additional exposure. This would be consistent with our finding that accurate predictions resulted in strong episodic memory. If memory is solely driven by the total encoding time, we would expect that the time spent looking at the move after it was shown should predict memory performance, potentially even more strongly than prediction accuracy. We conducted a mixed effect logistic regression predicting memory from fixation on the correct move during the encoding period after the move appeared, with a subject random intercept. We found that longer fixations on the move were not significantly associated with better memory in either experiment ( $p > .427$ ). We additionally performed a Bayesian analysis, comparing the logistic regression model with fixation duration (and a per-subject random intercept) to a null model with only per-subject random intercepts. The Bayes factor for the fixation-duration model was 0.181 in Exp. 1 and 0.055 in Exp. 2, providing moderate to

strong evidence in favor of the null model. This suggests that there is something special about pre-stimulus fixations to the correct move location in the prediction phase that drive subsequent memory.

## **2.4 Discussion**

When experiencing events that unfold over time in naturalistic settings (Chen et al., 2017; H. Lee & Chen, 2021), our episodic memories are scaffolded by our knowledge of the world that we have built through repeated experiences (Baldassano et al., 2018; Masís-Obando et al., 2021). One core benefit of having a schema for an event is that it enables us to make predictions, which can have important implications for memory (Antony et al., 2021; Rouhani et al., 2018). This study aimed to measure the specific impact of successful and unsuccessful predictions on subsequent memory by deconfounding prediction accuracy from stimulus probability. We accomplished this using a novel paradigm in which we measured predictions using real-time eye-tracking and generated stimuli that were consistent or inconsistent with these predictions. Overall, the results support a model in which prediction accuracy and stimulus probability contribute to better memory through different mechanisms; accurate prediction facilitates the formation of a precise episodic memory that is recalled directly, while probable stimuli were more likely to be reconstructed through a schema-based process at retrieval.

### **2.4.1 Accurate predictions facilitate memory**

We found that confirming a participant's prediction improved later memory, whether or not the predicted stimulus was actually probable according to the gameplay model. Decades of research studies have examined how schemas help memory (J. R. Anderson, 1981). For schema-consistent information, like a pan on a stove, schemas not only facilitate rapid consolidation

during encoding (Sommer et al., 2022), but also allow people to come up with guesses at retrieval that could serve as cues for recognition (R. C. Anderson & Pichert, 1978; Watkins & Gardiner, 1979). Recent work has argued for a central role of prediction in memory, showing that merely making a schema-consistent prediction is associated with better memory (Huang et al., 2023), and that how schema-inconsistent information is remembered depends on the strength of expectation and prediction error (Quent et al., 2021). Our results show that schematic knowledge can in fact improve memory through two separable mechanisms: by enhancing precise memory encoding through more accurate predictions, and by steering retrieval processes toward likely outcomes.

We found no evidence for improved memory from large prediction errors as reported in previous studies (Antony et al., 2023; Bein et al., 2021; A. Greve et al., 2017; Rouhani et al., 2018). One potential explanation for the lack of effect is the element of surprise in the study, where participants' predictions were violated about 50% of the time. It is possible that in previous studies prediction errors benefit memory because they are rarer and more surprising. However, in a recent study which found a significant benefit of incongruity on memory (Quent et al., 2022), the amount of congruent and incongruent trials was balanced, and the lack of prediction error effect is unlikely solely due to the prediction errors being more surprising in previous studies. It is important to note, however, that many studies showing benefits of prediction error use reward paradigms (Jang et al., 2019; Rouhani et al., 2018), in which prediction errors provide a critical learning signal for improving mental models or action policies. Prediction error in the current study does not provide any new information about rewards or the rules of the game, and therefore may not trigger processes that enhanced memory in these studies. Additionally, outside the context of reinforcement learning, the effects of

prediction error on memory have been less clear. In research on schema, participants are often shown object in a congruent or incongruent context (Quent et al., 2022; Van Kesteren et al., 2013), where incongruent pairs are thought to generate prediction errors. Although some studies have shown better memory for schema-incongruent pairs (Quent et al., 2022), many have found the opposite result (Höltje & Mecklinger, 2022; Ortiz-Tudela, Nolden, et al., 2023; Poskanzer et al., 2025; Van Kesteren et al., 2013). In addition, it is worth noting that research on both reinforcement learning and schema tend to use recognition or forced-choice memory tests (A. Greve et al., 2017, 2019), and research has shown better memory for unexpected items during recognition, but not recall (Lew & Howe, 2017). This paper therefore supports the view that the impact of prediction error on memory is more nuanced than previously assumed (Bein et al., 2023), and further research is needed to find the contexts in which prediction error benefits memory. One factor that could impact memory effects is the level of cortical hierarchy in the brain in which prediction errors occur. Previous studies have demonstrated that information accumulates at increasing timescales as it moves from sensory cortex into higher-level regions like prefrontal cortex (Hasson et al., 2015), and that there is a corresponding increase in the timescale of predictions (C. S. Lee et al., 2021; Tarder-Stoll et al., 2024). The current study relies on predictions that likely rely on higher-order regions such as medial prefrontal cortex (Hasson et al., 2015), and the effect of prediction error on memory might be different for lower-level predictions such as perceptual oddballs (Strange & Dolan, 2001). An alternative account for the observed benefit of prediction accuracy on memory is that this arises purely by increasing the effective amount of encoding time for the move stimulus, which is well known to improve memory in general (Murdock, 1974). It is possible that part of the mechanism through which prediction improves memory is by effectively allowing additional "pre-stimulus encoding" of the

move and relevant visual features if the stimulus can be accurately anticipated before it appears. Prediction can also allow a participant to spend more time fixating on the move once it appears, which could allow for more extensive encoding (through processes not related to prediction per se). However, we found that this additional fixation time on the stimulus itself was not the route by which prediction impacted memory in our study; the amount of time spent fixating on the move after it appeared was not related to subsequent memory. This suggests that, if additional encoding time is playing a role, it is only the early pre-activation of stimulus content generated by predictive processes. We therefore argue that a generic encoding-time explanation fails to account for the specific impact of pre-stimulus anticipation that we observed in our study.

#### **2.4.2 Accurate predictions led to reduced reliance on schema at retrieval**

Past literature has shown that eye-movements during retrieval represent meaningful temporal contexts of the memory representation (Kragel & Voss, 2021). Using eye movements at retrieval to detect when participants were using a schema-based reconstruction strategy, we found that this strategy was used more often for moves that were poorly predicted during encoding. We interpret the findings as evidence that that prediction confirmation leads to an enhanced episodic encoding process (Ramey et al., 2022), perhaps through facilitated processing of the expected stimulus (Sommer et al., 2022), such that they were recalled without having to search through schema-consistent possible moves. In addition, the positive emotional response evoked by correct predictions might make the stimulus more salient (A. Y. Lee & Sternthal, 1999). Another potential mechanism could be related to how predicted and probable information were consolidated (Van Kesteren et al., 2013). Future research could further investigate how memory encoding process and the resulting memory representation differs for predicted and

unpredicted items, to better understand how accurate predictions facilitate precise episodic memory.

The finding that accurate prediction promotes forming a robust episodic memory is particularly relevant to theories of how people remember schema-consistent and schema-inconsistent information. For example, the “Schema-Linked Interactions between Medial prefrontal and Medial temporal lobe” model (van Kesteren et al., 2012) proposes that schema-consistent information will be remembered through reactivation of the schema whereas schema-inconsistent information will be remembered through retrieval of a specific instance memory. Our results are partially consistent with this model, in this sense that moves with high probability (consistent with the game schema) relied more on a generate-and-recognize strategy during recall (Watkins & Gardiner, 1979), though this effect was found only for accurately predicted moves in Exp. 2. This potentially saves attentional resources and allows them to focus on episodically encoding improbable moves (which cannot be easily accessed via a schema-based strategy at retrieval). This suggests that participants might create strong episodic memories only for improbable moves that they know will be difficult to generate at retrieval. This strategy may be especially useful when the task is more difficult, explaining why it is used more in Exp. 1 (which requires memorization of many more low-probability moves). Our findings are also consistent with prior work showing efficient encoding of schema-consistent information, but at the cost of memory precision (Bellana et al., 2021).

However, we also showed that accurate predictions (which are more likely to occur for schema-consistent moves) created strong episodic memories that could be retrieved without engaging schematic processes. Our findings could potentially explain why Quent et al. (2022) found that both schema-consistent and inconsistent items were associated with better

recollection, if participants were able to make accurate predictions for the schema-consistent items. Future studies should consider the impact of prediction confirmation when designing experiments, since parameters such as the interval between context and item might facilitate or inhibit predictions and impact schema effects on memory.

### **2.4.3 Methodological implications**

Finally, this work developed two methodological advances that can provide new insights in memory research and related fields. Although gaze-contingent paradigms have been used to study vision (Rayner, 1975) and social cognition (Q. Wang et al., 2020; Wilms et al., 2010), and eye-tracking methods have been used in memory research (Clewett et al., 2020; Ramey et al., 2022; Wynn et al., 2019, 2020), adaptive paradigms with real-time eye-tracking have not previously been applied to memory research. Our work demonstrates that it is possible to track predictions over a large space of potential actions (on a 4x9 game board) to control when prediction errors occur, isolating the impact of prediction errors from factors such as outcome likelihood or reward magnitude. A common challenge in studying the impact of schematic knowledge is that it is difficult to disentangle encoding-time and retrieval-time mechanisms, motivating manipulations such as changing the schema between encoding and retrieval (R. C. Anderson et al., 1983; Bransford & Johnson, 1972). Our study established a novel method of disentangling these two kinds of processes, and also showed that eye movements during retrieval can index not only episodic memory of the item (Wynn et al., 2019), but also the degree to which schematic knowledge is being used to search for a memory. Both of these methods can be applied to fields including attention, learning and decision making. For example, real-time eye-tracking could be used to dissociate reward and prediction error in learning research (Rouhani et al., 2018), and eye movements during decision making could potentially reveal what strategies

(e.g., episodic memory, model-free learning, model-based learning) were used, in addition to modeling the behavioral choice people made (Nicholas et al., 2022). Similarly, in studies of visual search (Castelhano & Heaven, 2011; Wynn et al., 2020), our modeling approach could provide insights into how different sources of information might be used.

To conclude, the current study used a complex board game in combination with real-time eye-tracking to test how prediction accuracy and stimulus probability separately contribute to memory. We found that both prediction accuracy and stimulus probability lead to better memory, but through different mechanisms: prediction accuracy boosts the formation of episodic memory, whereas stimulus probability benefits memory by through schema-based inference at retrieval. This study is part of a recent movement to use games to study cognition (Allen et al., 2023), since they probe more complex processes than traditional designs while still allowing for precise quantitative modeling. There are especially exciting possibilities for studying schematic prediction and memory with these paradigms, since both participants and computational models can make meaningful predictions about upcoming moves, even for novel board positions. Since moves correspond to spatial positions in 4-in-a-row, eye-tracking can provide new insight into predictive processes and adaptive experimental designs. Future work in this field should continue to explore the advantages of using game-based paradigms to study the perception and memory of naturalistic sequences.

# **Chapter 3: Distinct neural representation of different types of predictions and prediction errors**

## **3.1 Introduction**

Events in daily life are often highly predictable. Many of our experiences, such as our commute to work, occur repeatedly with the same structure across episodes. Even when we do something that is not routine, such as going to a new restaurant, we typically have some prior experience, or script for what we expect to happen at this kind of location. These routine and scripts allow us to, consciously or unconsciously, make predictions about what might happen next. Prediction is a top-down process by which prior experience at different timescales can influence how incoming stimuli are processed. Prediction has been considered one of the crucial functions of the brain (Bar, 2009; Clark, 2013; Friston, 2010), necessary for processing the upcoming stream of sensory information. Predictions are observed across different levels of the cortical hierarchy (Hasson et al., 2015; C. S. Lee et al., 2021), from the basic sensory prediction that can be observed in mice (Finnie et al., 2021), to predictions based on memory during repeatedly viewing the same stimuli multiple times (C. S. Lee et al., 2021), to novel information that either fits or does not fit into the current context and the prior knowledge that we have (Quent et al., 2021, 2022; van Kesteren et al., 2012).

As discussed in Chapter 2, a few past studies that used relatively simple paradigms have found improved memory for schema-inconsistent information in some cases, and have attributed it to a benefit from “prediction error” (e.g., Quent et al., 2021, 2022). In Huang et al. (2025), we took one step at breaking down the idea of prediction, looking at the relationship between

schema and prediction in a more subtle lens. Specifically, we challenged the idea in the literature that schema-consistent information is always “predicted” whereas schema-inconsistent information is always a “prediction error” (e.g., Quent et al., 2021, 2022). By separating how much the stimuli are probable given the context and how much the stimuli are predicted by the participant before they appear (with real-time eye-tracking to manipulate prediction accuracy), we showed that there were distinct (and additive) impacts of probability and prediction accuracy. Contrary to a common view in the field about the mnemonic advantages of surprising stimuli (Bein et al., 2021), both prediction and schema-consistency benefited rather than worsened memory, though they were associated with different kinds of retrieval strategies. However, the neural mechanisms engaged by prediction accuracy/error and schema-consistency/inconsistency, and the extent to which they diverge or overlap, are not well understood.

In addition to making predictions based on general schematic knowledge, we can often make predictions based on specific past episodes. These predictions based on past episodes are in fact a more common topic of investigation in past research (Bein et al., 2021; Lee et al., 2021; Poskanzer et al., 2025; Tarder-Stoll et al., 2024), because it is relatively easy to create an episodic memory and later evoke a prediction error by repeating this stimulus with a different outcome. Recent work has provided evidence that different neural systems underlie schematic vs. episodic predictions. For example, in Varga et al (2025), participants were shown videos of everyday activities (such as using a washing machine) that either had a typical ending (schema-consistent, such as putting clothes into the machine) or an atypical ending (schema-inconsistent, such as putting flowers into the machine). Participants were then shown the beginnings of the same videos that ended the same or differently from the ones that they had previously seen. Across three studies, the hippocampus showed higher activity when participants

saw a video that was different from what they had seen previously, regardless of whether the videos had typical endings or not. These results suggest that the hippocampus specifically responds to mismatch or prediction errors that are based on episodic memory, rather than schematic knowledge.

In the current study, we looked at the neural mechanisms of different processes that might be happening under the general term “prediction”. We continued using the 4-in-a-row game that was used in the two previous chapters as a useful tool for studying processes of prediction. We had participants remember and recall sequences of moves in fMRI while having their eye movements tracked with an eye-tracker. In Huang et al. (2023), we established that participants spontaneously make predictions when encoding sequences, which can be indexed in eye-movement, where they look at probable next moves. Like Huang et al. (2025), for each move, we measured how probable it was according to the gameplay model and how accurately predicted it was (with the eye-tracking data), and related that to the activity in the brain. Additionally, we included a task where participants made predictions on boards that were either shown previously in the sequence memory task or novel boards that were never shown. This allowed us to create situations where participants could use memory to make a prediction vs. situations where they could use only their schema to make a prediction, and we looked for potential differences in the neural activities when predictions were based on schema vs. episodic memory. We found that move probability and prediction accuracy are associated with distinct activation patterns in the brain, both during encoding and retrieval of these moves, consistent with our hypothesis that these are separable cognitive processes. Additionally, we found that episodic memory-based predictions in the prediction task showed higher univariate activity than schema-based prediction in the retrosplenial cortex. These results further support the idea that

prediction encompasses many subtle processes that separately impact memory encoding and retrieval.

## **3.2 Methods**

### *Participants*

The current dataset consists of 19 participants (8 males and 11 females), with a mean age of 25.8 years old (SD = 4.26). The racial makeup of the participants is 16 White, 2 Asian, and 1 Black. Participants were paid \$80 at the end of the experiment. For both versions of the study, we recruited participants through personal contacts. All participants were over 18 years of age with normal or corrected-to-normal vision (with contacts) and gave informed consent for the study. The experimental protocol was approved by the Institutional Review Board of Columbia University (AAAS0252).

### *Stimuli*

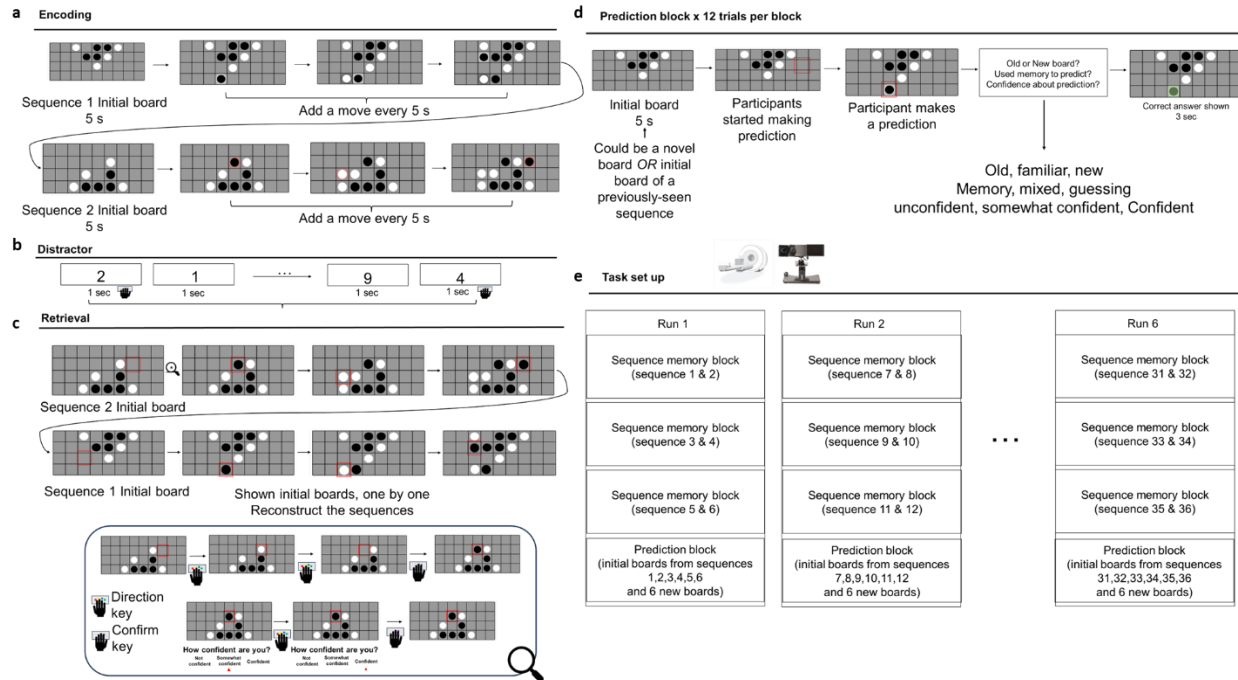
In the sequence memory task, the sequences shown to the participants were taken from the middle of a game between two AI players as in the previous chapters. In the prediction task, six boards were taken from the first board of each sequence from the same run (and the correct move is the same as the first move of the sequence). Six boards were new boards not shown to the participants in the sequence memory task, taken from the middle or the end of a game with two AIs playing against each other.

### *Experimental design*

Before coming into the scanning session, participants first completed a 1-hour gameplay, where they played 80 games against an AI opponent staircased to be stronger when the participant wins a game and weaker when the participant loses.

On the day of the scan, participants were first given a practice run of the task outside of the scanner on the laptop, which consists of three sequence memory blocks and one prediction block. The goal of the practice is to familiarize participants with the task, the response, and make sure that they understand the instructions. Participants in the experiment responded in the scanner with one button box on each hand, each with four keys laid out in one line. In the practice, participants learned the to use direction keys on the laptop with the same layout (6,7,8,9 as the direction key for left, up, down, right, respectively). The experiment starts with an anatomical scan, during which eye-tracker calibration and subsequent validation were done using a nine-point grid. Participants were also given the practice task again, just to familiarize themselves with using the direction key on the button box.

Eye tracking data were collected using an Eyelink 1000 Plus infrared video-based system by SR Research. The camera tracked the participant's right eye at 1000 Hz in a head-stabilized position. Participants were stabilized in the head coil and an MRI-compatible long-range enabled Eyelink 1000 Plus system was positioned outside the scanner bore and beneath the MRI display such that the participant's eye could be tracked through the head coil-mounted mirror.



**Figure 3.1 Overview of the task structure.** **a.** Encoding task. Participants were shown two sequences of three moves, and their task was to remember the sequences. They were first shown an initial board for 5 sec, then a move is added to the board every 5 sec (the two most recently added moves were highlighted). **b.** Distractor. Participants were given a distractor task where they saw numbers flashing on the screen for 1 sec, and they had to make a button press if they saw an even number. **c.** Retrieval. Participants were shown the initial board of one of the sequences, with a red square that was randomly placed on the board. They moved the red square with four direction keys to recall each move in the sequence, and pressed a button at each move location to confirm their choice. They provided a confidence judgment for each move by using the arrow keys to select from 3 confidence levels. **d.** Prediction block. In each trial, participants were shown a board for 5 sec. Their task was to make a prediction about the upcoming move on this board, responding as in the prediction task. Half of the boards were the initial boards of the sequences in the prior three encoding blocks of this run, and participants were told that they could use their memory to make a prediction if they have seen the board before. After they made a prediction, they were asked whether the board was old or new, whether they used memory or guessing (schema) to make a prediction, and whether they felt confident about their prediction. After they answered these questions, they were shown the correct next move for that board. **e.** Overall task structure. There were 6 runs in total in fMRI, each consisting of 3 blocks of sequence memory task (2 sequences per block) and 1 block of prediction task.

Each run of the formal experiments (with fMRI) started with calibration and validation, and consisted of three sequence memory blocks and one prediction block (Figure 3.1e). In each

sequence memory block, participants remember two sequences of three moves each. Initially, an initial board was shown to the participants for 5 sec. Then a new move was added to the board every 5 sec. The newly added moves were highlighted (Figure 3.1a). After two sequences were shown, participants complete a distractor task where they were asked to make a button press when they see an even number by pressing the confirm key (Figure 3.2b). During retrieval (Figure 3.1c), which came after the distractor task, participants were shown an initial board with a red square on a random non-occupied square on the board, and participants' task was to move the red square to the location of the next move on this board by using four direction keys on the button box. They then confirm their choice using the confirm key on their other hand. After that, a prompt occurred on the bottom of the screen asking how confident they were about their predictions, with three options (not confident, somewhat confident, confident). Participants can choose from these three by moving a red triangle to the option they would like to choose with a direction key and press the confirm key.

After three sequence memory blocks, participants were given the prediction task (Figure 3.1d). Participants were first shown an initial board for 5 sec, which is either the initial board of one of the six sequences shown previously in the sequence memory task, or a new board that was never shown to the participants. A red square showed up after the 5 sec, and participants move the red square with the direction keys to make a prediction about the next move on the board. They were told beforehand that if they saw a board which they have seen previously in the sequence memory task, the correct move would be the same as in the memory task, and that they should use their memory to make a prediction in these cases. After they made their prediction by pressing the confirm key, a prompt showed up on the bottom of the screen to ask whether the board is a new or old board (answer: new, unsure, old); whether they used episodic memory or

their knowledge to make a prediction (answer: knowledge, both, memory); and whether they felt confident about their prediction (not confident, somewhat confident, confident). They answer by navigating the red triangle, as described in the previous paragraph.

### ***Gameplay models and eye-movement measures***

The gameplay model and the eye-tracking measures used in the current study were the same as described in previous chapters.

### ***MRI Acquisition***

Whole-brain data were acquired on a 3 Tesla Siemens Magnetom Prisma scanner equipped with a 64-channel head coil at Columbia University. Whole-brain, high-resolution (1.0 mm iso) T1 structural scans were acquired with a magnetization-prepared rapid acquisition gradient-echo sequence (MPRAGE) at the beginning of the scan session. Functional measurements were collected using a multiband echo-planar imaging (EPI) sequence (repetition time = 1.5s, echo time = 30ms, in-plane acceleration factor = 2, multiband acceleration factor = 3, voxel size = 2mm iso). Sixty-nine oblique axial slices were obtained in an interleaved order. All slices were tilted approximately -20 degrees relative to the AC-PC line. There were 6 functional runs in each scan, consisting of 3 sequence memory tasks and 1 prediction task in each run.

### ***fMRI preprocessing***

Results included in this manuscript come from preprocessing performed using fMRIPrep 23.0.2 (Esteban et al. (2019); Esteban et al. (2018); RRID:SCR\_016216), which is based on Nipype 1.8.6 (Esteban et al., (2022); Gorgolewski et al., (2011); RRID:SCR\_002502).

### ***Anatomical data preprocessing***

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008, RRID:SCR\_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 6.0.5.1:57b01774, RRID:SCR\_002823, (Zhang, Brady, and Smith, 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.3.2, RRID:SCR\_001847, (Dale, Fischl, Sereno, 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR\_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization and accessed with TemplateFlow (23.0.0, Ciric et al., 2022): ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al., (2009), RRID:SCR\_008796; TemplateFlow ID: MNI152NLin2009cAsym].

### ***Preprocessing of B0 inhomogeneity mappings***

A total of 3 fieldmaps were found available within the input BIDS structure for this particular subject. A deformation field to correct for susceptibility distortions was estimated based on fMRIPrep's fieldmap-less approach. The deformation field is that resulting from co-

registering the EPI reference to the same-subject T1w-reference with its intensity inverted (Wang et al., 2017; Huntenburg, 2014). Registration is performed with `antsRegistration` (ANTs 2.3.3), and the process regularized by constraining deformation to be nonzero only along the phase-encoding direction, and modulated with an average fieldmap template (Treiber et al., 2016).

### ***Functional data preprocessing***

For each of the 18 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of `fMRIPrep`. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcfliirt` (FSL 6.0.5.1:57b01774, Jenkinson et al., 2002). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al., (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., (2002)). FD and DVARS are calculated for each functional run, both using their implementations in `Nipype` (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (`CompCor`, Behzadi et al., 2007).

Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the  $k$  components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by (Patriat Reynolds, and Birn, 2017). The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in

MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): fsaverage6. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.9.1 (Abraham et al., 2014, RRID:SCR\_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

### ***ROI and searchlight definition***

We used ROIs in the default mode network: angular gyrus, medial prefrontal, and retrosplenial cortex. These ROIs were originally derived from a resting-state network atlas on the fsaverage6 surface (Thomas Yeo et al., 2011).

### ***Univariate analysis***

Spatial smoothing was conducted on the surface post-fMRIPrep, averaging the signal with the neighboring vertices. Whole-brain univariate analysis was conducted on each vertex for each run on the data. Because each run includes encoding, retrieval, and prediction tasks, three task regressors were used as a baseline for other regressors related to measures used in the study. This makes sure the regressors of interest that we included account for more than just the task activity. The parametric regressors, described below, were included in the ridge regression:

Encoding (activity during the encoding period after a move shows up): move probability and prediction accuracy at each move

Retrieval (activity during which participants were navigating the red square): move probability and prediction accuracy of the correct move, eye-movement strategy coefficients (looking at probable moves and looking at correct moves)

Prediction activity (during the 5 sec when participants were looking at the board during the prediction task before the red square shows up): Three regressors corresponding to participants answering old, new, and familiar in the recognition question, and two regressors corresponding to whether episodic memory is used (participants responded with “memory” or “both”) and whether schema is used (participants responded with “guessing” or “both”) in the question about how the moves were recalled.

The coefficient for each regressor is obtained through a ridge regression that include all of the regressors for each run. The significance was calculated through one-sample t-test, comparing the coefficient against zero in each run. For the ROI analysis, we obtained the activity timecourse in the region by averaging across all the voxels in the ROI, and conduct the same ridge regression. We averaged the coefficient of all the runs within a participant for each ROI, and then One-sample t-test was conducted to determine the significance.

### **3.3 Results**

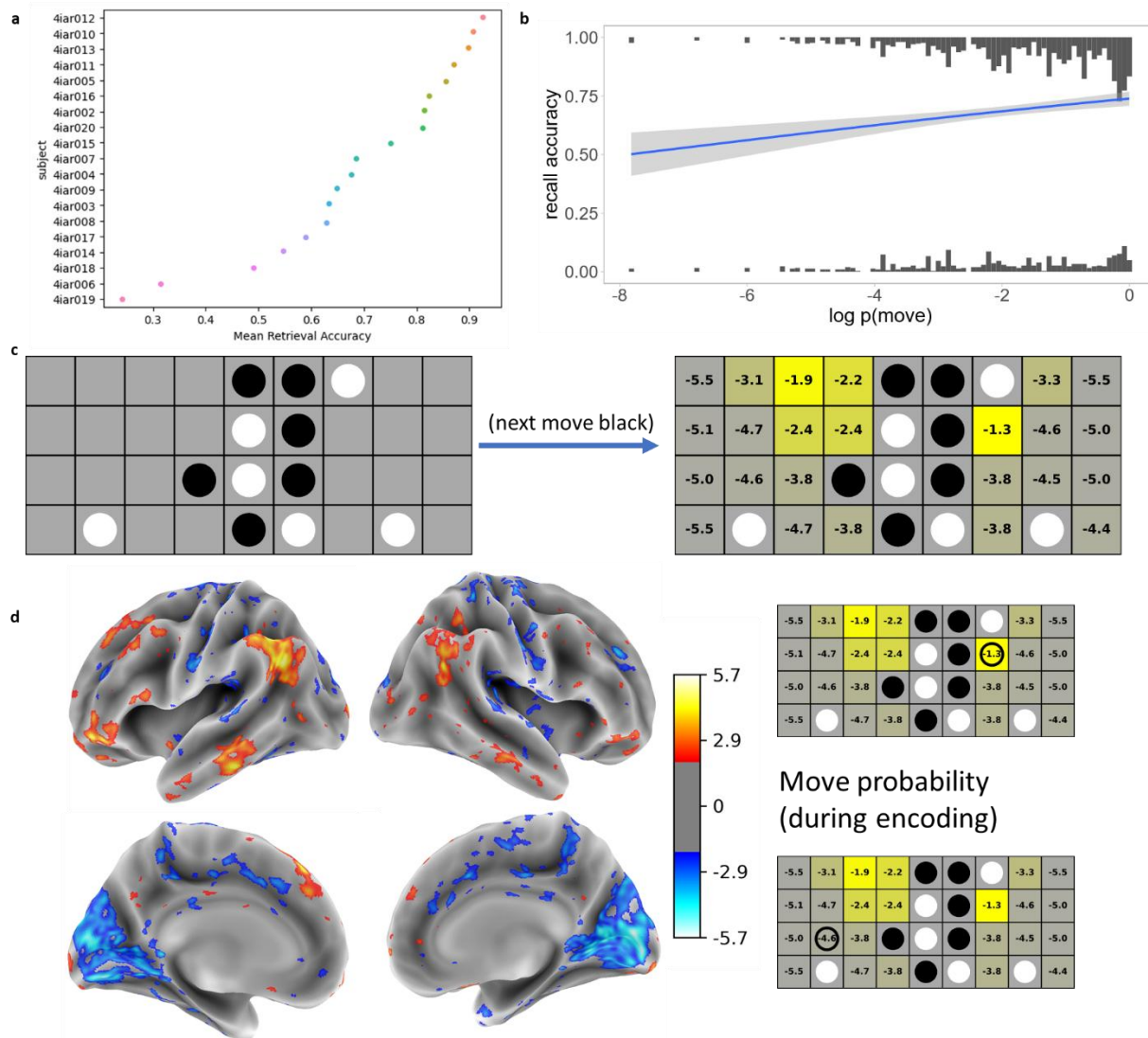
After learning to play four-in-a-row and completing a practice run of the experimental tasks, 19 participants were scanned while they viewed and recalled move sequences in the game. In sequence memory block, participants were first shown two sequences to remember from the

4-in-a-row game, with 3 moves in each sequence (figure 3.1a). After that, participants completed a distractor task (figure 1b), followed by retrieval. At retrieval, participants were shown the initial board, and were asked to recall the sequence. They were also asked to indicate their confidence about their memory after placing each move (figure 3.1c). After three sequence memory blocks, participants completed a prediction block. Participants were shown an initial board, which comes either from the first board of a previous sequence, or from a novel board that they have never seen before. Afterwards, they made a prediction on the board. They then indicated whether the board they saw was old or new, whether they used episodic memory (as opposed to using knowledge/schema of the game) to make a prediction, and how confident they felt about their prediction. They were then shown the correct answer (figure 3.1d). Each prediction block had 12 trials. In the scanner, participants completed 6 runs, each consisting of three sequence memory blocks and one prediction block (figure 3.1e). Their eye movements were monitored throughout the runs with an eye-tracker.

We first looked at participants' memory for the sequences. Figure 3.2a shows the overall sequence memory accuracies for each participant in the experiment. On average, participants recalled 69% (SD = 18.9%) of the moves correctly and in the right order. As can be seen from the figure, most participants achieved above 50% accuracy, demonstrating that participants were able to remember the sequences from the game in the current set up. Like previous chapters, we used a gameplay model to generate the probability distribution of potential next moves on the board (figure 3.2c), and get a measure of the probability of the moves shown to the participants (figure 3.2d, right). We conducted a mixed effect logistic regression, predicting whether a move is recalled correctly from its probability, with a random subject intercept. We show that,

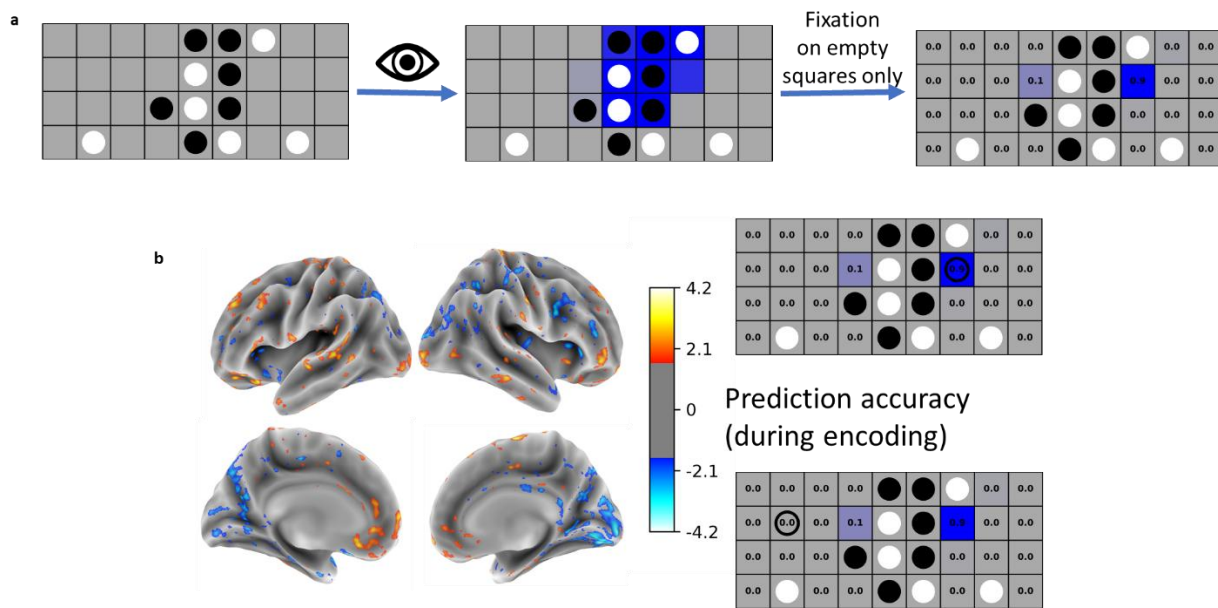
replicating previous studies, more probable moves were more likely to be remembered (beta = 0.163,  $z = 4.66$ ,  $p < .001$ ).

Next, we looked at the univariate activity when participants were encoding moves with high vs. low probabilities. Whole brain voxel-wise univariate regression revealed that encoding more probable moves are associated with higher activity, most predominantly, in the angular gyrus. On the other hand, encoding improbable moves was associated with activity in the primary and secondary visual cortex (figure 3.2d, left).



**Figure 3.2 Memory for the sequences and the relationship with move probability.** **a.** Mean memory accuracy of all the participants in the sequence memory task. **b.** The relationship between move probability and subsequent memory. The histogram on the top and bottom of the figure shows the distribution of individual move datapoints and the line shows the predicted probability of remember the move with the model, with the error bar representing the standard errors of the estimate. **c.** Illustration of the model used for generating the probability distribution of the move given a board. These probabilities are on a logarithmic scale, with higher values (more yellow colors) being more probable moves. **d.** The relationship between univariate activity during encoding a move and the probability of the move ( $p < 0.05$ , uncorrected). The right panel illustrates an example of high and low probability moves, where the hollow black circles represent the moves shown to participants.

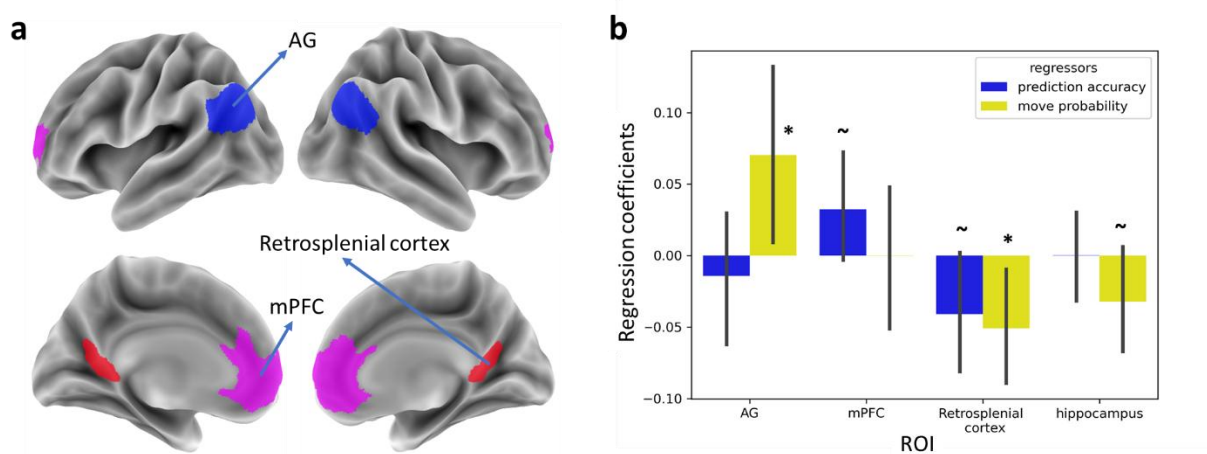
We subsequently looked at eye movements during encoding, measuring how accurately each participant was able to predict each move by calculating the percentage of time spent on the square where the move would occur divided by the total time spent looking at empty squares (figure 3.3, top). The relationship between encoding univariate activity and prediction accuracy was weaker than that of move probability, but as can be seen in the map, higher prediction accuracy is associated with higher univariate activity in mPFC during encoding.



**Figure 3.3 Measurement of prediction accuracy and its correlate with brain activity during encoding.** **a.** When participants were looking at a board, their eye-movements in the 5 sec period were captured by the eye-tracker and converted into a heatmap. Examining fixation times on empty squares, we calculated prediction accuracy as the percentage of the total time looking at empty squares that was spent looking at the correct move. **b.** The correlation between prediction accuracy (illustrated in the right, where given this board and the distribution of eye-movement, if, for example, the top move was selected, it has a high prediction accuracy) and univariate activity in the brain ( $p < 0.10$  uncorrected). Some activation of the medial prefrontal cortex is observed, although it is less clear.

We now test these whole-brain results in the ROI analyses, looking at three ROIs in the default mode network: AG, mPFC, retrosplenial cortex (figure 3.4a), and hippocampus. As can be seen in the figure 4b, higher move probability is associated with significantly higher

univariate activity in AG ( $t(18) = 2.20, p = .04$ ); on the other hand, higher prediction accuracy is associated with higher activity in mPFC, although it is only marginally significant ( $t(18) = 1.71, p = .10$ ). These results further support the idea demonstrated by the previous chapter that remembering probable information in the context might be different from remembering something that was accurately predicted. In retrosplenial cortex, both move probability and prediction accuracy were associated with lower univariate activity (move probability:  $t(18) = -2.51, p = .02$ ; prediction accuracy:  $t(18) = -1.82, p = .078$ ) In hippocampus, high move probability, but not prediction accuracy, is associated with lower univariate activity, although it is only marginally significant ( $t(18) = -1.70, p = .11$ ).

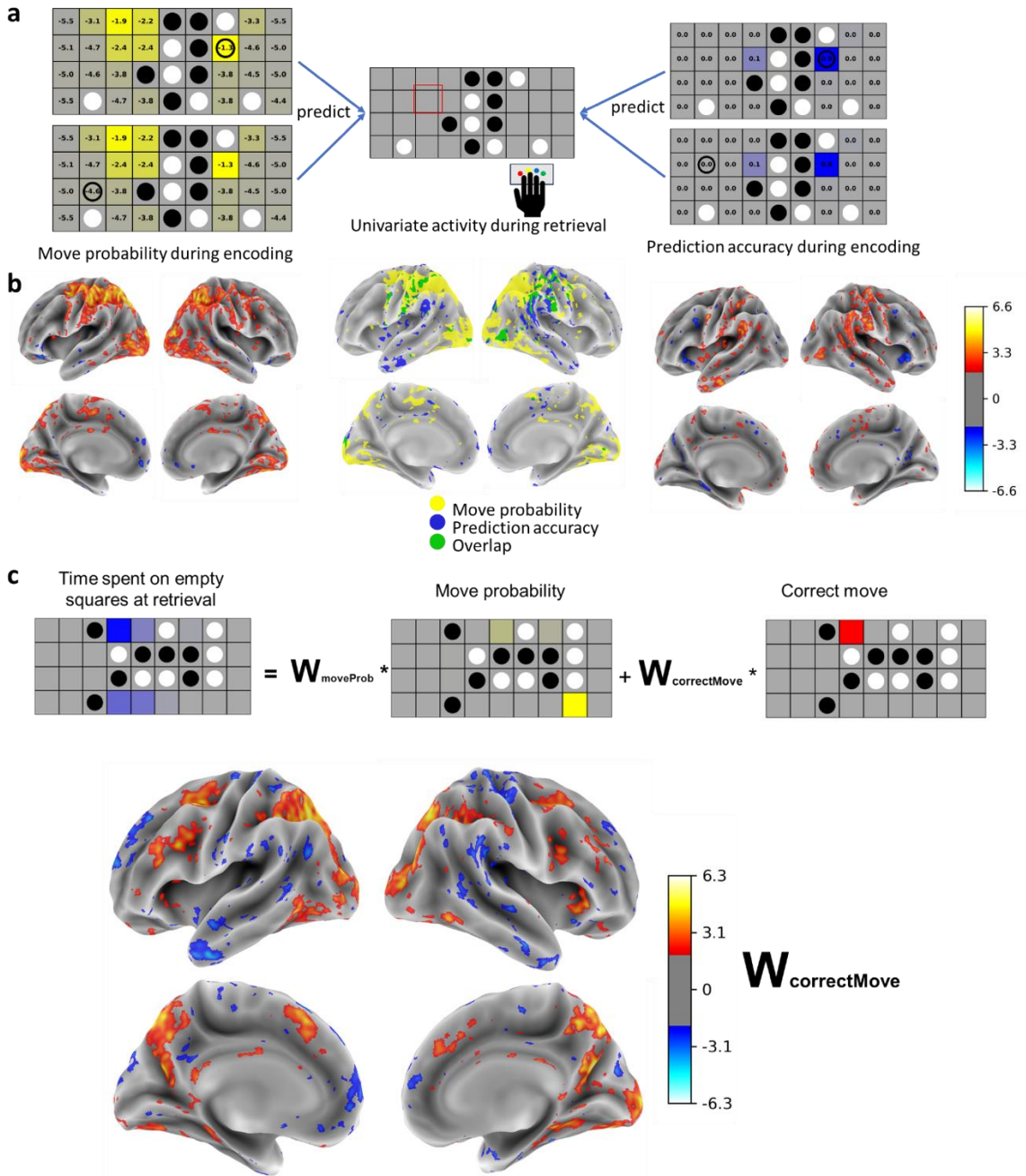


**Figure 3.4. ROI analysis of coefficient of move probability and prediction accuracy in predicting univariate activity in the brain. a.** ROI definition. **b.** The regression coefficients of move probability and prediction accuracy in each of the ROIs. AG and mPFC were differently engaged when processing moves that were probable vs. accurately predicted. Error bars denote 95% confidence interval. \*  $p < .05$ , ~  $p < .11$ .

We then related move probability and prediction accuracy to brain activity at retrieval, using an approach analogous to the retrieval eye-movement analysis in the previous Chapter. Specifically, we predicted univariate activity when participants were recalling a specific move from the probability of the move and the prediction accuracy of the move (during encoding)

(figure 3.5a, top). This generated partially overlapping but largely separated maps, such that recalling a probable move engaged the dorsal attention network and recalling an accurately predicted move engaged AG (figure 3.5a, bottom). We next looked at retrieval eye movements as they indexed the strategy potentially used during retrieval. By predicting retrieval eye movements from a move probability distribution and a correct move distribution (figure 3.5b, top), we computed a measure of the extent to which people were using their schema and their episodic memory when recalling a move. Here, we see that when the participants were looking at the correct move, indicating retrieval from episodic memory, the dorsal attention network and

retrosplenial cortex show higher activation.

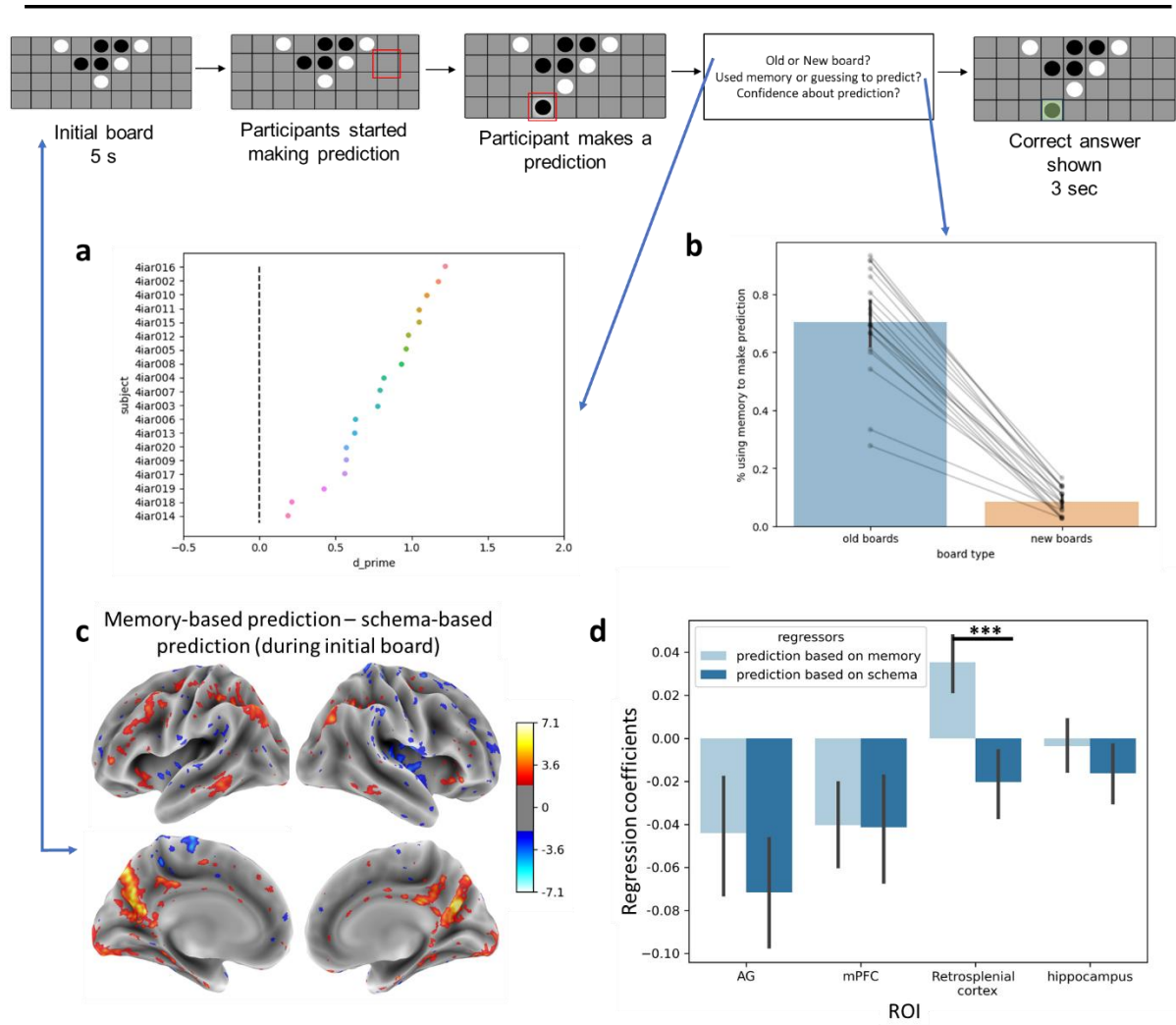


**Figure 3.5 Analysis of retrieval activity.** **a.** Illustration of the analysis conducted. We looked at the move probability and prediction accuracy of the stimuli (a particular move) during encoding, and use them to predict the univariate activity in the brain when participants were recalling the move. **b.** Brain maps showing the coefficient of move probability and prediction accuracy on retrieval univariate activity. When recalling a more probable move (left), higher activities were shown most dominantly in the dorsal attention network, whereas when recalling predicted moves (right), higher activities are shown in AG ( $p < 0.05$  uncorrected). The map in the

middle shows the overlapping (green) and unique (yellow and blue) regions that respond positively to high move probability and prediction accuracy during retrieval. **c.** Analysis of retrieval eye-movement and its correlate to the brain. We looked at the time spent looking at empty squares during retrieval period, and derive two potential strategies that could be used: looking at probable moves to try to cue a memory, and looking at the correct move directly. The usage of each strategy can be calculated by fitting a regression model to predict retrieval eye-movement on empty squares from move probability distribution and the correct move distribution (1 on the correct move and 0s everywhere else). Higher coefficients of the correct move regressor suggest greater use of episodic memory to directly retrieve the move, and was associated with higher univariate activity in the dorsal attention network and retrosplenial cortex (bottom).

Finally, we looked at the data from the prediction task. We first validated the task by showing that participants could correctly distinguish old boards from new boards, indexed by the  $d$ -prime when answering the question about whether the board is old or new (figure 3.6a). They were following the instructions about using memory and schema to make a prediction, because they were much more likely to say they used memory to make a prediction when the board is objectively old (figure 3.6b). We then compared the univariate activity when participants used memory to make a prediction and when participants used schema to make a prediction. Note that this is done by constructing two regressors, one showing whether memory is used, and the other whether schema is used. If participants indicated that they used both schema and memory, both regressor would have a value of 1 on that trial. The contrast between the maps of episodic memory-based prediction and schema-based prediction showed, most significantly, higher activity in retrosplenial cortex when the prediction was episodic memory-based (figure 3.6c). The ROI results further confirmed this finding (figure 3.6d) showing that there is a significant difference between episodic memory-based prediction and schema-based prediction ( $t(18) = 4.77, p < .001$ ). This result is consistent with the previous finding that if participants were looking specifically at the correct move during retrieval, which indicates a more direct retrieval process, they showed higher activity in retrosplenial cortex. In AG and hippocampus, higher

activity was also observed when prediction is based on memory, but the differences between memory- and schema-based predictions were not statistically significant (AG:  $t(18) = 1.38$ ,  $p = .184$ ; hippocampus:  $t(18) = 1.15$ ,  $p = .263$ ).



**Figure 3.6. Behavioral and brain results of the prediction task (top).** **a.** Recognition performance on discriminating old vs. new boards. **b.** The percentage of times participants reported using memory (reporting either using “memory” or “both”) to make a prediction given the boards are objectively new or objectively old. **c.** Contrast in univariate activity when memory was used to make a prediction and when schema is used to make a prediction. Retrosplenial cortex showed the most obvious contrast in memory- vs. schema-based predictions ( $p < .05$ , unthresholded). **d.** ROI version of the analysis, again supporting the searchlight results of retrosplenial cortex showing higher activity for episodic memory-based predictions. \*\*\*  $p < .001$

### 3.4 Discussion

The current study aimed at investigating the neural mechanism of prediction with the 4-in-a-row game, with the subtle differences between different types of prediction in mind. In particular, we were interested in two different distinctions in how memory was used to understand a complex stimulus. The first, extending work in Huang et al. (2025), examined brain responses to stimuli that were probable given the context and, separately, that were accurately predicted before the stimulus showed up. The second, similar to Varga et al., (2025), looked at how the brain makes predictions based on schema vs. episodic memory. We indeed found distinct neural representations underlying both processes.

As discussed in Huang et al. (2025), probability and prediction accuracy have been typically treated as synonymous, and they are indeed highly correlated with each other. In the previous study, we made a specific attempt to separate move probability and prediction accuracy with real-time eye-tracking, by showing moves where participants were not looking, but might still be probable. This revealed that both move probability and prediction accuracy independently contribute to better memory. While the current study did not intentionally separate move probability from prediction accuracy with real-time eye-tracking, different patterns were shown during encoding (and retrieving) moves that are probable vs. predicted. When encoding moves that were probable, AG shows higher activity, whereas when encoding moves that are predicted, mPFC showed higher activity. Both regions are part of the DMN, which are considered to be dealing with internal processes (Menon, 2023), and usually “silenced” when participants are completing a difficult task. However, these regions have also been implicated in encoding memory for naturalistic stimuli (Chen et al., 2017), which are highly structured and could benefit from prior knowledge (Baldassano et al., 2018). The finding that these regions are involved in

the encoding of sequences might suggest that with the help with prior knowledge, encoding the abstract sequences like in the current study became more like encoding a narrative.

Past research has pointed to mPFC as the key region for schema processing (Brod et al., 2015; Masis-Obando et al., 2021; Preston & Eichenbaum, 2013; Raykov et al., 2021; van Kesteren et al., 2012). For example, van Kesteren et al., (2012) showed that for schema-consistent pairs, mPFC showed higher activity during encoding if the pair is subsequently correctly remembered. The current study provides additional insight on the role of mPFC in schema processing. Namely, mPFC might be specifically involved with processing the outcome consistent with schema-based prediction during encoding. On the other hand, AG is more often associated with semantic processing and language comprehension (Seghier, 2013). It makes sense that a move that is more probable would involve the semantic processing in the brain, because instead of remembering the move by its absolute position or nearby landmarks, a probable move can be remembered as, for example, a move that blocks the opponent from winning.

We additionally found that during retrieval, remembering probable and predicted moves also showed different brain patterns. Recalling probable moves is associated with higher activity in the dorsal attention networks, whereas recalling predicted moves involved temporal parietal junction. In Huang et al., (2025), we showed that predicted moves were remembered more directly, which leads to less reliance on schema during retrieval. That is, during retrieval, participants could directly look at the correct move, instead of having to search through different probable moves on the board to potentially help them remember. On the other hand, retrieving probable moves involve more exploration and looking at other probable moves on the board, suggesting a potential use of generate-recognize strategy (Watkins & Gardiner, 1979), where

participants generate potential moves as cues for subsequent recognition. This might explain the activation of the dorsal attention network, which is responsible for visual search (Leonards et al., 2000) when recalling probable moves.

In a recent opinion paper, Ortiz-Tudela et al., (2023) proposed a framework to attempt to categorize at predictions along five non-orthogonal dimensions – flow of information (sequential or recursive), mnemonic origins (episodic or semantic/schematic), specificity (single potential input or multiple inputs), complexity (low-level or high-level), and temporal precision (whether the prediction includes precise timing about when the events would happen). This framework provides a useful way to differentiate different predictive processes that could potentially explain discrepancies in the previous studies of prediction error and memory. That is, prediction error might influence memory in opposite ways depending on the types of predictions being investigated. The second question of interest in the current study was in one of the dimensions proposed by Ortiz-Tudela et al., (2023) – the differences in neural representation of memory- and schema-based predictions.

Most of the past research on prediction has focused on predictions based on episodic memory (Bein et al., 2021; Lee et al., 2021; Poskanzer et al., 2025; Tarder-Stoll et al., 2024), and one common criticism of this approach is that these predictions are just episodic memory retrieval rather than “true” prediction. The current study showed strong activation of retrosplenial cortex when participants report making predictions based on memory, but not when it is based on schema. Interestingly, activity in retrosplenial cortex was also correlated with the extent to which participants look directly at the correct move during sequence retrieval, a potential indication of strong episodic memory. This is consistent in some prior work showing the role of retrosplenial cortex in prediction and episodic memory retrieval (See Alexander et al.,

(2023), for a review) and that damage in retrosplenial cortex is could produce amnesia (Aggleton, 2010). It is also important to note that retrosplenial cortex is strongly connected to the hippocampus, a core memory region in the brain. This potentially suggests that episodic memory-based prediction and schema-based prediction are indeed reliant on different processes, where episodic memory-based prediction is more similar to episodic memory retrieval.

Hippocampus is a core region in the memory system, responsible for the encoding and retrieval of novel information (Scoville & Milner, 1957). In the research of schema and memory, hippocampus was associated with correctly remembering schema-inconsistent information (van Kesteren et al., 2012; Van Kesteren et al., 2013). Consistent with these findings, we showed lower hippocampal activity when encoding moves with higher probability. We found lower activity in the hippocampus when participants made predictions based on schema, but not when predictions are based on memory. However, the difference between the two was not significant. One potential reason for the this might be limited power, since only 19 participants were collected for the study. Additionally, previous studies demonstrating the non-mnemonic functions typically used multivariate analysis, which could be a potential next step for the current study. One study conceptually very similar to the current one is Varga et al. (2025), which looked at hippocampal involvement in detecting prediction error, cleanly showing that the hippocampus is only involved in reacting to prediction error when participants saw a different ending of a video they have seen previously, regardless of whether the ending makes sense given the context or not. Further analysis with the current study could look at the period when participants were shown the correct answer at the prediction task, to see whether hippocampus would be more activated in responding to errors when prediction was based on memory.

The span over multiple timescales is one of the strengths of the idea of prediction as a unifying principle of the brain with different functional specializations. However, this might also raise potential concerns in the studies of prediction, especially in generalization across studies, both in the behavioral domain and neuroimaging domain. For example, a common belief in the field is that surprising events, or events that elicit prediction errors tend to be better remembered, which is supported by some empirical evidence (e.g., Bein et al., 2021; Quent et al., 2022; Wilcocks, 1928). However, some studies have also found the opposite results (Höltje & Mecklinger, 2022; Huang et al., 2023, 2025; Ortiz-Tudela et al., 2021; Poskanzer et al., 2025). These inconsistent findings are partially due to the distinct processes the term prediction might be describing. A study showing that people remembered better a letter that has a color different from other letters in the set (Wilcocks, 1928) can be considered as showing benefit of surprise / prediction error on memory, but it is unclear whether the conclusions that prediction error is good for memory can generalize to different contexts, such as remembering a weird move in a game in a sequence of moves (Huang et al., 2023). Similarly, in the neural domain, a study showing predictive representation in hippocampus when people are navigating a virtual environment (Brown et al., 2016; Tarder-Stoll et al., 2024) does not mean that hippocampus is involved in predictions based on lower-level sensory information or general knowledge of the world. The current study provided evidence for distinct neural mechanisms underlying different types of prediction, supporting the caution needed to generalize findings from work on prediction, and providing a framework for categorizing prediction in future research.

To conclude, the current study provides evidence supporting the idea that the concept of prediction needs to be scrutinized more carefully, and better categorization of what we mean by prediction is necessary for generalizing and understanding this process in the brain. We showed

that probability and prediction accuracy, which are highly correlated and have been considered as synonyms in the literature on schema, are instantiated differently in the brain. Additionally, memory- and schema-based prediction also showed different patterns in the brain.

## **Chapter 4: Binding items to contexts through conjunctive neural representations with the Method of Loci**

*A version of this paper was preprinted as:*

*Huang, J., Manglik, A., Dutra, N., Tarder-Stoll, H., Chamberlain, T., Ajemian, R., ... & Baldassano, C. (2024). Binding items to contexts through conjunctive neural representations with the Method of Loci. bioRxiv, 2024-12.*

## 4.1 Introduction

Decades of research on human memory have sought to probe the functional architecture of the memory system by using memorization tasks with word lists and pictures (Howard & Kahana, 2002; Polyn et al., 2009; Puff, 1979). Paradoxically, participants often struggle to recall these kinds of simple items (Murdock, 1974), while showing impressive ability to remember much more complex stimuli such as movie events (Chen et al., 2017). Stimuli drawn from familiar real-world settings allow us to draw on *schemas*, our prior knowledge about the structure of the world and how events unfold over time. This prior knowledge can scaffold memory processes in various ways. Past behavioral and modeling research has studied how schemas facilitate memory during the encoding process (Bartlett, 1932; Chandra et al., 2025; Chase & Simon, 1973; Gasser & Davachi, 2023; Gobet & Waters, 2003; Huang et al., 2025; Masís-Obando et al., 2024) and how schemas aid memory through providing a scaffold at retrieval (J. R. Anderson, 1981; R. C. Anderson & Pichert, 1978; Huang et al., 2023; Watkins & Gardiner, 1979). For this scaffolding to be effective, the details to be remembered need to be “attached” to the schema; that is, there must be a meaningful relationship between the current episode and schematic knowledge that is formed during encoding and accessible during retrieval. Despite extensive past research, this crucial process of how event-specific details are combined with schemas to form a robust memory representation remains relatively unexplored. The main aim of the current study is to investigate the neural mechanism behind this binding of schemas and event details.

Past research on associative memory has shown that effectively linking items to an externally-presented context requires more than simply experiencing the item in that context; the item must also interact with the context in a meaningful way (Eich, 1985; Murnane et al., 1999;

Shin et al., 2021). This should also be true when the context arises internally through the activation of structured knowledge, but it is difficult to study this process in realistic events. Schemas and event details are often tightly intertwined (e.g., a metal detector is closely linked to an airport schema), and this inherent integration makes it challenging to disentangle the details from the schema and understand how they are combined in memory. Thus, even though recent neuroimaging studies have highlighted the role of the Default Mode Network (DMN) regions in representing schemas (Baldassano et al., 2018; De Soares et al., 2024; Gilboa & Marlatte, 2017; Masís-Obando et al., 2021; van Kesteren et al., 2012; Van Kesteren et al., 2013), the neural mechanisms underlying the interaction between schemas and details in an ongoing event are still unclear.

We hypothesized that when schemas and event details are interactively combined together, the resulting representation should be conjunctive (O'Reilly & Rudy, 2001), meaning that the memory formed is more than a linear sum of its constituent components. For example, when a chess grandmaster sees a chess board, they build a mental representation that is not simply a list of the pieces and their positions, but also includes the relationships between the pieces and the potential dangers and opportunities afforded by these relationships. Some previous work with fMRI has shown signatures of conjunctive representations when perceiving individual objects created from simple features (Erez et al., 2016; Liang et al., 2020), images of human-object interaction (Baldassano et al., 2017), or scenes with different features of environment, objects, and people (van den Honert et al., 2017). While these studies examined conjunctive representations during perception, here we investigated the role of conjunctive representations in linking contexts and items to form a robust episodic memory that can be reinstated through schema-based retrieval.

We developed a new paradigm for studying conjunctive representations based on an ancient mnemonic technique called the Method of Loci (MoL, Figure 1a). The technique involves building and consolidating a spatial layout (memory palace) in the mind, with ordered locations (loci) in the memory palace that serve as a schematic scaffold. During encoding, each item to be remembered is combined with each of the loci in order by forming a meaningful connection between the two, such as imagining an event involving that item occurring at that locus. During retrieval, people mentally retrace their steps through the memory palace in order, using each locus as a cue to recall its associated item. A critical skill for using this technique is the ability to create and elaborate on a relationship between the item and locus, since adding a meaningful interaction is highly effective at improving associative memory (reviewed in (Higbee, 1979) Mnemonists will in fact strategically choose loci with high "associability" (i.e. that have a wide range of features, attributes, and associations) in order to facilitate the creation of these interactions (Bellezza, 1996).

While some past behavioral and neuroimaging research has studied MoL, this work has largely focused on the effectiveness and potential practical applications of the technique (McCabe, 2015; Ondřej, 2025; Qureshi et al., 2014; Reggente et al., 2020; Twomey & Kroneisen, 2021) and the activity pattern while the technique was used and how training changes patterns and connectivity (C. Liu et al., 2022; Maguire et al., 2003; Wagner et al., 2021). In one of the first neuroimaging works on MoL, (Maguire et al., 2003) showed increased hippocampal activity while memory experts remembered information using MoL. More recent research has focused on fMRI functional connectivity differences between experts and novices, finding that connectivity in novices becomes more similar to experts' after training (Wagner et al., 2021). In the current study, we use MoL as a testbed for studying how novel information is encoded into a

well-consolidated schematic map, by measuring the neural representations of each item and locus on their own and then relating these to neural patterns found at encoding (when the item and locus are being bound together) and retrieval.

We recruited participants for a 4-week MoL training program that drastically improved their memory for word lists, and collected fMRI data in one session after week 2 and in two sessions after week 4. In the scanner, participants described their loci and imagined words on the screen. They also used the MoL technique to encode a list of words by combining each word with a locus in their memory palace. Finally, they recalled the list of words by describing each locus, item, and how they were combined together. We used multivariate analyses to investigate how loci and items are represented separately (in locus description and item imagination) and combined (in encoding and retrieval) during MoL. We hypothesized that the activity during encoding and retrieval should be conjunctive, i.e. that it contains information that goes beyond a linear combination of the locus and item. We measured the conjunctive representations in multivariate neural activity patterns (the component of neural activity not predictable from the locus pattern and item pattern) and also assessed the amount of conjunctivity in semantic space (the degree to which the verbal descriptions given by the participants incorporated semantic content beyond that of the locus alone and item alone). We hypothesized that the amount of conjunctive representation in both neural and semantic spaces should increase over the course of the training as participants developed expertise in the technique. Additionally, the two measures of conjunctive representation should be related to each other, with the degree of neural conjunctivity for a locus-item pair predicting the degree of semantic conjunctivity for that pair. Another mechanism in the development of MoL skill might stem from improvements in the ability to bring loci online during encoding and incorporate locus features into the memory trace,

allowing memories to be more easily retrieved through a locus-based strategy. If this is the case, we would expect that locus representations in retrieved memories would become stronger over the course of the training. This hypothesis is also not mutually exclusive with greater conjunctivity: retrieved locus-item representations could be made increasingly robust during training by incorporating both strong locus patterns and conjunctive patterns specific to each locus-item combination.

Past research on conjunctive representations has focused on the role of the hippocampus in forming conjunctive representations to associate two novel, arbitrary, and independent events in memory (Squire et al., 1989). However, forming conjunctive representations of real-life events supported by schemas might require involvement of the cortex. Previous research has proposed that remembering details consistent with a schema relies on cortical regions like medial prefrontal cortex (Brod et al., 2015; Ghosh & Gilboa, 2014; Gilboa & Marlatte, 2017; Z.-X. Liu et al., 2017; Preston & Eichenbaum, 2013; Raykov et al., 2020, 2021; Reagh & Ranganath, 2023; Tse et al., 2007; van Kesteren et al., 2012; Van Kesteren et al., 2013; van Kesteren et al., 2020), potentially because episodic memory can be consolidated rapidly through interaction with the activated associative schema (Morris, 2006). The MoL provides a way to make *any* item schema-consistent, by having participants imagine a situation or find a dimension in which the item is connected to the locus (generating additional details as necessary to reinforce the schematic relationship they identified). In addition, studies have found representations of naturalistic schemas (Baldassano et al., 2018; De Soares et al., 2024; Reagh & Ranganath, 2023) in the DMN. Thus, we hypothesized that we would find conjunctive representations of locus-item pairs in DMN regions, including medial prefrontal cortex (mPFC), which has been shown to be most sensitive to the top-down internal activation of naturalistic schemas (De Soares et al., 2024). In

line with this prediction, we found that conjunctive representations were present through the DMN (and beyond) – memory representations were in fact *dominated* by this conjunctive information rather than pure locus or item representations. The amount of conjunctive representation increased in the novices over the course of training, and was related to the amount of additional creative details added to the story that go beyond descriptions of the locus and item on their own. Overall, these findings point to a crucial role of the conjunctive representation in the DMN for MoL, and the importance of DMN more broadly in forming conjunctive associations that support robust memory.

## **4.2 Methods**

### ***Participants***

26 novice participants passed our initial screening (see below), and were enrolled in a 4-week training program that included three fMRI scanning sessions. One participant was removed from the study for not demonstrating proper use of the MoL during training, resulting in a final sample of N=25 (15 participants were female and 10 were male). The participants' age range from 18 to 48, with a mean age of 26.32 (SD = 7.46). The racial makeup of the participants were 16 White, 7 Asian, 1 Black or African American, and 1 mixed. Four of the participants were Hispanic or Latino. All participants were proficient in English. Participants were compensated \$235 for the completion of the study.

The experimental protocol was approved by the Institutional Review Board of Columbia University (AAAS0252).

### ***Stimuli***

Participants self-generated 40 loci with the guidance of the coach, an expert user of MoL who led the training sessions. The loci were objects in a place familiar to the individual participants (e.g., light switch or kitchen stove in their room). In the first two scans, in the locus task, participants imagined and described these loci; in the encoding task, they used them to remember 40 words. The locus for a given trial was always generated internally by the participant based on their pre-practiced memory palace; they were not presented with the name / image of the loci during any of the fMRI tasks.

Participants also learned a standardized memory palace with 20 loci. The standardized memory palace consisted of five 2D virtual reality environments created in the game engine Unity. For each of the five environments, four distinct locations were selected, resulting in a total of 20 loci arranged in a fixed sequence. The order of the loci was the same for all participants. The standardized memory palace was taught to the participants through video clips created by rotating a virtual camera through the environments and stopping at each locus in the order. Each locus was marked by a number (1 to 20) and an arrow pointing to it, along with a brief written description (e.g., “You turn the corner into the tavern and see an elevated table”, where elevated table is locus 1).

40 concrete nouns, 20 animate and 20 inanimate, were selected as the item stimuli used in all the scans and screening. These words were used in the item and encoding tasks in the first two scans, always appearing in a different random order. A subset of 20 words was selected and used in the item and encoding task in the last scan. In the item task the order of these 20 words was randomized, but in the encoding task, these words were presented to all the participants in the same order.

### ***Screening***

Participants signed up through Columbia's RecruitMe website, and went through a screening procedure, which consisted of a sequence of tasks similar to those used in the main fMRI experiment. They first completed two runs of the Item task, where a word was presented for 10 sec and participants reported whether the word was animate or inanimate during the final 5 sec of each trial. The purpose of including the item task was to better match the protocol used in the scanner, where two item localizers were conducted before completing the encoding task. Participants then completed an encoding task, where each word was presented for 12 sec and they were instructed to remember the words in the right order. After that, participants attempted to retrieve the words one by one, in order. Participants then completed the fMRI safety form and a demographic form. Participants were selected based on availability for the 4-week memory training, and demographics (to ensure diversity in age and gender). After the first cohort of participants, we also selected participants based on memory performance in the screening part of the task (recalling fewer than 20 words in the right order). After participants were selected, they had a virtual meeting with the experimenter to complete the fMRI-related paperwork and schedule the fMRI scans.

### ***Training and scanning schedule***

The experiment is conducted with groups of 3-5 people at a time. In the first two weeks, participants had four 1.5 hour interactive lectures with the coach, where they were trained to use the MoL by creating a personal memory palace with 40 anchors. After the training, they came in for the first fMRI scan (W2 scan). After the fMRI scan, participants went home and completed 10 daily practices. The daily practices consisted of an encoding and retrieval task that lasted 15 minutes each. They also meet with the coach one-on-one for check-in/questions before the next

scan. They then came in for two consecutive days for two fMRI scans (W4D1, W4D2). After the W4D1 scan, they learned a standardized memory palace with 20 loci in it (as described above).

### ***Locus task***

Participants were instructed to describe each locus in detail, and to keep talking until instructed to move to the next locus. Once the scan started, text appeared on the screen for 1 sec instructing the participants to start describing the first locus in their memory palace, then a fixation cross was shown on the center of the screen for 10 sec while participants gave their verbal description. After that, text appeared on the center of the screen for 1 sec instructing the participants to move to the next locus. This was repeated until all of the 40 or 20 loci were described.

### ***Item task***

Participants were instructed to vividly imagine each word shown to them on the screen and make a judgment of whether the word is animate or inanimate. They were also instructed to not try to remember the words. Each word was shown to the participants for 10 sec, and 5 sec after the word was presented a text prompt appeared under the word asking participants to press button “1” if the word is animate and “2” if it is inanimate. After 10 sec, a fixation cross appeared on the center of the screen for 1 sec, followed by the next word.

### ***Encoding task***

Participants were instructed to remember a list of words in order by using MoL. Each word was shown on the screen for 12 sec, followed by a fixation cross of 1 sec after each word.

### ***Retrieval task***

After the scan started, participants were shown a fixation cross on the center of the screen. They were given unlimited time to recall all the words in order. They were told to talk

about each locus and the item in detail, as well as how they associated the locus to the item. The recall was self-paced, but they were encouraged to spend at least 10 sec on each locus-item pair, even if they could not recall the item associated with a locus.

### ***Standardized memory palace learning and review***

After the W4D1 scan, participants learned the 20 standardized loci by watching the videos of the standardized memory palace loci six times. In the first two repetitions (loci learning), participants were introduced to the loci one by one and had unlimited time to study each locus. They were told to press a button when they were ready to move on to the next locus. In the subsequent four repetitions (loci generation), participants viewed the video clips and text descriptions again but were prompted to type the name of the upcoming locus (e.g., elevated table) before advancing to the next one. After the six repetitions, participants were told to recall the names of all twenty loci in order. Before the W4D2 scan, participants reviewed the standardized memory palace by completing three rounds of loci generation. The experimenter then asked participants to verbally describe the 20 loci in order to ensure they could report all 20 without errors.

### ***Behavioral processing***

For the locus and retrieval tasks, we used Open AI's speech-to-text "Whisper" API (Radford et al., 2023) to obtain a transcript. Using the transcript and the audio recordings, we identified the locus described in each 10-sec time window in the locus task. For the retrieval task, we identified the locus-item pair (or in case participants forgot the word, we identified the locus spoken) and the start and end time when participants were describing each pair.

Participants wrote down their list of loci in order prior to the scan. Because participants were relatively new to the technique and their memory palace, they sometimes made mistakes in

transitioning between their loci during the tasks. The most common was to skip a locus in either the locus, encoding, or retrieval task, and participants also occasionally misordered loci. To accommodate these errors, we used results from the retrieval task to (retrospectively) determine what locus was used at encoding. For example, consider the scenario where a participant's loci were swing-grass-slide, and they were asked to remember the words apple-dog-pencil; in this case, if the participant recalled swing-apple, slide-dog, we assumed that they skipped a locus (grass) at encoding and they used slide to remember dog. Consequently, in the subsequent analyses we used the representation of the locus that was recalled (slide) to predict encoding/retrieval representation, rather than the locus that the participant should have used (grass).

### ***Performance scoring***

To score participants' performance in the memory task (which required participants to recall in the correct order), we identified the words recalled in the order they were spoken (repeated words were removed) and found the serial positions of these words from the encoding list. The word was considered to be recalled in the correct order if the serial position of the word was larger than the serial position of the previously recalled word.

### ***MRI Acquisition***

Whole-brain data were acquired on a 3 Tesla Siemens Magnetom Prisma scanner equipped with a 64-channel head coil at Columbia University. Whole-brain, high-resolution (1.0 mm iso) T1 structural scans were acquired with a magnetization-prepared rapid acquisition gradient-echo sequence (MPRAGE) at the beginning of the scan session. Functional measurements were collected using a multiband echo-planar imaging (EPI) sequence (repetition

time = 1.5s, echo time = 30ms, in-plane acceleration factor = 2, multiband acceleration factor = 3, voxel size = 2mm iso). Sixty-nine oblique axial slices were obtained in an interleaved order. All slices were tilted approximately -20 degrees relative to the AC-PC line.

There were 6 functional runs in each scan: two runs of the locus task, two runs of item task, and one run of encoding task, and one run of retrieval task.

### ***fMRI preprocessing***

Results included in this manuscript come from preprocessing performed using fMRIPrep 23.0.2 (Esteban et al., 2018, 2019)RRID:SCR\_016216), which is based on Nipype 1.8.6 (Esteban et al., 2022; Gorgolewski et al., 2011)); RRID:SCR\_002502).

### ***Anatomical data preprocessing***

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008), RRID:SCR\_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 6.0.5.1:57b01774, RRID:SCR\_002823,(Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.3.2, RRID:SCR\_001847, (Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR\_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through

nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization and accessed with TemplateFlow (23.0.0, (Cicic et al., 2022): ICBM 152 Nonlinear Asymmetrical template version 2009c [(Fonov et al., 2009), RRID:SCR\_008796; TemplateFlow ID: MNI152NLin2009cAsym].

### ***Preprocessing of B0 inhomogeneity mappings***

A total of 3 fieldmaps were found available within the input BIDS structure for this particular subject. A deformation field to correct for susceptibility distortions was estimated based on fMRIPrep's fieldmap-less approach. The deformation field is that resulting from co-registering the EPI reference to the same-subject T1w-reference with its intensity inverted (S. Wang et al., 2017; Huntenburg, 2014). Registration is performed with antsRegistration (ANTs 2.3.3), and the process regularized by constraining deformation to be nonzero only along the phase-encoding direction, and modulated with an average fieldmap template (Treiber et al., 2016).

### ***Functional data preprocessing***

For each of the 18 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1:57b01774, (Jenkinson et al., 2002)). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were

mapped on to the reference EPI using the transform. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (D. N. Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, (Power et al., 2014)) and Jenkinson (relative root mean square displacement between affines, (Jenkinson et al., 2002)). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by (Power et al., 2014)). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, (Behzadi et al., 2007)). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's `aseg` segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the  $k$  components

with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by (Patriat et al., 2017). The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): fsaverage6. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using mri\_vol2surf (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.9.1 ((Abraham et al., 2014), RRID:SCR\_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

### ***ROI and searchlight definition***

We used ROIs in the default mode network previously found to be responsive to schematic content (Baldassano et al., 2018): angular gyrus (1868 vertices), medial prefrontal cortex (mPFC; 2069 vertices), and posterior medial cortex (PMC; 2495 vertices). These ROIs were originally derived from a resting-state network atlas on the fsaverage6 surface (Thomas Yeo et al., 2011).

Searchlight ROIs were defined as circular regions on the cortical surface, by identifying all vertices within 11 edges of a center vertex along the fsaverage6 mesh. Since the average edge length between vertices is 1.4mm, searchlights had a radius of approximately 15mm. We defined a circular searchlight around every vertex on a hemisphere, and then iteratively removed the most redundant searchlights (i.e. those whose vertices were covered by the most other searchlights). We stopped removing searchlights when doing so would cause some vertices to be covered by fewer than six searchlights. This yielded approximately 1000 searchlights on each hemisphere.

### ***Activity pattern extraction***

After fMRIPrep, the data (now in fsaverage6 and MNI152 space) were further preprocessed by a custom python script that removed from the data (via linear regression) any variance related to the six degrees of freedom motion correction estimate and their derivatives, mean signals in the CSF and white matter, motion outlier timepoints (defined above), and a cosine basis set for high-pass filtering w/ 0.008 Hz (125s) cut-off. We then z scored each run to have zero mean and standard deviation of 1. All subsequent analyses, described below, were performed using custom python and R scripts.

To obtain the locus, item and encoding patterns, we conducted a GLM predicting whole-brain univariate activity from the design matrices for the corresponding tasks based on the timing of each locus/item. For the retrieval task, the design matrix was based on the manual identification of the start and end time of describing each locus-item pair. The coefficients from fitting this GLM were used as the values for defining the voxel patterns of the locus/item/pair.

***Pattern similarity analyses: Locus and item representations (Figure 2)***

To look at where loci were represented in the brain, for the loci that were described in both locus runs, we correlated the representations of the pairs of loci between the two runs for each participant in each session. This produced a correlation matrix of the representational similarity between each locus in run 1 with each locus in run 2 for each searchlight. We then computed the representational similarity of the same loci across two runs (the average of the diagonal of the correlation matrix). For assessing statistical significance, the similarity between different loci (the mean off-diagonal of the correlation matrix) was then subtracted from the mean similarity of the same loci. Once we obtained the difference, we randomly shuffled the rows of the correlation matrix 1000 times to compute a null distribution of this diagonal vs off-diagonal difference, and computed the z-score for the searchlight

$$z = \frac{\text{difference}_{\text{true}} - \mu(\text{difference}_{\text{shuffled}})}{\sigma(\text{difference}_{\text{shuffled}})}$$

which represents the degree to which there was a locus-specific representation across runs. A z value was computed for each surface vertex as the average of z values from all searchlights that included that vertex. We then combined maps across all subjects and all sessions, and ran a one-sample t-test of the z values against 0. The r values in the voxels with  $q <$

.05 after false discovery rate (FDR) correction from the one-sample t-test were plotted in the map.

For the locus-encoding and item-encoding similarity analysis, the approach was the same as described above. Here, we looked at the similarity between the average of representations of the two locus/item runs and the encoding representation where the locus/item is used. If participants only described a locus in one of the locus runs, then the representation of the locus from the single run was used.

### ***Relating the similarity of encoding residuals to the similarity of stories (Figure 3)***

First, for each locus-item pair on week 4 day 2 (separately for each participant and each brain region), we obtained the residual of the encoding representation by regressing out the representations of the locus and item that were used during encoding. Then, separately for each brain region, we generated a *neural correlation matrix* for each item-locus pair by computing the Pearson's correlation of each participant's encoding residual with each other participant's encoding residual for that pair. Next, we took the verbal stories that were generated on week 4 day 2 for each locus-item pair, and we generated a *semantic correlation matrix* for that pair using the cross-encoder version of the Sentence-BERT language model (Reimers & Gurevych, 2019) takes two text passages as input and generates a score (0 - 1) quantifying the semantic similarity between the two passages. Then the lower triangles of the neural and semantic correlation matrices for all the locus-item pairs were each flattened into a long vector. Finally, a Spearman correlation was computed between the neural and semantic correlation vector for each pair. A permutation test was conducted, shuffling the semantic vectors between locus-item pairs 1000 times to obtain a chance level neural-semantic correlation. We averaged across the 20 locus-item pairs in each permutation, generating 1000 null correlations. For each ROI, the significance level

was determined by counting the percentage of times that the 1000 random correlations were more extreme than the true correlation.

***Identifying conjunctive representations by predicting retrieval representations (Figure 4)***

For each item-locus pair (separately for each participant and searchlight), we ran a GLM predicting the representation of that locus-item pair at retrieval from the locus by itself, the item by itself, and the encoding residual representation for the locus-item pair; this analysis incorporated data from all of the sessions. To make sure the weights reflected item-specific reinstatement rather than general task-related patterns, we adjusted the weights against a null distribution in which retrieval of each locus-item pair was predicted by the locus, item, and encoding residual representations for a different locus-item pair. The labels for the retrieval representations within a session were shuffled 100 times. For each locus-item pair, we calculated the adjusted weight,  $\beta_{task}$  :

$$\beta_{task} = \frac{\beta_{true} - \mu(\beta_{shuffled})}{\sigma(\beta_{shuffled})}$$

The weights were averaged across all the pairs from each session for each participant for each searchlight. A  $\beta$  value was computed for each surface vertex as the average of  $\beta$  values from all searchlights that included that vertex. For each vertex, a one-sample t-test compared the participant-level average weights (across all three sessions) against 0. To compare the difference in amount of representation between encoding residual and locus/item, we conducted a paired t-test on the difference in average weights of encoding residual and locus/item within participants in each session in each participant. FDR correction was used to identify regions with  $q < .05$  for the brain map.

We also ran the above analysis for the three ROIs; here, we compared  $\beta_{task}$  against 0 for each ROI for each timepoint (week 2 or 4) using a one-sample t-test. For across-region

comparisons, we conducted paired t-tests between each pair of regions, combining week 2 and week 4 data.

For looking at differences between week 2 and 4 in the amount of conjunctive representation, we conducted a linear mixed effect regression. We controlled for speaking duration as a nuisance regressor to ensure that differences between weeks were not being driven solely by longer recall durations for each item (which provide more timepoints for estimating retrieval patterns and therefore less-noisy representations). We used the following formula:

$$\beta_{encoding\ residual} \sim \text{training stage} + \text{speaking duration} + (1|\text{subject id})$$

### ***Semantic conjunctive representation (Figure 5)***

We conceptualized semantic conjunctive representation as the amount of additional details added to the locus and item pairs. To measure this, we computed *story deviation*, the semantic distance between the locus-item pair on its own and the whole story, using the same cross-encoder model described above. After we obtained the similarity score, we subtract the score from 1 to generate the story deviation measure. To look at changes in story deviation, we conducted a linear mixed effect regression with the following formula:

$$\text{Story deviation} \sim \text{training stage} + (1|\text{subject id})$$

For each ROI, we looked at how univariate activity during encoding is related to story deviation, with a linear mixed effects regression:

$$\text{Story deviation} \sim \text{univariate activity} + (1|\text{subject id}) + (1|\text{session})$$

We also looked at how weights of encoding residuals are related to story deviation, with a linear mixed effects regression:

$$\text{Story deviation} \sim \beta_{encoding\ residual} + \text{speaking duration} + (1|\text{subject\_id}) + (1|\text{session})$$

### ***Predicting memory from neural representations***

To predict recall success from encoding residuals, we use the following formula:

$$\text{Recall correct} \sim \beta_{\text{encoding residual}} + (1|\text{subject id}) + (1|\text{item})$$

To predict recall success from item-encoding correlation and locus-encoding correlation, we use the following formula:

$$\text{Recall correct} \sim r(\text{encoding representation, locus representation}) + r(\text{encoding representation, item representation}) + (1|\text{subject\_id}) + (1|\text{item})$$

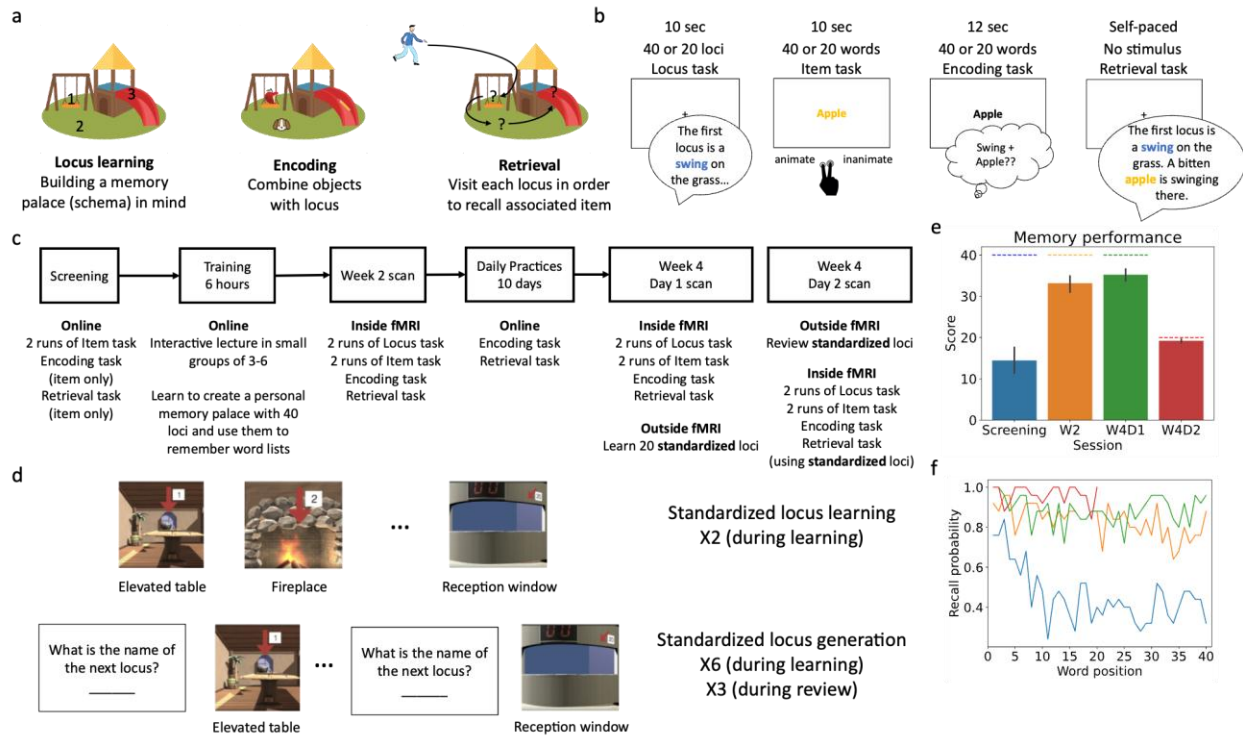
### **4.3 Results**

An overview of the four-week study paradigm is presented in Figure 1c. After passing an online screening, novice participants (N = 25) first participated in an online training program for 2 weeks, in which they were taught the MoL technique and created a memory palace consisting of 40 loci. During the training program, participants were also familiarized with the 40 items (20 animate, 20 inanimate) that they would be associating with the loci in subsequent phases of the study. Participants were given their first fMRI scan in week 2 (W2), which consisted of four types of tasks (Figure 1b). In the locus task (repeated twice), participants verbally described each of their loci for 10 sec, before being prompted to describe the next locus. In the item task (repeated twice), participants saw a word for 10 sec and were asked to imagine the word vividly and then make a judgment of whether or not the word was a living object. In the encoding task, participants used MoL to encode each word for 12 sec, by forming an association between the item and the locus in the memory palace. In the retrieval task, participants described each locus in their memory palace, the item that the locus was connected to, and the story they created to link the locus to the item. The recall was self-paced, but they were encouraged to spend at least

10 sec for each pair. After the first scan, participants completed 10 daily online practices for two more weeks, in which they were presented with 40 words to remember using MoL, each for 12 sec, and then attempted to recall the words in the correct order. In week 4, they were scanned on two consecutive days.

On week 4 day 1 (W4D1), they completed a scan just like the week 2 scan. After that, they were taught (outside of the scanner) a new 20-locus memory palace (Figure 1d), presented as videos on a laptop. We call the loci in the new memory palace “standardized” loci, because this set of 20 loci was the same for all participants (in contrast to the 40 idiosyncratic loci developed individually by each participant, used in the previous encoding sessions). On week 4 day 2 (W4D2), they reviewed the standardized loci and then completed a scan with the same structure as the previous two, now using the standardized loci to remember 20 words (a subset of the 40 words used previously). Note that the same 20 words were presented in the same order to all the participants, so they were each attempting to create a binding between the same set of item-locus pairs, allowing us to make comparisons across participants. Participants’ performance in the memory task showed a massive improvement compared to the baseline (Figure 1e), from 14.44 (SD = 7.93) at baseline to 33.16 (SD = 5.07) to 35.24 (SD = 3.72) in Week 2 and Week 4 respectively. Participants were also near ceiling on the standardized memory task, with a mean score of 19.2 (SD = 1.02) out of 20 words in total. Participants’ serial position curves were markedly different after training (Figure 1f), shifting from showing a standard primacy effect (Murdock, 1974) to near-uniform recall of words from all serial positions. It should be noted that over the course of the study, participants were exposed to the same set of words 3 times (twice in item task, once in encoding task) in each session, which might have contributed to the

performance improvement. Therefore, caution should be taken with regard to interpreting the change in performance over time.

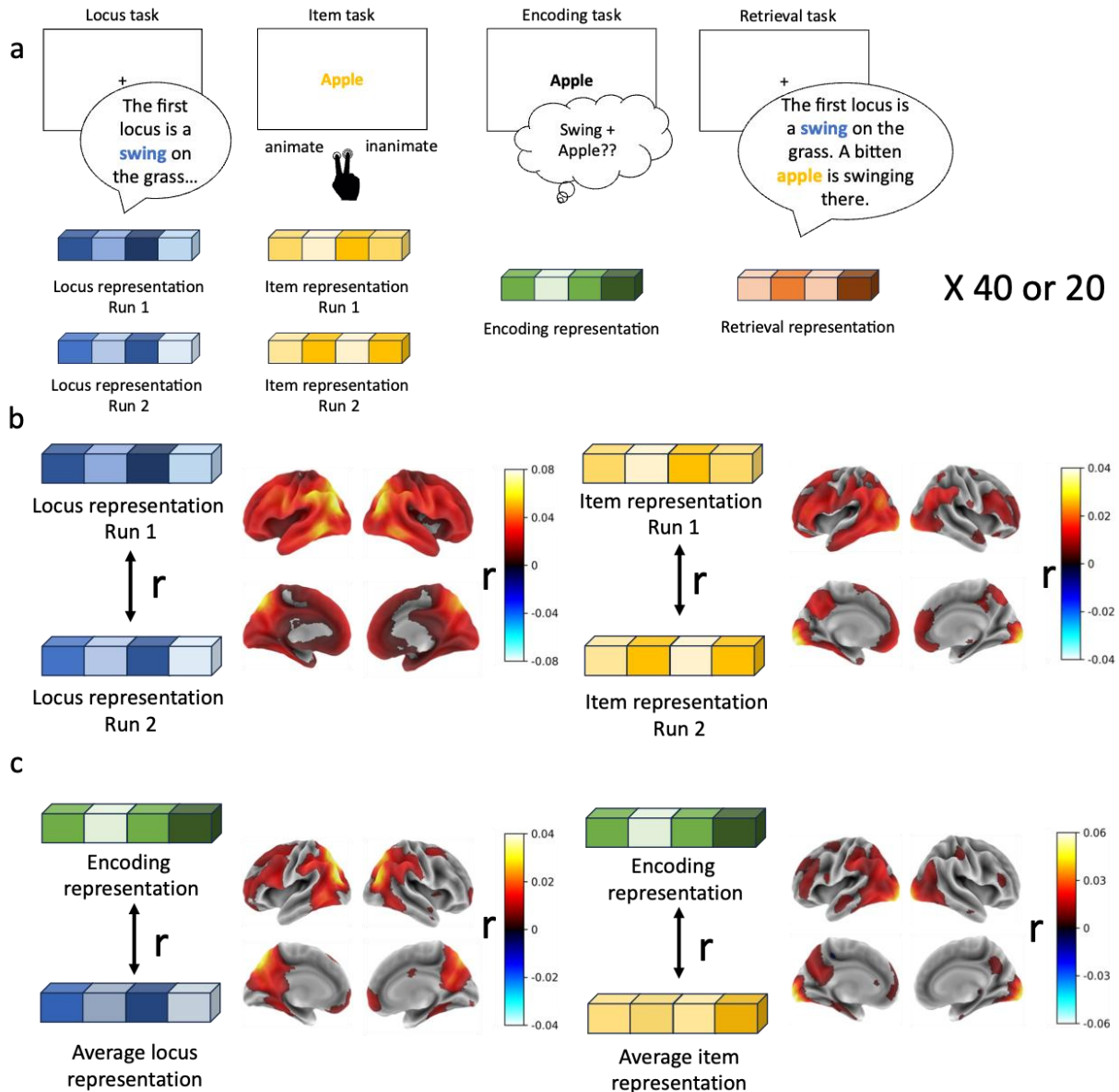


**Figure 4.1 Illustration of the paradigm and behavioral performance.** **a.** Demonstration of the Method of Loci, in which items to be remembered are attached to sequential locations along an imagined spatial map. **b.** The tasks participants completed in the scanner, capturing: neural representations of loci alone, items alone, encoding an item at each locus, and retrieving an item at each locus. **c.** Illustration of the four-week training and data collection schedule for participants. **d.** Illustration of how participants learned the standardized memory palace with 20 loci for the W4D2 scan. All participants first watched videos moving through a 3D environment with locations marked and a label for each locus provided. Participants then learned by generating the name of the next locus after watching videos of the previous locus. **e.** Memory performance (number of words correctly recalled in order) for participants at different timepoints during the study. Dashed lines represent the maximum score possible in each session. Error bars represent the 95% confidence interval. Participants demonstrated substantial improvement after receiving two weeks of training, continued to improve at week 4, and were able to generalize to the standardized loci on W4D2 with near-ceiling accuracy. **f.** Probability of recalling a word given the word's position in the encoding list. During screening (blue), participants showed a strong primacy effect (they were much more likely to recall words in the beginning of the list). However, in later sessions (W2: orange, W4D1: green, W4D2: red) this effect was greatly reduced.

### 4.3.1 Widespread representations of locus and item by themselves and during encoding

Using a generalized linear model (GLM), we obtained the representation of each locus and item alone in each run of the locus and item task and the representation of the locus-item pair during encoding and retrieval (Figure 2a). We then measured the pattern correlation (within searchlights on the cortical surface) between corresponding loci or items in the two runs of the locus and item tasks (while accounting for overall task similarity not specific to particular loci or items; see Methods). For similarity between locus/item and encoding, we used a similar approach, measuring the pattern correlation between the locus by itself (e.g., “swing”) or item by itself (e.g., “apple”) with the locus-item (e.g., “swing-apple”) pair where the locus/item was used.

We showed that a large portion of the brain represented imagined loci during encoding, most notably in angular gyrus (AG) and posterior medial cortex (PMC) (Figure 2b, left). The same analysis for the item task (during which each item was presented as written word) showed item representation in visual cortex, AG, PMC, and mPFC (Figure 2b, right). Because using MoL requires a combination of locus and item during encoding, we looked for representations of locus and item information during encoding (Figure 2c). Both loci and items were represented in large scale brain networks during encoding, with some overlapping regions. Locus reactivation was observed in the default mode network regions, including AG, PMC, and mPFC, while item information was represented primarily in the visual cortex and PMC. These results demonstrate that participants were successfully reinstating locus-specific patterns throughout a broad network of regions during encoding, while simultaneously maintaining a representation of the presented item to be remembered.



**Figure 4.2. Brain regions representing loci and items alone and during encoding. a.**

Illustration of how representations were obtained. For each searchlight on the cortical surface, the multivariate activity pattern was measured for each locus (in each two runs), item (in each of two runs), and locus-item pair during encoding and retrieval of 40 words (in the first two sessions) or 20 words (in the final session). **b.** Representation of locus and item in the brain. Almost the whole brain showed significant locus-specific activation (pattern similarity for corresponding loci in the two locus runs), with the strongest effects in angular gyrus (AG) and posterior medial cortex (PMC). For two item runs, visual cortex, AG, PMC, and mPFC all showed item-specific activation patterns. **c.** Representation of locus and item during encoding. The current (imagined) locus was represented in AG, PMC, and mPFC during encoding, while items (presented as words) were represented in visual cortex and PMC. The brain maps were thresholded based on results of a permutation test, at  $q < .05$  with FDR correction.

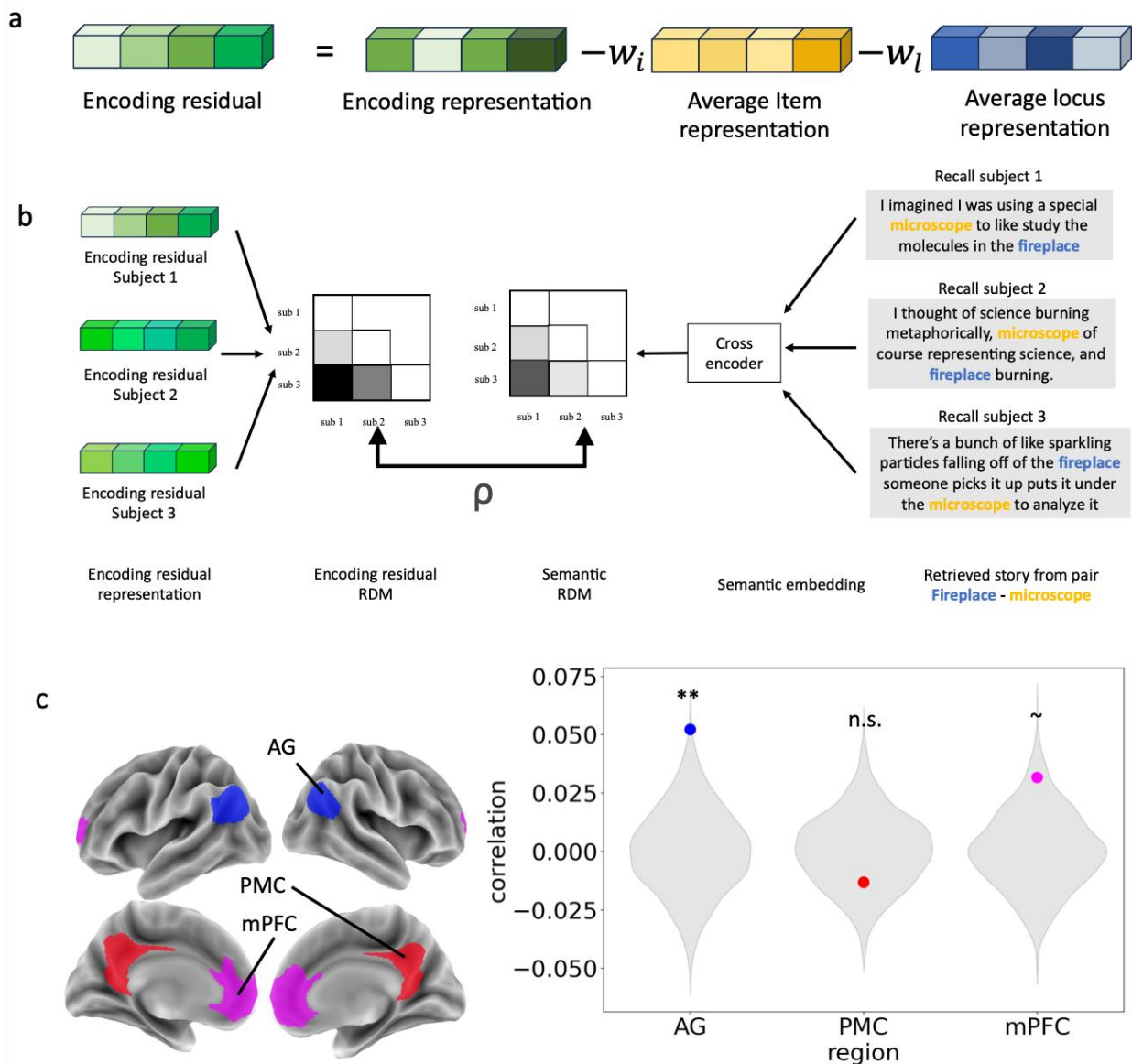
### 4.3.2 Encoding residuals track idiosyncratic semantic combinations of loci and items

When people use MoL to remember a word, in addition to thinking simultaneously about the item and the locus, they also add in additional details to forge a link between the two, creating a conjunctive representation between the locus and the item. We therefore sought to measure the conjunctive component of the encoding and retrieval representations by decomposing them into a representation of the locus, the item, and the conjunction between the locus and the item. For each locus-item pair (separately for each participant and brain region), we regressed out the representation of locus and item from encoding (Figure 3a). Note that the locus representation was derived from the task in which participants sequentially moved through their locus sequence just like in the encoding task. Therefore, if the representation at a specific locus has been reconfigured through repeated practice or has incorporated features of recent or upcoming loci (Tarder-Stoll et al., 2024) these changes would still be captured as part of the “isolated locus” pattern in this regression. The resulting encoding residuals contain the conjunctive information (the pattern for this locus-item pair that is not linearly related to the isolated locus and isolated item patterns) and the representations of irrelevant (e.g. off-task) mental state and measurement noise not reflective of neural activity. One way to test if the encoding residual contains meaningful conjunctive information (vs noise) is to check whether it is related to the semantic content of participants’ verbal recall. For this analysis, we focused on the W4D2 session, in which participants used the same memory palace to remember the same set of words in the same order but often came up with very different ways of relating a locus to an item (see Figure 3b, rightmost column for examples). In three key regions of interest in the DMN, found in previous studies to represent schematic knowledge (Figure 3c, (Baldassano et al., 2018; Thomas Yeo et al., 2011)), we investigated whether higher similarity of the encoding residual

representations in these regions correlates with semantic similarity in participants' verbal descriptions. Note that this is a across-subject analysis for each object, allowing us to assess whether participants who generated similar stories for an object also have similar neural representations; the across-subject nature of this analysis avoids potential issues related to temporal proximity that can arise when measuring within-run neural representations (Dimsdale-Zucker & Ranganath, 2018). For all the ROI analyses presented below, we also conducted the same analyses in anterior, posterior, and the whole hippocampus and found no significant effects in all of the analyses (Supplementary Figure B.1). Additionally, False Discovery Rate (FDR) correction with Benjamini-Hochberg Procedure was conducted for each analysis. For each locus-item pair, we constructed a between-subject representational similarity matrix of the (neural) encoding residual (Kriegeskorte et al., 2008); we also computed a semantic similarity matrix based on people's verbal recall of the locus-item pair using a cross-encoder with sentence-BERT (Reimers & Gurevych, 2019) (Figure 3b); we then computed the Spearman correlation between the lower triangles of these matrices for each locus-item pair. We also shuffled the participant order for the semantic similarity matrix 1000 times for each locus-item pair, which generated a null distribution of the mean Spearman correlation of the 20 locus-item pairs. The mean of the true Spearman correlation for the 20 locus-item pairs were compared to the null distribution. In AG if two people showed similar neural patterns in the encoding residual, they were also more likely to tell similar stories (Mean  $\rho = .052$ , SD = 0.052,  $z = 2.84$ ,  $p = .004$ ,  $q = .011$ ); The same effect was found in mPFC, but it was only marginally significant (Mean  $\rho = .032$ , SD = 0.102,  $z = 1.83$ ,  $p = .066$ ,  $q = .099$ ). This effect was not found for PMC (Mean  $\rho = -.01$ , SD = 0.078,  $z = -0.85$ ,  $p = .402$ ,  $q = .402$ ) (Figure 4d). This demonstrates that the encoding residual representations in

parts of the DMN indeed track idiosyncratic content used to link a locus to an item, supporting the idea that they contain information about the conjunction of the locus and the item.

Because the W4D2 scan happened the day after the W4D1 scan and made use of overlapping words, a potential concern is that the association formed in the W4D1 scan might create interference when the same word is presented in W4D2. If there is interference, we would expect that during encoding of a word in W4D2 (e.g., fireplace-microscope), the locus used to encode the word in W4D1 (e.g., swing, if swing-microscope) was a W4D1 pair) would be activated. We therefore looked at the correlation between the W4D2 encoding representation and the corresponding W4D1 locus representation (and, for comparison, the W4D2 locus representation). Consistent with Figure 2c (where the most locus reactivation was observed in AG and PMC), we found significant reactivation of the W4D2 locus representation during W4D2 encoding in AG and PMC (AG: Mean  $r = 0.015$ ,  $t(24) = 2.73$ ,  $p = .012$ ,  $q = .018$ ; PMC: mean  $r = 0.013$ ,  $t(24) = 4.27$ ,  $p < .001$ ,  $q = .002$ ), but not in mPFC (mean  $r = 0.004$ ,  $t(24) = 1.08$ ,  $p = .291$ ,  $q = .291$ ). However, we found no evidence of W4D1 locus representation in W4D2 encoding in any regions ( $p > .428$ ) suggesting that any interference effects are small.



**Figure 4.3 Encoding residuals track semantic similarity across stories.** **a.** Illustration of how the encoding residual is computed. Locus and item representations were removed from the encoding representation via linear regression, and the resulting encoding residual contained the conjunctive representation of the linkage between locus and item. **b.** Illustration of the RSA analyses. The analysis is based on week 4 day 2, when participants used the same memory palace to remember the same words, allowing us to see the idiosyncratic item-in-locus story each person generated. Pairs of recalls from different subjects for the same story were input to a cross-encoder language model, generating a semantic representational similarity matrix quantifying the semantic similarity between each pair of stories. Similarly, the encoding residuals were compared across pairs of participants to create a neural representational similarity matrix. We then looked at the similarity of the neural representational similarity matrix to the semantic representational similarity matrix. **c.** Demonstration of location of the

ROIs on the cortical surface. **d.** Neural-semantic correlation in the three ROIs. The grey violin plot represents the null distribution generated from shuffling the subject correspondence between the two measures. The dots show the true correlation between the neural representational similarity matrices and semantic representational similarity matrices. In AG and mPFC, similar neural representations track similar semantic representations of stories. (n.s. not significant,  $\sim q < .10$ ,  $** q < .01$ )

### **4.3.3 Widespread and robust conjunctive representation in the brain**

The encoding residual contains not just the conjunctive representation, but also irrelevant representations and noise, making it difficult to estimate the strength of the conjunctive representation based solely on the norm of the encoding residual. We can instead look at the extent to which the encoding residual is reinstated at recall: only the conjunctive component of the residual should be reactivated when retrieving the memory using the locus as a cue. For each locus-item pair (separately for each participant and brain region), we ran a regression to predict retrieval representation from the locus representation, item representation, and encoding residual representation (Figure 4a, bottom). To ensure that these regression weights were specific to the locus-item pair (rather than reflecting a generic task-related representation), the weights were adjusted relative to a null distribution created by permuting which retrieval pattern was matched to the item, locus, and encoding residual patterns (see Methods). Importantly, the weights for the encoding residual capture the amount of conjunctive representation for the pair.

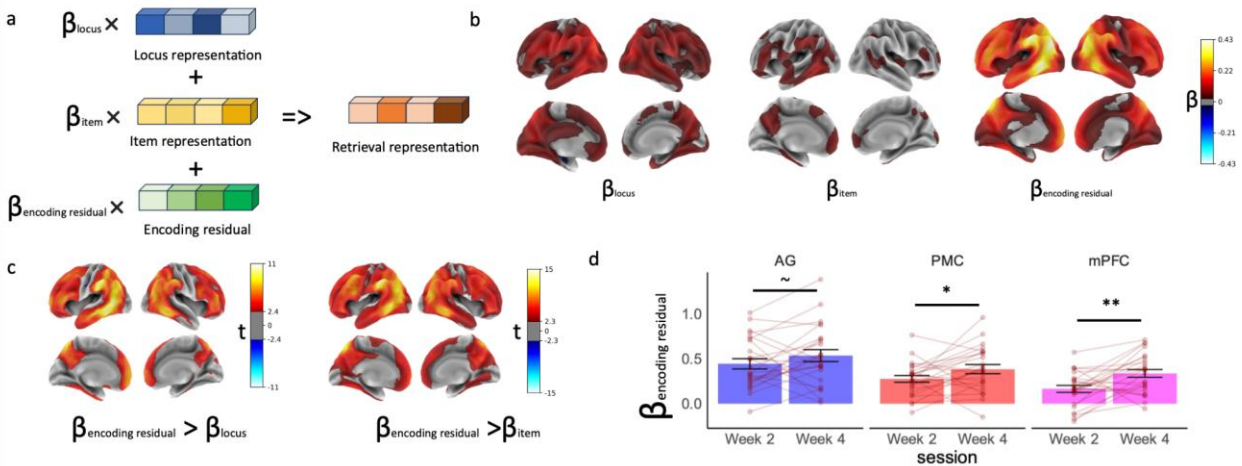
For the locus and item weights in the retrieval regression, we found similar results to the locus-encoding correlation and item-encoding correlation described above, with locus representations in AG, PMC, and item representations in AG, PMC, and mPFC during retrieval (Figure 3b). Strikingly, we found that the encoding residual is represented very strongly in AG, PMC, and mPFC during retrieval, showing the importance of the conjunctive representation during the MoL. We conducted a paired t-test to compare the weight of the encoding residual to the weight of the locus and item, and showed that in largely overlapping regions including AG,

PMC and mPFC, the encoding residual was represented more strongly than the locus or item by themselves (Figure 3c).

#### **4.3.4 Relationship between conjunctive representation in the DMN and training and behavior**

We next explored whether the conjunctive representation was related to experience with MoL and recall behavior. Given the importance of the conjunctive representation to neural representations and the centrality of locus-item binding in the technique of MoL, we would expect differences in the amount of conjunctive representations over the course of training. Comparing each subject's average weight for the encoding residual (across items) against 0, we found that three ROIs (AG, PMC, and mPFC) demonstrated a significant weight of encoding residual in both week 2 and week 4 (all  $p < .001$ ), with a higher weight in AG than mPFC (Mean difference = 0.239, 95% CI = [0.312, 0.166],  $t(49) = 6.58$ ,  $p < .001$ ,  $q = .002$ ) and PMC (Mean difference = 0.160, 95% CI = [0.100, 0.219],  $t(49) = 5.39$ ,  $p < .001$ ,  $q = .002$ ) and a higher weight in PMC than mPFC (Mean difference = 0.080, 95% CI = [0.021, 0.138],  $t(49) = 2.71$ ,  $p = .009$ ,  $q = .009$ ). To look at the change in weight of the encoding residual with experience with the MoL, we used a mixed-effects linear model to predict the residual weight from stage of training (week 2 vs. week 4) while controlling for the duration of recall (to account for the possibility of increased encoding-recall similarity due solely to retrieval pattern estimates being more stable for longer recalls) with a random subject intercept. We found that, across all ROIs, there was an increase in the weight of encoding residual from week 2 to week 4 (Figure 4b), which was significant in PMC and mPFC (PMC:  $\beta = 0.101$ ,  $SE = 0.045$ ,  $t(2145.0) = 2.24$ ,  $p = .025$ ,  $q = .038$ ; mPFC:  $\beta = 0.171$ ,  $SE = .045$ ,  $t(2116.5) = 3.77$ ,  $p < .001$ ,  $q = .003$ ) and marginally significant in AG ( $\beta = 0.087$ ,  $SE = 0.046$ ,  $t(2108.9) = 1.88$ ,  $p = .06$ ,  $q = .06$ ),

providing evidence that representations become more conjunctive with increasing experience in MoL. As a control, we ran similar linear mixed-effect regressions to test whether the contribution of locus or item representations to either encoding or retrieval representations varied with training, and did not find any significant effects of experience (all  $p > .131$ ).



**Figure 4.4 Conjunctive representation in the brain.** **a.** Illustration of how the strength of the conjunctive representation was measured. When using locus, item, and encoding residual to predict retrieval representation, the weight of the encoding residual indicated the amount of conjunctive representation. **b.** The weight for locus, item, and encoding residual when predicting recall patterns in each searchlight. Widespread and robust representations were found for locus and the encoding residual during retrieval, with the strongest locus effects in AG and PMC and the strongest residual (conjunctive) effects in AG, PMC, and mPFC. Item representations were also found in AG, PMC, and mPFC, though with a more limited overall extent of significant effects. **c.** Comparisons between weights for encoding residual and locus (left) and encoding residual and item (right). Encoding residual was more strongly reinstated than either locus or item, especially in AG, PMC, and mPFC. In b and c, color indicates significant difference at  $q < .05$  with FDR correction. **d.** Weight of encoding residual in week 2 and week 4 in the three ROIs. Error bars represent standard error of the mean. Points and connections represent individual participants. All ROIs had significantly positive weights for the encoding residual and the weight of encoding residual increased from week 2 to week 4 in all ROIs ( $\sim q < 0.10$ ,  $* q < .05$ ,  $** q < .01$ ).

Because participants were trained to near-ceiling memory performance even in week 2, we did not have sufficient statistical power to examine subsequent memory effects at the session or participant level. We could examine subsequent memory at a trial level by running a mixed

effect model predicting memory success based on the weight of encoding residuals, which showed that in all DMN ROIs, the amount of neural conjunctive representations is higher for correctly remembered pairs (AG:  $\beta = 3.30$ ,  $SE = 0.59$ ,  $z = 5.63$ ,  $p < .001$ ,  $q = .002$ ; PMC:  $\beta = 2.89$ ,  $SE = 0.93$ ,  $z = 3.07$ ,  $p = .002$ ,  $q = .002$ ; mPFC:  $\beta = 2.90$ ,  $SE = 0.95$ ,  $z = 3.04$ ,  $p = .002$ ,  $q = .002$ ). However, because we quantify the amount of conjunctive representation by measuring how much of the encoding pattern is reactivated during the retrieval period, this finding is difficult to interpret; it is not clear if the low conjunctive representation on failed trials was caused by low conjunctive representation during encoding or a failure to recall the conjunctive representation at retrieval.

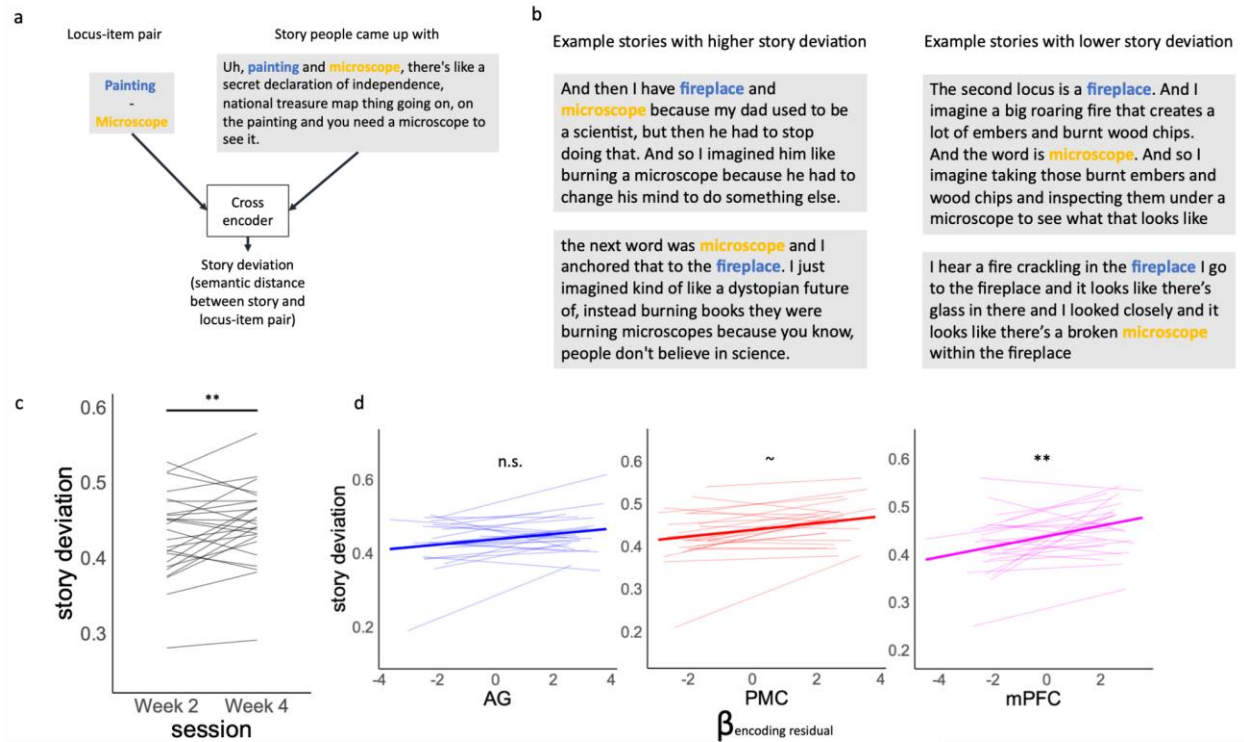
It is possible to run a subsequent memory analysis for locus and item content during encoding, by correlating the encoding representation with the locus and item templates (as in Figure 2c). This correlation provides a measure of how much isolated locus and item content is present at encoding that does not depend on successful retrieval, and can therefore be compared between successful and unsuccessful trials. Running this analysis in our three ROIs, we found that in PMC only, item-encoding similarity significantly predicted whether the pair is later correctly recalled, but it did not survive FDR correction ( $\beta = 1.82$ ,  $SE = 0.80$ ,  $z = 2.27$ ,  $p = .023$ ,  $q = .138$ ). The effect was not found in any other regions or for locus-encoding similarity.

We conceptualized conjunctive representation as the additional details participants added for linking the locus to the item, which in the neural space is measured as the weight of encoding residual, taking into account the locus and item representations. We can define a similar measure for verbal recall, reflecting the extent to which the individual stories generated by each person

added new semantic content not present in the locus or item alone. We quantified this by computing the semantic distance of the story to the locus-item pair using the cross-encoder described above (where the semantic distance is equal to 1 minus the similarity from the cross-encoder) – we refer to this measure as *story deviation*. If this story deviation measure is high, the generated story is more different from just the locus and item pair (Figure 5a), as can be observed in example stories with high and low story deviations (Figure 5b). Comparing week 2 and week 4, we conducted a linear mixed-effects regression predicting story deviation from session, with a random subject and item intercept. We found a significant increase in story deviation across sessions (beta = 0.012, SE = 0.005,  $t(2083.3) = 2.55$ ,  $p = .011$ ) (Figure 5c). Across subjects, increase in story deviation between week 2 and week 4 is positively correlated with increase in the weight of encoding residual in all the ROIs (AG:  $r = .402$ , 95% CI = [0.008, 0.688],  $t(23) = 2.11$ ,  $p = .046$ ,  $q = .046$ ; PMC:  $r = .540$ , 95% CI = [0.184, 0.771],  $t(23) = 3.08$ ,  $p = .005$ ,  $q = .015$ ; mPFC:  $r = .463$ , 95% CI = [0.083, 0.726],  $t(23) = 2.51$ ,  $p = .020$ ,  $q = .03$ ). These results provide converging evidence to suggest that forming a good conjunctive representation is a skill that is associated with training of MoL.

We next looked at whether brain activity measures (univariate activity, encoding residual weights) could predict the amount of the conjunctive representation in the verbal recalls of individual stories (operationalized using the story deviation measure described above). We first conducted mixed-effects linear regressions predicting the story deviation from univariate activity in each of the ROIs with random subject and session intercepts. In all three ROIs, univariate activity predicted higher story deviation values (AG: beta = .051, SE = .014,  $t(2089.8) = 4.29$ ,  $p < .001$ ,  $q = .003$ ; PMC: beta = .058, SE = .02,  $t(2081.2) = 2.84$ ,  $p = .004$ ,  $q = .006$ ; mPFC: beta =

0.055, SE = 0.02,  $t(2082.8) = 2.66$ ,  $p = .008$ ,  $q = .008$ ). Finally, we compared our *neural* conjunctivity measure (the weight of the encoding residual) to our *behavioral* conjunctivity measure (story deviation), conducting a linear mixed-effects regression predicting story deviation from the weight of encoding residual, with a random subject and session intercept (and controlling for duration of recall). We found that in PMC and mPFC, a higher encoding residual weight for a locus-item pair predicted that the generated story for this pair would have a higher story deviation value (PMC:  $\beta = 0.005$ , SE = 0.002,  $t(2088.0) = 2.01$ ,  $p = .036$ ,  $q = .054$ ; mPFC:  $\beta = 0.007$ , SE = 0.002,  $t(2038.0) = 3.06$ ,  $p = .002$ ,  $q = .006$ ), although the effect in PMC did not survive FDR correction. The effect in the same direction was found but not statistically significant in AG ( $\beta = 0.003$ , SE = 0.002,  $t(2078.0) = 1.29$ ,  $p = .198$ ,  $q = .198$ ). This analysis provides further evidence that novel patterns that are formed at encoding (not linearly related to locus or item patterns alone), and reinstated at retrieval, reflect the generation of new semantic details bridging between the item and locus.



**Figure 4.5. Measuring the novelty of generated stories and linking this measure to brain activity.** **a.** How story deviation (semantic distance between the story and the locus-item pair) was measured using a cross-encoder language model. **b.** Example stories with high and low story deviation. In a and b, locus is highlighted in blue and item is highlighted in orange. **c.** How story deviation changes from week 2 to week 4. Each line represents a participant. (\*\*  $p < .01$ ) **d.** Correlation between weight of encoding residual for a locus-item pair and story deviation for that pair. In all ROIs, the weight of encoding residual significantly predicted story deviation. Thick blue lines represent the overall trend and thin lines represent individual participants (in all three sessions). (n.s. not significant,  $\sim q < 0.10$ , \*\*  $q < 0.01$ )

## 4.4 Discussion

In the current study, we trained a group of naive participants in the Method of Loci and used fMRI to measure the item-locus memories they formed during 3 sessions over the course of a month. Consistent with prior work (Wagner et al., 2021), all participants showed great improvement in their ability to remember word lists after learning MoL, demonstrating that the technique is highly effective and does not require users to have any exceptional baseline memory abilities. Our results highlight the role played by the mPFC and other DMN regions in

representing schematic knowledge (loci) and items, both by themselves and in combination. Crucially, we found evidence that these regions contain conjunctive representations that track the content of the unique story details generated by each participant. The degree of neural conjunctive representation for locus-item pairs increased with training, and tracked the amount of novel details (going beyond the locus and item on their own) that were described by participants for each locus-item pair. Overall, these results point to a central role of conjunctive coding in the DMN for creating robust associative memories.

#### **4.4.1 A new approach to study conjunctive representation and MoL**

Past research has looked at the integration of components over the course of learning (Fernandez et al., 2023; Ritvo et al., 2019; Schlichting et al., 2015), which is related to but different from the conjunctive representation studied in the current paper. In these prior studies, integration has been defined as increasing representational similarity between the component pieces of an association after making a connection between them. For example, (Milivojevic et al., 2015)) showed that, when participants became aware of a connection between two seemingly unrelated events, the representations of the two events became more similar to each other in mPFC and posterior hippocampus. On the other hand, our analysis of conjunctive representation looks at how the combined representation relates to the individual components; a conjunctive representation requires a new pattern to be added into the combined representation, rather than adjusting the representations of the individual item and locus patterns to be more similar (e.g. by blending their representations in some way). Another important distinction is that participants linked a well consolidated schema (locus in their memory palace) to a relatively novel item, whereas the components being integrated in previous studies were equally unfamiliar to the participants.

Past behavioral research has highlighted the importance of forming a conjunctive representation between novel events and their contexts to create durable memories (Eich, 1985; Murnane et al., 1999; Shin et al., 2021), and past neuroimaging studies have looked at schema representation in the brain in a naturalistic context (Baldassano et al., 2018; Masís-Obando et al., 2021). However, these past studies could not elucidate the neural mechanisms underlying the creation of conjunctive memories for naturalistic events, because it is difficult to disentangle schematic information from event-specific details in movie or story stimuli. In the domain of perception, where the separate representation of different objects can be assessed individually and combined, evidence for conjunctive representation has indeed been found, yet the impact of such conjunctive representations on memory was not tested (Baldassano et al., 2017; Erez et al., 2016; Liang et al., 2020). Our approach addressed these two concerns, and allowed us to not only look at schemas separately from the novel event details, but also how they were combined together to form a durable memory.

This study also looked at the MoL in ways that had not been previously done in the literature. While anecdotally MoL involves creating narratives for combining a locus and item, the nature of these connections can vary substantially across individuals and items (as shown in Figure 5b) and past behavioral work has not quantified the content of individual locus-item connections. Similarly, past neuroimaging work shows that teaching people MoL changes connectivity and activation patterns (C. Liu et al., 2022; Maguire et al., 2003; Wagner et al., 2021), but not how these narratives are created. Our approach opens up possibilities of looking at strategies during MoL in more subtle ways. The neural and semantic conjunctivity measures developed in the current study allowed us to show, for the first time, the importance of making conjunctive representations in MoL. Our findings suggest that optimizing memory for a locus-

word pair involves generating and encoding extra details; counterintuitively, these details are semantically distinct from both the locus and word, rather than simply reinforcing features of the to-be-remembered word. Future modeling work can investigate the role of these novel details in creating durable memories; we hypothesize that these details are generated in order to create a context in which the item has a strong relationship to the locus, granting the memory benefits for schema-consistent associations (Bein et al., 2015, p. 202; Brod & Shing, 2019; Buuren et al., 2014; Huang et al., 2025; Quent et al., 2022) to arbitrary locus-item pairs. We did not find evidence that incorporating locus features more strongly into the encoded memories is a key factor in MoL training, since the proportion of the retrieved memory representation related purely to the locus was small and did not increase from week 2 to week 4.

#### **4.4.2 Conjunctive representations in the DMN**

Conjunctive representation involves combining stimulus elements into a representation that is more than the simple sum of its elements. In the past literature, forming conjunctive representations between arbitrary elements has been suggested to be one of the main functions of the hippocampus (McClelland et al., 1995; Rudy & Sutherland, 1995). In the current study, where people formed conjunctive representations between novel items and a well-established internal schema, we did not find evidence for the involvement of the hippocampus. Instead, we found widespread conjunctive representations in the cortex, particularly the DMN, when people connected episodic information (words) with a well-consolidated schema (their memory palace). We found that these regions, which have been previously implicated in processing schemas and forming memory consistent with schemas (Baldassano et al., 2018; Ben-Yakov & Dudai, 2011; Brod et al., 2015; Masís-Obando et al., 2021; Preston & Eichenbaum, 2013; Raykov et al., 2020,

2021; van Kesteren et al., 2012; Van Kesteren et al., 2013; van Kesteren et al., 2016), play a key role in representing conjunctions between schemas and episodic details.

We found significant amounts of neural conjunctive representation in all three core regions of the DMN: AG, PMC, and mPFC. These regions have been previously implicated in work with memory for naturalistic events tied to familiar contexts (Chen et al., 2017; Zadbood et al., 2017), showing strong reinstatement of encoding patterns during retrieval. Our results suggest that conjunctive representations in these regions may also serve to “glue” schemas to event details when we form memories of naturalistic events (Baldassano et al., 2018; Masís-Obando et al., 2021).

In line with prior work, we found that mPFC played a central role in building schema-based memories, with significant effects across our analyses: the degree of conjunctive representation in mPFC robustly increased between week 2 and week 4; the neural similarity of conjunctive representations across pairs in mPFC tracked the semantic similarity of people’s stories; and the amount of mPFC conjunctive representation for individual pairs was strongly associated with the amount of conjunctive semantic detail in participants' verbal recalls. Past lesion studies (Ghosh et al., 2014) have shown that vmPFC patients with confabulation symptoms had difficulty judging whether a word belonged to a script or not, suggesting that mPFC plays a role in relating current items to prior knowledge. Thus, mPFC may play an especially important role in the item-locus association step of MoL, drawing on prior knowledge to find a plausible interaction between the item and locus and elaborating on this interaction with new integrative details. For example, one participant (Figure 4.5b) linked a knight to a lavender candle by identifying the linking concept of a lavender field that the knight could walk through. This may draw on the same cognitive mechanisms as the "free generation of remote associates"

task, in which participants attempt to generate a word with a meaningful but unusual relationship with a cue word, which is also known to be impaired by vmPFC lesions (Bendetowicz et al., 2018).

On the other hand, we found no evidence for the engagement of hippocampus in forming conjunctive representations. Previous studies showed that, while novel associations consistent with a schema could rapidly become independent of the hippocampus, the encoding and early consolidation of these associations still depends on the hippocampus (Tse et al., 2007). It is possible that semantic representations are present in the hippocampus during this task and our design and scanning protocol were not sensitive enough to detect them. An alternative possibility is that the hippocampal representation is episode-specific in nature, reflecting a unique code for each locus-item pair that does not generalize to semantically-similar pairs (Reagh & Ranganath, 2023).

#### **4.4.3 Relating conjunctive coding to creativity and concept combination**

In the current study, participants had complete freedom in how they chose to bind items to loci, and different participants came up with very different relationships even when provided with the same loci and items in the standardized-loci task. Although we were primarily interested in using MoL as a tool for studying memory, this experimental task could also be useful for research that involves concept combination, such as research in creativity. Creativity paradigms generally involve relatively constrained tasks like identifying a common word connecting two or three seemingly unrelated words (Bowden & Jung-Beeman, 2003) or finding alternative uses for a tool (Beaty et al., 2015), while our paradigm encouraged more open-ended divergent thinking (Runco & Yoruk, 2014) to generate semantic associates of the locus and/or the item. The current study thus provides a framework to study creativity across different semantic domains or across

different individuals in a relatively unconstrained and spontaneous manner. The measurements developed in the study, including both the neural and semantic measure of conjunctive representation, could be relevant for providing insights into the neural mechanisms of creativity. Our finding that the amount of conjunctive representation in mPFC and PMC tracks the amount of semantic elaboration (as indexed by our story deviation measure) provides support for an important role of these regions in the creative process (Aziz-Zadeh et al., 2013; Shamay-Tsoory et al., 2011).

We hypothesize that MoL is so effective because conjunctively combining an item and locus draws on a fundamental cognitive function of building rich concepts from simpler primitives (Lake et al., 2015). While some of previous research has looked at the importance of the DMN in conceptual combination (Frankland & Greene, 2020), past research typically involved studying relatively straightforward combinations (like combining “old” and “woman”) and has focused more on the comprehension and perception of these ideas (Baron & Osherson, 2011). Our results argue for a role of the DMN in combining ideas together in a more elaborate way. While here we focus on how this conjunctive process supports episodic memory, future work can elucidate how conjunctive representations in the brain might support other processes such as concept learning (Zhou et al., 2024), episodic simulation and imagination (Spreng et al., 2018), and semantic-episodic linkage in trivia experts (Thieu et al., 2024).

One limitation of the current study is that, because we wanted to ensure that the novice participants would be able to form associations and that we would be able to accurately measure neural representations with fMRI, we provided 12 seconds for encoding each item. This is longer than most memory experiments with word list learning (e.g., (Murdock, 1974)) and other experiments involving MoL (e.g., (Wagner et al., 2021)), and is not how MoL is used by experts

in a competition context, when an association needs to be formed in ~2 sec. The long encoding duration, in combination with other factors such as the relatively smaller total number of words (compared to that in memory competition), also led to performance being at ceiling even after just two weeks of training, making it infeasible to perform analyses relating conjunctivity to performance. Future research could employ other designs, such as measuring conjunctive representations at a short delay and then testing memory at a long delay, to avoid these issues and allow for additional analyses of subsequent memory. Another direction is to look at whether quickly-formed item-locus associations are supported by the same brain regions as in our slow-encoding design, as well as the meta-cognitive question of how mnemonists assess whether they have spent sufficient time adding conjunctive information for a specific association.

In conclusion, we used MoL as a window into studying the process of how people combine details with schemas in the context of memory formation. We found evidence of strong conjunctive representation in the DMN, which increased with training and tracked the semantic details added to the locus-item pair. These results demonstrate the importance of conjunctive representation in meaningfully combining schema with novel items and shed light on the neural mechanisms of creativity and concept combination.

## Conclusion

In this dissertation, I presented a series of behavioral and neuroimaging experiments to investigate how prior knowledge influences memory. In Chapter 1-3, I looked at memory for “circles appearing on a grid”. In Chapter 4, I studied memory for word lists. These kinds of stimuli have been used extensively in the literature. However, by giving people prior knowledge as a scaffold, these stimuli become “kind of naturalistic”, in that people are able to use their complex knowledge that they have built to remember these simple stimuli in ways that are more similar to how they remember events in real life.

In Chapter 1-3, I taught participants about a board game to study how complex prior knowledge like knowledge of a board game is related to prediction and memory. In Chapter 1, I studied the process by which a schema of the game is developed and subsequently influences prediction and sequence memory. At first, participants did not know the rules of the game and therefore could only use pure episodic memory to remember the sequences. After they learned the rules of the game, they were able to better remember moves that have higher probability according to a gameplay model. Interestingly, eye-tracking revealed that participants also became more likely to look at probable empty squares during sequence encoding. We also showed that making such predictions is good for memory for the upcoming move, regardless of whether the upcoming move is probable or not. The study demonstrated, for the first time to our knowledge, how the development of a schema could influence prediction, and showed that enabling complicated prediction is one potential important mechanism by which schemas influence memory.

In Chapter 2, I proposed that in research on schema, there are in fact two different processes that are highly correlated and hard to separate in experimental paradigms: one where a

prediction is made and the stimulus (outcome) is compared with the prediction; the other where the stimulus is evaluated in terms of its probability given the context. With real-time eye-tracking, I was able to show that these are indeed separable processes that both facilitate memory, but in different ways: accurate prediction facilitates episodic memory that could be retrieved directly, whereas probable moves are remembered in a potentially gist-like fashion, which rely on more schema during retrieval.

In Chapter 3, I used fMRI to delve deeper into what I explored in Chapter 2, to see how what is called prediction in the field could be further broken down into different processes. In Chapter 2, I made distinction between accurate prediction and high probability given the context. In Chapter 3, I showed that these two processes are associated with different patterns of neural representation during encoding and retrieval. Encoding probable moves engage AG, while encoding predicted moves engage mPFC. Retrieving probable moves engage dorsal attention network while retrieving predicted moves engage temporal parietal junction. Additionally, I looked at another dimension of prediction – predictions based on episodic memory vs. predictions based on schema. I found that episodic memory-based prediction is associated with increased neural activity in the retrosplenial cortex, which is connected to the hippocampus and past research has shown that damage to retrosplenial cortex could produce amnesia (Aggleton, 2010). Overall, Chapter 2 and 3 demonstrate that the term prediction indeed incorporates multiple distinct processes that should be looked at more carefully when trying to make generalizations.

In Chapter 4, I looked at an ancient technique called method of loci (MoL), an ancient mnemonic technique. MoL allows people to encode a list of random words in a way more similar to encoding real-life events. In particular, real-life events are highly structured, with a script of

what typically happen in the situation, and an event can be encoded by incorporating the unique details in the episode to the script. With MoL, people use their memory palace to provide a structure such that random words (the unique details) can be combined with loci in the memory palace (the script). I showed that the memory representation formed with this technique is not just a linear combination of the word and the loci, but a conjunctive representation in multiple parts of the DMN that is more than the sum of its parts. One particularly important region is the mPFC, which showed the biggest changes in the amount of conjunctive representation as people became better at the technique, and strongly tracked the amount of additional semantic details added to the story.

While the schemas in 4-in-a-row and MoL seem very different from each other, with one being the rule of an abstract board game, and the other being a sequence of fixed locations in the memory palace, both can be considered as a type of “script” (Schank & Abelson, 2013). A script provides a structure about how an event typically evolve, like a restaurant script that describe the typical event that happen in a restaurant (being seated, viewing the menu, ordering food, food arriving, paying the check). In 4-in-a-row, there is also a script about how the game evolves (forming a 2-in-a-row, forming a 3-in-a-row, then winning with a 4-in-a-row), although it is more complicated and much more flexible. In the MoL, the locations in the memory palace resemble a script in that they follow a specific order that allowed the details to be combined. In both cases, there is a temporal predictive structure, in which we can use our past experience to make (potentially probabilistic) predictions about what will happen next.

These studies provide novel insights in the study of how schemas influence memory and the underlying neural mechanisms. It has theoretical, empirical, and methodological implications and opens the door for exciting new directions for future research in this field.

### *How prior knowledge scaffold memory*

The core question investigated by the current dissertation is how prior knowledge provides a scaffold for memory. This question is by no means new in the field of cognitive psychology, and many mechanisms have been proposed and tested on how prior knowledge scaffolds memory. As reviewed in the general introduction, many studies have showed ways that prior knowledge scaffolds memory, such as by allowing novel information to be represented more meaningfully (Bransford & Johnson, 1972), by chunking (Chase & Simon, 1973), by providing retrieval cues (R. C. Anderson & Pichert, 1978; Watkins & Gardiner, 1979). By taking advantage of the new, “kind of naturalistic” paradigms, the current dissertation provided some additional insights into the role prior knowledge plays in episodic memory. Although predictive representations have been seen in experts (Didierjean & Marmèche, 2005), Chapter 1 was the first study to demonstrate the development of prediction as people acquire a novel schema, as well as how prediction is a mechanisms by which schema could scaffold episodic memory.

Past research has also provided invaluable insight into the neural mechanisms underlying such memory benefits, pointing to the importance of the interaction between the mPFC and hippocampus (Baldassano et al., 2018; De Soares et al., 2024; Van Kesteren et al., 2013). In Chapter 3, I examine the neural mechanisms of this scaffolding process in the context of the 4-in-a-row game. While it was a challenging memory task that is generally associated with activation of the multiple demand network (Duncan, 2010), I found activation of the DMN, particularly in AG and mPFC when the move is probable or predicted, which are generally associated with internal processing and processing of naturalistic stimuli (Baldassano et al., 2018; Chen et al., 2017; Menon, 2023). This suggests that this paradigm indeed sits in between

the traditional experimental paradigm and paradigms with naturalistic stimuli, offering a bridge to understanding the role of prior knowledge in different contexts.

In Chapter 4, I further demonstrated that in DMN the neural representation in schema-facilitated memory is conjunctive, meaning it is more than a simple linear sum of the schema and the specific details, but rather contains additional semantic details that are reinstated during retrieval. This could potentially provide insights into how these DMNs are used in naturalistic context such as movie viewing.

### ***Different types of predictions and effect of prediction on memory***

A popular belief in the field is that prediction is a core function of the brain, and as a result it is a process that has been studied extensively. Because predictions function at so many different levels of the hierarchy in the brain, it is important to think about predictions not as a singular process, but rather a general term to describe processes where prior experience influences current perception through potentially different mechanisms. One implication of this idea is that careful consideration should be taken when interpreting and generalizing results from one study about the impact or neural mechanisms of prediction. Chapter 2 and 3 specifically looked this issue, and show that, for example, distinct processes might be underlying how we encode information that are improbable given the context, or that are (probable but) not accurately predicted. Both could be called prediction errors, but they both independently and differently influence memory for the stimuli, and the brain is differently engaged when encoding and retrieving them. Additionally, Chapter 3 also found differences in schema- vs. memory-based predictions in the brain.

It is also important to consider the ontology of the cognitive processes given these findings – for example, are memory-based predictions just episodic memory retrieval? This is

relevant to, for example, the question of the functional role of the hippocampus, which some believe is restricted to memory. Chapter 3 did not find strong evidence of the involvement of the hippocampus in either memory- or schema-based prediction. While I refrain from over-interpreting null results, studying prediction and different types of prediction could be a way forward into looking at the modularity of the brain.

By considering prediction as an umbrella term that includes many sub-processes, and by creating better categorization for different types of predictive processes, we could really investigate some of the debated questions in the research on prediction. For example, a commonly held belief in the field is that prediction error benefits memory, which has been demonstrated in numerous studies (Bein et al., 2021; Jang et al., 2019; Quent et al., 2022), but the opposite has also been found (Höltje & Mecklinger, 2022; Huang et al., 2023, 2025; Ortiz-Tudela, Nolden, et al., 2023; Poskanzer et al., 2025; Van Kesteren et al., 2013). A better categorization of prediction might provide insights into the condition in which prediction errors benefit vs. harm memory.

### ***Potentials of “kind of naturalistic” paradigms***

The studies presented in the current dissertation used paradigms that sit in the middle of the traditional experimental paradigms that allow maximum control and modellability, and naturalistic paradigms that are more similar to real-life experiences. This approach allows me to take advantage of the efficiency and the level of manipulation of an experimental paradigm – creating situations where people could make simple response for prediction and memory; allowing people to develop a novel, complex schema quickly; modeling schemas easily; manipulating prediction error through real-time eye-tracking. It would take very hard work to create naturalistic stimuli and paradigms that check these boxes. An additional advantage of

these paradigms is that they are in general more fun for the participants. More than 700 hours' worth of data were collected from the participants in the studies conducted in the 4 chapters, and they have all been at least 2 sessions, but the attrition rate has always been low and participants often told me that they had fun in my study.

At the same time, I was able to study processes that are more similar to events happening in real-life – spontaneously making probabilistic predictions about a sequence of events; simultaneously encoding the current stimuli and predicting the future; giving meaning to the stimuli with the help of a schema as they come in; creating stories to help us remember things. These complex and dynamic processes are difficult to study in a lab setting.

Another exciting aspect of this kind of paradigms is that the insights generated from the them could be applicable and tested by making it more experimental or more naturalistic. For example, ongoing work in our lab looks at (1) how people make schema- vs. memory-based predictions from YouTube instruction videos. (2) how people respond to stories that have probable but unpredicted endings. On the other side, another project I am involved in is looking at how mouse movement during event perception of simple images could be an indication of learning of event structures, and how that influences memory.

### ***Concluding remarks***

To conclude, the current dissertation looked at how prior knowledge scaffolds memory with a “kind of naturalistic” paradigm. I show that (1) one way prior knowledge scaffolds memory is by enabling complicated predictions; (2) making prediction, especially accurate ones, are good for memory; (3) prediction is not a singular process and different processes depend on different brain regions; (4) the DMN plays an important role by which prior knowledge scaffolds

memory, potentially through forming a conjunctive representation; (5) these kind of naturalistic paradigms have a lot of potential in answering questions across the board.

## References

- J., Gopnik, A., Griffiths, T. L., Hartshorne, J., Hauser, T. U., Ho, M. K., Leeuw, J. de, Ma, W. J., Murayama, K., Nelson, J. D., Opheusden, B. van, Pouncy, H. T., Rafner, J., Rahwan, I., Rutledge, R., ... Schulz, E. (2023). *Using Games to Understand the Mind*. PsyArXiv. <https://doi.org/10.31234/osf.io/hbsvj>
- Aly, M., & Turk-Browne, N. B. (2016). Attention promotes episodic encoding by stabilizing hippocampal representations. *Proceedings of the National Academy of Sciences*, *113*(4), E420–E429. <https://doi.org/10.1073/pnas.1518931113>
- Aly, M., & Turk-Browne, N. B. (2017). How Hippocampal Memory Shapes, and Is Shaped by, Attention. In D. E. Hannula & MDuff (Eds.), *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition* (pp. 369–403). Springer International Publishing. [https://doi.org/10.1007/978-3-319-50406-3\\_12](https://doi.org/10.1007/978-3-319-50406-3_12)
- Anderson, J. R. (1981). Effects of prior knowledge on memory for new information. *Memory & Cognition*, *9*(3), 237–246. <https://doi.org/10.3758/BF03196958>
- Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*(1), 1–12. [https://doi.org/10.1016/S0022-5371\(78\)90485-1](https://doi.org/10.1016/S0022-5371(78)90485-1)
- Anderson, R. C., Pichert, J. W., & Shirey, L. L. (1983). Effects of the reader's schema at different points in time. *Journal of Educational Psychology*, *75*(2), 271–279. <https://doi.org/10.1037/0022-0663.75.2.271>
- Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougale, S. D., & Norman, K. A. (2021). Behavioral, Physiological, and Neural Signatures of Surprise

- during Naturalistic Sports Viewing. *Neuron*, 109(2), 377-390.e7.  
<https://doi.org/10.1016/j.neuron.2020.10.029>
- Antony, J. W., Van Dam, J., Massey, J. R., Barnett, A. J., & Bennion, K. A. (2023). Long-term, multi-event surprise correlates with enhanced autobiographical memory. *Nature Human Behaviour*, 1–17. <https://doi.org/10.1038/s41562-023-01631-8>
- Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41.  
<https://doi.org/10.1016/j.media.2007.06.004>
- Aziz-Zadeh, L., Liew, S.-L., & Dandekar, F. (2013). Exploring the neural correlates of visual creativity. *Social Cognitive and Affective Neuroscience*, 8(4), 475–480.  
<https://doi.org/10.1093/scan/nss021>
- Bakkour, A., Palombo, D. J., Zylberberg, A., Kang, Y. H., Reid, A., Verfaellie, M., Shadlen, M. N., & Shohamy, D. (2019). The hippocampus supports deliberation during value-based decisions. *eLife*, 8, e46080. <https://doi.org/10.7554/eLife.46080>
- Baldassano, C. (2023). Studying waves of prediction in the brain using narratives. *Neuropsychologia*, 189, 108664. <https://doi.org/10.1016/j.neuropsychologia.2023.108664>
- Baldassano, C., Beck, D. M., & Fei-Fei, L. (2017). Human–Object Interactions Are More than the Sum of Their Parts. *Cerebral Cortex*, 27(3), 2276–2288.  
<https://doi.org/10.1093/cercor/bhw077>
- 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>

- Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1235–1243.  
<https://doi.org/10.1098/rstb.2008.0310>
- Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *NeuroImage*, 55(4), 1847–1852.  
<https://doi.org/10.1016/j.neuroimage.2011.01.066>
- Bartlett, S. F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48.  
<https://doi.org/10.18637/jss.v067.i01>
- Beatty, R. E., Benedek, M., Barry Kaufman, S., & Silvia, P. J. (2015). Default and Executive Network Coupling Supports Creative Idea Production. *Scientific Reports*, 5(1), 10964.  
<https://doi.org/10.1038/srep10964>
- Behzadi, Y., Restom, K., Liao, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1), 90–101.  
<https://doi.org/10.1016/j.neuroimage.2007.04.042>
- Bein, O., Gasser, C., Amer, T., Maril, A., & Davachi, L. (2023). Predictions transform memories: How expected versus unexpected events are integrated or separated in memory. *Neuroscience & Biobehavioral Reviews*, 153, 105368.  
<https://doi.org/10.1016/j.neubiorev.2023.105368>
- 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), e0115624.  
<https://doi.org/10.1371/journal.pone.0115624>
- Bein, O., Plotkin, N. A., & Davachi, L. (2021). Mnemonic prediction errors promote detailed memories. *Learning & Memory*, *28*(11), 422–434. <https://doi.org/10.1101/lm.053410.121>
- Bellana, B., Mansour, R., Ladyka-Wojcik, N., Grady, C. L., & Moscovitch, M. (2021). The influence of prior knowledge on the formation of detailed and durable memories. *Journal of Memory and Language*, *121*, 104264. <https://doi.org/10.1016/j.jml.2021.104264>
- Bellezza, F. S. (1996). Chapter 10—Mnemonic Methods to Enhance Storage and Retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 345–380). Academic Press.  
<https://doi.org/10.1016/B978-012102570-0/50012-4>
- Bendetowicz, D., Urbanski, M., Garcin, B., Foulon, C., Levy, R., Bréchemier, M.-L., Rosso, C., Thiebaut de Schotten, M., & Volle, E. (2018). Two critical brain networks for generation and combination of remote associations. *Brain*, *141*(1), 217–233.  
<https://doi.org/10.1093/brain/awx294>
- Ben-Yakov, A., & Dudai, Y. (2011). Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. *Journal of Neuroscience*, *31*(24), 9032–9042. <https://doi.org/10.1523/JNEUROSCI.0702-11.2011>
- Biderman, N., Bakkour, A., & Shohamy, D. (2020). What Are Memories For? The Hippocampus Bridges Past Experience with Future Decisions. *Trends in Cognitive Sciences*, *24*(7), 542–556. <https://doi.org/10.1016/j.tics.2020.04.004>
- Bonasia, K., Sekeres, M. J., Gilboa, A., Grady, C. L., Winocur, G., & Moscovitch, M. (2018). Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic

- events at short and long delays. *Neurobiology of Learning and Memory*, 153, 26–39.  
<https://doi.org/10.1016/j.nlm.2018.02.017>
- Bonnen, T., Yamins, D. L. K., & Wagner, A. D. (2021). When the ventral visual stream is not enough: A deep learning account of medial temporal lobe involvement in perception. *Neuron*, 109(17), 2755–2766.e6. <https://doi.org/10.1016/j.neuron.2021.06.018>
- Bowden, E. M., & Jung-Beeman, M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic Bulletin & Review*, 10(3), 730–737.  
<https://doi.org/10.3758/BF03196539>
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, 11(2), 177–220. [https://doi.org/10.1016/0010-0285\(79\)90009-4](https://doi.org/10.1016/0010-0285(79)90009-4)
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 717–726. [https://doi.org/10.1016/S0022-5371\(72\)80006-9](https://doi.org/10.1016/S0022-5371(72)80006-9)
- Brod, G. (2021). Predicting as a learning strategy. *Psychonomic Bulletin & Review*, 28(6), 1839–1847. <https://doi.org/10.3758/s13423-021-01904-1>
- Brod, G., Lindenberger, U., Werkle-Bergner, M., & Shing, Y. L. (2015). Differences in the neural signature of remembering schema-congruent and schema-incongruent events. *NeuroImage*, 117, 358–366. <https://doi.org/10.1016/j.neuroimage.2015.05.086>
- Brod, G., & Shing, Y. L. (2019). A boon and a bane: Comparing the effects of prior knowledge on memory across the lifespan. *Developmental Psychology*, 55(6), 1326–1337.  
<https://doi.org/10.1037/dev0000712>
- Brown, T. I., Carr, V. A., LaRocque, K. F., Favila, S. E., Gordon, A. M., Bowles, B., Bailenson, J. N., & Wagner, A. D. (2016). Prospective representation of navigational goals in the

- human hippocampus. *Science*, 352(6291), 1323–1326.  
<https://doi.org/10.1126/science.aaf0784>
- Brunec, I. K., & Momennejad, I. (2022). Predictive Representations in Hippocampal and Prefrontal Hierarchies. *The Journal of Neuroscience*, 42(2), 299–312.  
<https://doi.org/10.1523/JNEUROSCI.1327-21.2021>
- Buckner, R. L. (2010). The Role of the Hippocampus in Prediction and Imagination. *Annual Review of Psychology*, 61(1), 27–48.  
<https://doi.org/10.1146/annurev.psych.60.110707.163508>
- Buuren, M. van, Kroes, M. C. W., Wagner, I. C., Genzel, L., Morris, R. G. M., & Fernández, G. (2014). Initial Investigation of the Effects of an Experimentally Learned Schema on Spatial Associative Memory in Humans. *Journal of Neuroscience*, 34(50), 16662–16670.  
<https://doi.org/10.1523/JNEUROSCI.2365-14.2014>
- Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, 18(5), 890–896. <https://doi.org/10.3758/s13423-011-0107-8>
- Chandra, S., Sharma, S., Chaudhuri, R., & Fiete, I. (2025). Episodic and associative memory from spatial scaffolds in the hippocampus. *Nature*, 1–13. <https://doi.org/10.1038/s41586-024-08392-y>
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81.  
[https://doi.org/10.1016/0010-0285\(73\)90004-2](https://doi.org/10.1016/0010-0285(73)90004-2)
- Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20(1), 115–125. <https://doi.org/10.1038/nn.4450>

- Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience & Biobehavioral Reviews*, *60*, 82–89. <https://doi.org/10.1016/j.neubiorev.2015.11.011>
- Cheung, O. S., & Bar, M. (2012). Visual prediction and perceptual expertise. *International Journal of Psychophysiology*, *83*(2), 156–163. <https://doi.org/10.1016/j.ijpsycho.2011.11.002>
- Ciric, R., Thompson, W. H., Lorenz, R., Goncalves, M., MacNicol, E. E., Markiewicz, C. J., Halchenko, Y. O., Ghosh, S. S., Gorgolewski, K. J., Poldrack, R. A., & Esteban, O. (2022). TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. *Nature Methods*, *19*(12), 1568–1571. <https://doi.org/10.1038/s41592-022-01681-2>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, *11*(1), Article 1. <https://doi.org/10.1038/s41467-020-17851-9>
- Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E., & Murty, V. P. (2021). Memory consolidation as an adaptive process. *Psychonomic Bulletin & Review*, *28*(6), 1796–1810. <https://doi.org/10.3758/s13423-021-01978-x>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>

- Dalmajjer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. <https://doi.org/10.3758/s13428-013-0422-2>
- De Soares, A., Kim, T., Mugisho, F., Zhu, E., Lin, A., Zheng, C., & Baldassano, C. (2024). Top-down attention shifts behavioral and neural event boundaries in narratives with overlapping event scripts. *Current Biology*, *34*(20), 4729-4742.e5. <https://doi.org/10.1016/j.cub.2024.09.013>
- Den Ouden, H. E., Kok, P., & De Lange, F. P. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. <https://doi.org/10.3389/fpsyg.2012.00548>
- Didierjean, A., & Marmèche, E. (2005). Anticipatory representation of visual basketball scenes by novice and expert players. *Visual Cognition*, *12*(2), 265–283. <https://doi.org/10.1080/13506280444000021A>
- Dimsdale-Zucker, H. R., & Ranganath, C. (2018). Chapter 27 - Representational Similarity Analyses: A Practical Guide for Functional MRI Applications. In D. Manahan-Vaughan (Ed.), *Handbook of Behavioral Neuroscience* (Vol. 28, pp. 509–525). Elsevier. <https://doi.org/10.1016/B978-0-12-812028-6.00027-6>
- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. <https://doi.org/10.1016/j.tics.2010.01.004>
- Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 764–770. <https://doi.org/10.1037/0278-7393.11.1-4.764>

- Elo, A. E. (1978). *The rating of chessplayers, past and present*. Arco Pub.
- Erez, J., Cusack, R., Kendall, W., & Barense, M. D. (2016). Conjunctive Coding of Complex Object Features. *Cerebral Cortex*, *26*(5), 2271–2282.  
<https://doi.org/10.1093/cercor/bhv081>
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. <https://doi.org/10.1038/s41592-018-0235-4>
- Esteban, O., Markiewicz, C. J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., Berleant, S., Ellis, D. G., Pinsard, B., Madison, C., Waskom, M., Notter, M. P., Clark, D., Manhães-Savio, A., Clark, D., Jordan, K., Dayan, M., Halchenko, Y. O., Loney, F., ... Ghosh, S. (2022). *nipy/nipype: 1.8.3* (Version 1.8.3) [Computer software]. Zenodo.  
<https://doi.org/10.5281/ZENODO.596855>
- Esteban, O., Markiewicz, C. J., Goncalves, M., Provins, C., Kent, J. D., DuPre, E., Salo, T., Ciric, R., Pinsard, B., Blair, R. W., Poldrack, R. A., & Gorgolewski, K. J. (2018). *fMRIPrep: A robust preprocessing pipeline for functional MRI* (Version 23.0.2) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.7863421>
- Finnie, P. S. B., Komorowski, R. W., & Bear, M. F. (2021). The spatiotemporal organization of experience dictates hippocampal involvement in primary visual cortical plasticity. *Current Biology*. <https://doi.org/10.1016/j.cub.2021.06.079>

- Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. [https://doi.org/10.1016/S1053-8119\(09\)70884-5](https://doi.org/10.1016/S1053-8119(09)70884-5)
- Frank, D., & Kafkas, A. (2021). Expectation-driven novelty effects in episodic memory. *Neurobiology of Learning and Memory*, *183*, 107466. <https://doi.org/10.1016/j.nlm.2021.107466>
- Frankenstein, A. N., McCurdy, M. P., Sklenar, A. M., Pandya, R., Szpunar, K. K., & Leshikar, E. D. (2020). Future thinking about social targets: The influence of prediction outcome on memory. *Cognition*, *204*, 104390. <https://doi.org/10.1016/j.cognition.2020.104390>
- Frankland, S. M., & Greene, J. D. (2020). Concepts and Compositionality: In Search of the Brain's Language of Thought. *Annual Review of Psychology*, *71*(Volume 71, 2020), 273–303. <https://doi.org/10.1146/annurev-psych-122216-011829>
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. <https://doi.org/10.1038/nrn2787>
- Gasser, C., & Davachi, L. (2023). Cross-Modal Facilitation of Episodic Memory by Sequential Action Execution. *Psychological Science*, *34*(5), 581–602. <https://doi.org/10.1177/09567976231158292>
- Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114. <https://doi.org/10.1016/j.neuropsychologia.2013.11.010>
- Ghosh, V. E., Moscovitch, M., Melo Colella, B., & Gilboa, A. (2014). Schema Representation in Patients with Ventromedial PFC Lesions. *The Journal of Neuroscience*, *34*(36), 12057–12070. <https://doi.org/10.1523/JNEUROSCI.0740-14.2014>

- Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, 21(8), 618–631. <https://doi.org/10.1016/j.tics.2017.04.013>
- Gobet, F., & Waters, A. J. (2003). The Role of Constraints in Expert Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1082–1094. <https://doi.org/10.1037/0278-7393.29.6.1082>
- Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., ... Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, 25(3), 369–380. <https://doi.org/10.1038/s41593-022-01026-4>
- Goodrich, B., Gabry, J., & Brilleman, S. (2022). *rstanarm: Bayesian applied regression modeling via Stan*. [Computer software]. <https://mc-stan.org/rstanarm/>
- Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, 5. <https://doi.org/10.3389/fninf.2011.00013>
- Gorman, A. D., Abernethy, B., & Farrow, D. (2011). Investigating the anticipatory nature of pattern perception in sport. *Memory & Cognition*, 39(5), 894–901. <https://doi.org/10.3758/s13421-010-0067-7>
- Gorman, A. D., Abernethy, B., & Farrow, D. (2012). Classical Pattern Recall Tests and the Prospective Nature of Expert Performance. *Quarterly Journal of Experimental Psychology*, 65(6), 1151–1160. <https://doi.org/10.1080/17470218.2011.644306>

- Graesser, A. C., & Nakamura, G. V. (1982). The Impact of a Schema on Comprehension and Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 16, pp. 59–109). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60547-2](https://doi.org/10.1016/S0079-7421(08)60547-2)
- Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. <https://doi.org/10.1016/j.jml.2016.11.001>
- Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, *148*(2), 325–341. <https://doi.org/10.1037/xge0000498>
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>
- Grossmann, I., Varnum, M. E. W., Hutcherson, C. A., & Mandel, D. R. (2024). When expert predictions fail. *Trends in Cognitive Sciences*, *28*(2), 113–123. <https://doi.org/10.1016/j.tics.2023.10.005>
- Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., & Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829–842. <https://doi.org/10.3758/s13428-015-0642-8>
- Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304–313. <https://doi.org/10.1016/j.tics.2015.04.006>

- Hemmer, P., & Steyvers, M. (2009). A Bayesian Account of Reconstructive Memory. *Topics in Cognitive Science*, 1(1), 189–202. <https://doi.org/10.1111/j.1756-8765.2008.01010.x>
- Higbee, K. L. (1979). Recent research on visual mnemonics: Historical roots and educational fruits. *Review of Educational Research*, 49(4), 611–629. <https://doi.org/10.2307/1169987>
- Höltje, G., & Mecklinger, A. (2022). Benefits and costs of predictive processing: How sentential constraint and word expectedness affect memory formation. *Brain Research*, 1788, 147942. <https://doi.org/10.1016/j.brainres.2022.147942>
- Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, 46(3), 269–299. <https://doi.org/10.1006/jmps.2001.1388>
- Huang, J., Furness, E., Liu, Y., Kenmoe, M.-J., Elias, R., Zeng, H. T., & Baldassano, C. (2025). Accurate Predictions Facilitate Robust Memory Encoding Independently From Stimulus Probability. *Open Mind*, 9, 940–958. <https://doi.org/10.1162/opmi.a.14>
- Huang, J., Velarde, I., Ma, W. J., & Baldassano, C. (2023). Schema-based predictive eye movements support sequential memory encoding. *eLife*, 12, e82599. <https://doi.org/10.7554/eLife.82599>
- Huntenburg, J. (2014). Evaluating nonlinear coregistration of BOLD EPI and T1w images. *Master's Thesis, Berlin: Freie Universität*. <http://hdl.handle.net/11858/00-001M-0000-002B-1CB5-A>.
- Hunter, D. R. (2004). MM algorithms for generalized Bradley-Terry models. *The Annals of Statistics*, 32(1), 384–406. <https://doi.org/10.1214/aos/1079120141>

- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376.  
<https://doi.org/10.1037/0033-295X.98.3.352>
- Jang, A. I., Nassar, M. R., Dillon, D. G., & Frank, M. J. (2019). Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*, *3*(7), Article 7. <https://doi.org/10.1038/s41562-019-0597-3>
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. <https://doi.org/10.1006/nimg.2002.1132>
- Kragel, J. E., & Voss, J. L. (2021). Temporal context guides visual exploration during scene recognition. *Journal of Experimental Psychology: General*, *150*(5), 873–889.  
<https://doi.org/10.1037/xge0000827>
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*.  
<https://www.frontiersin.org/article/10.3389/neuro.06.004.2008>
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.  
<https://doi.org/10.1126/science.aab3050>
- Lampinen, J. M., Faries, J. M., Neuschatz, J. S., & Toglia, M. P. (2000). Recollections of things schematic: The influence of scripts on recollective experience. *Applied Cognitive Psychology*, *14*(6), 543–554. [https://doi.org/10.1002/1099-0720\(200011/12\)14:6%253C543::AID-ACP674%253E3.0.CO;2-K](https://doi.org/10.1002/1099-0720(200011/12)14:6%253C543::AID-ACP674%253E3.0.CO;2-K)

- Lanczos, C. (1964). Evaluation of Noisy Data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. <https://doi.org/10.1137/0701007>
- Lavín, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience*, *7*.  
<https://doi.org/10.3389/fnbeh.2013.00218>
- Lee, A. Y., & Sternthal, B. (1999). The Effects of Positive Mood on Memory. *Journal of Consumer Research*, *26*(2), 115–127. <https://doi.org/10.1086/209554>
- Lee, C. S., Aly, M., & Baldassano, C. (2021). Anticipation of temporally structured events in the brain. *eLife*, *10*, e64972. <https://doi.org/10.7554/eLife.64972>
- Lee, H., & Chen, J. (2021). *Narratives as Networks: Predicting Memory from the Structure of Naturalistic Events* [Preprint]. Neuroscience. <https://doi.org/10.1101/2021.04.24.441287>
- Leonards, U., Sunaert, S., Van Hecke, P., & Orban, G. A. (2000). Attention Mechanisms in Visual Search—An fMRI Study. *Journal of Cognitive Neuroscience*, *12*(Supplement 2), 61–75. <https://doi.org/10.1162/089892900564073>
- Lew, A. R., & Howe, M. L. (2017). Out of place, out of mind: Schema-driven false memory effects for object-location bindings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(3), 404–421. <https://doi.org/10.1037/xlm0000317>
- Liang, J. C., Erez, J., Zhang, F., Cusack, R., & Barense, M. D. (2020). Experience Transforms Conjunctive Object Representations: Neural Evidence for Unitization After Visual Expertise. *Cerebral Cortex*, *30*(5), 2721–2739. <https://doi.org/10.1093/cercor/bhz250>
- Liu, C., Ye, Z., Chen, C., Axmacher, N., & Xue, G. (2022). Hippocampal Representations of Event Structure and Temporal Context during Episodic Temporal Order Memory. *Cerebral Cortex*, *32*(7), 1520–1534. <https://doi.org/10.1093/cercor/bhab304>

- Liu, Y., Hsueh, P.-Y., Lai, J., Sangin, M., Nussli, M.-A., & Dillenbourg, P. (2009). Who is the expert? Analyzing gaze data to predict expertise level in collaborative applications. *2009 IEEE International Conference on Multimedia and Expo*, 898–901.  
<https://doi.org/10.1109/ICME.2009.5202640>
- Liu, Z.-X., Grady, C., & Moscovitch, M. (2017). Effects of Prior-Knowledge on Brain Activation and Connectivity During Associative Memory Encoding. *Cerebral Cortex*, *27*(3), 1991–2009. <https://doi.org/10.1093/cercor/bhw047>
- Maguire, E. A., Valentine, E. R., Wilding, J. M., & Kapur, N. (2003). Routes to remembering: The brains behind superior memory. *Nature Neuroscience*, *6*(1), 90–95.  
<https://doi.org/10.1038/nn988>
- Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General*, *145*(3), 266–272. <https://doi.org/10.1037/xge0000140>
- Masís-Obando, R., Norman, K. A., & Baldassano, C. (2021). *Schema representations in distinct brain networks support narrative memory during encoding and retrieval* (p. 2021.05.17.444363). <https://doi.org/10.1101/2021.05.17.444363>
- Masís-Obando, R., Norman, K. A., & Baldassano, C. (2024). *How sturdy is your memory palace? Reliable room representations predict subsequent reinstatement of placed objects* (p. 2024.11.26.625465). bioRxiv. <https://doi.org/10.1101/2024.11.26.625465>
- McCabe, J. A. (2015). Location, Location, Location! Demonstrating the Mnemonic Benefit of the Method of Loci. *Teaching of Psychology*, *42*(2), 169–173.  
<https://doi.org/10.1177/0098628315573143>

- 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. <https://doi.org/10.1037/0033-295X.102.3.419>
- Menon, V. (2023). 20 years of the default mode network: A review and synthesis. *Neuron*, *111*(16), 2469–2487. <https://doi.org/10.1016/j.neuron.2023.04.023>
- Morris, R. G. M. (2006). Elements of a neurobiological theory of hippocampal function: The role of synaptic plasticity, synaptic tagging and schemas. *European Journal of Neuroscience*, *23*(11), 2829–2846. <https://doi.org/10.1111/j.1460-9568.2006.04888.x>
- Murdock, B. B. (1974). *Human memory: Theory and data* (pp. x, 362). Lawrence Erlbaum.
- Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403–415. <https://doi.org/10.1037/0096-3445.128.4.403>
- Naim, M., Katkov, M., Romani, S., & Tsodyks, M. (2020). Fundamental Law of Memory Recall. *Physical Review Letters*, *124*(1), 018101. <https://doi.org/10.1103/PhysRevLett.124.018101>
- Neuschatz, J. S., Lampinen, J. M., Preston, E. L., Hawkins, E. R., & Toglia, M. P. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Applied Cognitive Psychology*, *16*(6), 687–708. <https://doi.org/10.1002/acp.824>
- Nicholas, J., Daw, N. D., & Shohamy, D. (2022). Uncertainty alters the balance between incremental learning and episodic memory. *eLife*, *11*, e81679. <https://doi.org/10.7554/eLife.81679>

- Ondřej, J. (2025). The method of loci in the context of psychological research: A systematic review and meta-analysis. *British Journal of Psychology*, *n/a(n/a)*.  
<https://doi.org/10.1111/bjop.12799>
- O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311–345.  
<https://doi.org/10.1037/0033-295X.108.2.311>
- Ortiz-Tudela, J., Nicholls, V. I., & Clarke, A. (2023). Parameters of prediction: Multidimensional characterization of top-down influence in visual perception. *Neuroscience & Biobehavioral Reviews*, *153*, 105369.  
<https://doi.org/10.1016/j.neubiorev.2023.105369>
- Ortiz-Tudela, J., Nolden, S., Pupillo, F., Ehrlich, I., Schommartz, I., Turan, G., & Shing, Y. L. (2021). *Not what U expect: Effects of Prediction Errors on Episodic Memory*. PsyArXiv.  
<https://doi.org/10.31234/osf.io/8dwb3>
- Ortiz-Tudela, J., Nolden, S., Pupillo, F., Ehrlich, I., Schommartz, I., Turan, G., & Shing, Y. L. (2023). Not what u expect: Effects of prediction errors on item memory. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0001367>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, *32*.  
<https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>

- Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *NeuroImage*, *144*, 74–82.  
<https://doi.org/10.1016/j.neuroimage.2016.08.051>
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. <https://doi.org/10.1037/bul0000158>
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.  
<https://doi.org/10.1037/a0014420>
- Poskanzer, C., Tarder-Stoll, H., Javid, R., Spolaore, E., & Aly, M. (2025). *Successful prediction is associated with enhanced encoding*. [https://doi.org/10.31234/osf.io/scrxq\\_v2](https://doi.org/10.31234/osf.io/scrxq_v2)
- Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644–667. <https://doi.org/10.1037/a0033194>
- 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*, 320–341. <https://doi.org/10.1016/j.neuroimage.2013.08.048>
- Preston, A. R., & Eichenbaum, H. (2013). Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology*, *23*(17), R764–R773.  
<https://doi.org/10.1016/j.cub.2013.05.041>
- Preuschoff, K., 't Hart, B., & Einhauser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Frontiers in Neuroscience*, *5*.  
<https://www.frontiersin.org/articles/10.3389/fnins.2011.00115>
- Puff, C. R. (1979). *Memory organization and structure*. Academic Press.  
<https://cir.nii.ac.jp/crid/1130000797374336256>

- Quent, J. A., Greve, A., & Henson, R. N. (2022). Shape of U: The Nonmonotonic Relationship Between Object–Location Memory and Expectedness. *Psychological Science*, 095679762211091. <https://doi.org/10.1177/09567976221109134>
- Quent, J. A., Henson, R. N., & Greve, A. (2021). A predictive account of how novelty influences declarative memory. *Neurobiology of Learning and Memory*, 179, 107382. <https://doi.org/10.1016/j.nlm.2021.107382>
- Qureshi, A., Rizvi, F., Syed, A., Shahid, A., & Manzoor, H. (2014). The method of loci as a mnemonic device to facilitate learning in endocrinology leads to improvement in student performance as measured by assessments. *Advances in Physiology Education*, 38(2), 140–144. <https://doi.org/10.1152/advan.00092.2013>
- Radford, A., Kim, J. W., Xu, T., Brockman, G., Mcleavey, C., & Sutskever, I. (2023). Robust Speech Recognition via Large-Scale Weak Supervision. *Proceedings of the 40th International Conference on Machine Learning*, 28492–28518. <https://proceedings.mlr.press/v202/radford23a.html>
- Ramey, M. M., Henderson, J. M., & Yonelinas, A. P. (2022). Episodic memory processes modulate how schema knowledge is used in spatial memory decisions. *Cognition*, 225, 105111. <https://doi.org/10.1016/j.cognition.2022.105111>
- Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, 4(3), Article 3. <https://doi.org/10.1038/nrn1052>
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), Article 1. <https://doi.org/10.1038/4580>

- Raskin, V. (1984). *Semantic Mechanisms of Humor*. Springer Netherlands.  
<https://doi.org/10.1007/978-94-009-6472-3>
- Raykov, P. P., Keidel, J. L., Oakhill, J., & Bird, C. M. (2020). The brain regions supporting schema-related processing of people's identities. *Cognitive Neuropsychology*, *37*(1–2), 8–24. <https://doi.org/10.1080/02643294.2019.1685958>
- Raykov, P. P., Keidel, J. L., Oakhill, J., & Bird, C. M. (2021a). Activation of Person Knowledge in Medial Prefrontal Cortex during the Encoding of New Lifelike Events. *Cerebral Cortex*, bhab027. <https://doi.org/10.1093/cercor/bhab027>
- Raykov, P. P., Keidel, J. L., Oakhill, J., & Bird, C. M. (2021b). Activation of Person Knowledge in Medial Prefrontal Cortex during the Encoding of New Lifelike Events. *Cerebral Cortex*, bhab027. <https://doi.org/10.1093/cercor/bhab027>
- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*(1), 65–81. [https://doi.org/10.1016/0010-0285\(75\)90005-5](https://doi.org/10.1016/0010-0285(75)90005-5)
- Reagh, Z. M., & Ranganath, C. (2023). Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nature Communications*, *14*(1), Article 1. <https://doi.org/10.1038/s41467-023-36805-5>
- Reggente, N., Essoe, J. K. Y., Baek, H. Y., & Rissman, J. (2020). The Method of Loci in Virtual Reality: Explicit Binding of Objects to Spatial Contexts Enhances Subsequent Memory Recall. *Journal of Cognitive Enhancement*, *4*(1), 12–30. <https://doi.org/10.1007/s41465-019-00141-8>
- Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (No. arXiv:1908.10084). arXiv.  
<https://doi.org/10.48550/arXiv.1908.10084>

- Robin, J., & Moscovitch, M. (2017). Details, gist and schema: Hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences*, *17*, 114–123. <https://doi.org/10.1016/j.cobeha.2017.07.016>
- Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1430–1443. <https://doi.org/10.1037/xlm0000518>
- Rudy, J. W., & Sutherland, R. J. (1995). Configural association theory and the hippocampal formation: An appraisal and reconfiguration. *Hippocampus*, *5*(5), 375–389. <https://doi.org/10.1002/hipo.450050502>
- Ruiz, N. A., Meager, M. R., Agarwal, S., & Aly, M. (2020). The Medial Temporal Lobe Is Critical for Spatial Relational Perception. *Journal of Cognitive Neuroscience*, *32*(9), 1780–1795. [https://doi.org/10.1162/jocn\\_a\\_01583](https://doi.org/10.1162/jocn_a_01583)
- Runco, M. A., & Yoruk, S. (2014). The Neuroscience of Divergent Thinking. *Activitas Nervosa Superior*, *56*(1), 1–16. <https://doi.org/10.1007/BF03379602>
- Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. <https://doi.org/10.1016/j.neuroimage.2012.08.052>
- 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>

- Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160049. <https://doi.org/10.1098/rstb.2016.0049>
- Schliephake, L. M., Trempler, I., Roehe, M. A., Heins, N., & Schubotz, R. I. (2021). Positive and negative prediction error signals to violated expectations of face and place stimuli distinctively activate FFA and PPA. *NeuroImage*, 236, 118028. <https://doi.org/10.1016/j.neuroimage.2021.118028>
- Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, 80(1), 1–27. <https://doi.org/10.1152/jn.1998.80.1.1>
- Scoville, W. B., & Milner, B. (1957a). LOSS OF RECENT MEMORY AFTER BILATERAL HIPPOCAMPAL LESIONS. *Journal of Neurology, Neurosurgery & Psychiatry*, 20(1), 11–21. <https://doi.org/10.1136/jnnp.20.1.11>
- Scoville, W. B., & Milner, B. (1957b). 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>
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience*, 27(9), 2349–2356. <https://doi.org/10.1523/JNEUROSCI.5587-06.2007>
- Seghier, M. L. (2013). The Angular Gyrus: Multiple Functions and Multiple Subdivisions. *The Neuroscientist*, 19(1), 43–61. <https://doi.org/10.1177/1073858412440596>

- Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, *138*, 107307. <https://doi.org/10.1016/j.neuropsychologia.2019.107307>
- Shamay-Tsoory, S. G., Adler, N., Aharon-Peretz, J., Perry, D., & Mayseless, N. (2011). The origins of originality: The neural bases of creative thinking and originality. *Neuropsychologia*, *49*(2), 178–185. <https://doi.org/10.1016/j.neuropsychologia.2010.11.020>
- Sherman, B. E., & Turk-Browne, N. B. (2020). Statistical prediction of the future impairs episodic encoding of the present. *Proceedings of the National Academy of Sciences*, *117*(37), 22760–22770. <https://doi.org/10.1073/pnas.2013291117>
- Shin, Y. S., Masís-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2021). Context-dependent memory effects in two immersive virtual reality environments: On Mars and underwater. *Psychonomic Bulletin & Review*, *28*(2), 574–582. <https://doi.org/10.3758/s13423-020-01835-3>
- Skipper, J. I. (2015). The NOLB model: A model of the natural organization of language and the brain. In *Cognitive neuroscience of natural language use* (pp. 101–134). Cambridge University Press. <https://doi.org/10.1017/CBO9781107323667.006>
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604. <https://doi.org/10.1037/0278-7393.4.6.592>
- Sommer, T., Hennies, N., Lewis, P. A., & Alink, A. (2022). The Assimilation of Novel Information into Schemata and Its Efficient Consolidation. *Journal of Neuroscience*, *42*(30), 5916–5929. <https://doi.org/10.1523/JNEUROSCI.2373-21.2022>

- Spreng, R. N., Madore, K. P., & Schacter, D. L. (2018). Better imagined: Neural correlates of the episodic simulation boost to prospective memory performance. *Neuropsychologia*, *113*, 22–28. <https://doi.org/10.1016/j.neuropsychologia.2018.03.025>
- Squire, L. R., Shimamura, A. P., & Amaral, D. G. (1989). 12—Memory and the Hippocampus. In J. H. Byrne & W. O. Berry (Eds.), *Neural Models of Plasticity* (pp. 208–239). Academic Press. <https://doi.org/10.1016/B978-0-12-148955-7.50016-3>
- Squire, L. R., & Zola-Morgan, S. (1991). The Medial Temporal Lobe Memory System. *Science*, *253*(5026), 1380–1386. <https://doi.org/10.1126/science.1896849>
- Stark, S. M., Reagh, Z. M., Yassa, M. A., & Stark, C. E. L. (2018). What’s in a context? Cautions, limitations, and potential paths forward. *Neuroscience Letters*, *680*, 77–87. <https://doi.org/10.1016/j.neulet.2017.05.022>
- Strange, B. A., & Dolan, R. J. (2001). Adaptive anterior hippocampal responses to oddball stimuli. *Hippocampus*, *11*(6), 690–698. <https://doi.org/10.1002/hipo.1084>
- Tal, A., Bloch, A., Cohen-Dallal, H., Aviv, O., Schwizer Ashkenazi, S., Bar, M., & Vakil, E. (2021). Oculomotor anticipation reveals a multitude of learning processes underlying the serial reaction time task. *Scientific Reports*, *11*(1), 6190. <https://doi.org/10.1038/s41598-021-85842-x>
- Tarder-Stoll, H., Baldassano, C., & Aly, M. (2024). The brain hierarchically represents the past and future during multistep anticipation. *Nature Communications*, *15*(1), 9094. <https://doi.org/10.1038/s41467-024-53293-3>
- Thieu, M. K., Wilkins, L. J., & Aly, M. (2024). Episodic-semantic linkage for \$1000: New semantic knowledge is more strongly coupled with episodic memory in trivia experts.

- Psychonomic Bulletin & Review*, 31(4), 1867–1879. <https://doi.org/10.3758/s13423-024-02469-5>
- Thomas Yeo, B. 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>
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). **mediation**: R Package for Causal Mediation Analysis. *Journal of Statistical Software*, 59(5). <https://doi.org/10.18637/jss.v059.i05>
- Tomparry, A., Zhou, W., & Davachi, L. (2020). Schematic memories develop quickly, but are not expressed unless necessary. *Scientific Reports*, 10(1), Article 1. <https://doi.org/10.1038/s41598-020-73952-x>
- Treiber, J. M., White, N. S., Steed, T. C., Bartsch, H., Holland, D., Farid, N., McDonald, C. R., Carter, B. S., Dale, A. M., & Chen, C. C. (2016). Characterization and Correction of Geometric Distortions in 814 Diffusion Weighted Images. *PLOS ONE*, 11(3), e0152472. <https://doi.org/10.1371/journal.pone.0152472>
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and Memory Consolidation. *Science*, 316(5821), 76–82. <https://doi.org/10.1126/science.1135935>
- Tulving, E., Markowitsch, H. J., Kapur, S., Habib, R., & Houle, S. (1994). Novelty encoding networks in the human brain: Positron emission tomography data. *Neuroreport: An*

- International Journal for the Rapid Communication of Research in Neuroscience*, 5(18), 2525–2528. <https://doi.org/10.1097/00001756-199412000-00030>
- Tustison, N. J., Avants, B. B., Cook, P. A., Yuanjie Zheng, Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>
- Twomey, C., & Kroneisen, M. (2021). The effectiveness of the loci method as a mnemonic device: Meta-analysis. *Quarterly Journal of Experimental Psychology*, 74(8), 1317–1326. <https://doi.org/10.1177/1747021821993457>
- van Buuren, M., Kroes, M. C. W., Wagner, I. C., Genzel, L., Morris, R. G. M., & Fernandez, G. (2014). Initial Investigation of the Effects of an Experimentally Learned Schema on Spatial Associative Memory in Humans. *Journal of Neuroscience*, 34(50), 16662–16670. <https://doi.org/10.1523/JNEUROSCI.2365-14.2014>
- van den Honert, R. N., McCarthy, G., & Johnson, M. K. (2017). Holistic versus feature-based binding in the medial temporal lobe. *Cortex*, 91, 56–66. <https://doi.org/10.1016/j.cortex.2017.01.011>
- Van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, 51(12), 2352–2359. <https://doi.org/10.1016/j.neuropsychologia.2013.05.027>
- van Kesteren, M. T. R., Brown, T. I., & Wagner, A. D. (2016). Interactions between Memory and New Learning: Insights from fMRI Multivoxel Pattern Analysis. *Frontiers in Systems Neuroscience*, 10. <https://doi.org/10.3389/fnsys.2016.00046>

- van Kesteren, M. T. R., Rignanes, P., Gianferrara, P. G., Krabbendam, L., & Meeter, M. (2020). Congruency and reactivation aid memory integration through reinstatement of prior knowledge. *Scientific Reports*, *10*(1), 4776. <https://doi.org/10.1038/s41598-020-61737-1>
- van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. <https://doi.org/10.1016/j.tins.2012.02.001>
- van Opheusden, B., Galbiati, G., Kuperwajs, I., Bnaya, Z., li, Y., & Ji, W. (2021). *Revealing the impact of expertise on human planning with a two-player board game* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/rhq5j>
- van Opheusden, B., Kuperwajs, I., Galbiati, G., Bnaya, Z., Li, Y., & Ma, W. J. (2023). Expertise increases planning depth in human gameplay. *Nature*, *618*(7967), Article 7967. <https://doi.org/10.1038/s41586-023-06124-2>
- Varga, D. K., Raykov, P. P., Jefferies, E., Ben-Yakov, A., & Bird, C. M. (2025). Hippocampal mismatch signals are based on episodic memories and not schematic knowledge. *Proceedings of the National Academy of Sciences*, *122*(34), e2503535122. <https://doi.org/10.1073/pnas.2503535122>
- Wagner, I. C., Konrad, B. N., Schuster, P., Weisig, S., Repantis, D., Ohla, K., Kühn, S., Fernández, G., Steiger, A., Lamm, C., Czisch, M., & Dresler, M. (2021). Durable memories and efficient neural coding through mnemonic training using the method of loci. *Science Advances*, *7*(10), eabc7606. <https://doi.org/10.1126/sciadv.abc7606>
- Wahlheim, C. N., Eisenberg, M. L., Stawarczyk, D., & Zacks, J. M. (2022). Understanding Everyday Events: Predictive-Looking Errors Drive Memory Updating. *Psychological Science*, *33*(5), 765–781. <https://doi.org/10.1177/09567976211053596>

- Wang, Q., Hoi, S. P., Wang, Y., Song, C., Li, T., Lam, C. M., Fang, F., & Yi, L. (2020). Out of mind, out of sight? Investigating abnormal face scanning in autism spectrum disorder using gaze-contingent paradigm. *Developmental Science*, *23*(1), e12856. <https://doi.org/10.1111/desc.12856>
- Wang, S., Peterson, D. J., Gatenby, J. C., Li, W., Grabowski, T. J., & Madhyastha, T. M. (2017). Evaluation of Field Map and Nonlinear Registration Methods for Correction of Susceptibility Artifacts in Diffusion MRI. *Frontiers in Neuroinformatics*, *11*. <https://doi.org/10.3389/fninf.2017.00017>
- Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 687–704. [https://doi.org/10.1016/S0022-5371\(79\)90397-9](https://doi.org/10.1016/S0022-5371(79)90397-9)
- Wilcocks, R. W. (1928). *The Effect of an Unexpected Heterogeneity on Attention* (world). <https://www.tandfonline.com/doi/pdf/10.1080/00221309.1928.9920127>
- Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes—Using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *5*(1), 98–107. <https://doi.org/10.1093/scan/nsq024>
- Wynn, J. S., Ryan, J. D., & Moscovitch, M. (2020). Effects of prior knowledge on active vision and memory in younger and older adults. *Journal of Experimental Psychology: General*, *149*(3), 518–529. <https://doi.org/10.1037/xge0000657>
- Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye Movements Actively Reinstates Spatiotemporal Mnemonic Content. *Vision*, *3*(2), Article 2. <https://doi.org/10.3390/vision3020021>

- Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How We Transmit Memories to Other Brains: Constructing Shared Neural Representations Via Communication. *Cerebral Cortex (New York, N.Y.: 1991)*, 27(10), 4988–5000. <https://doi.org/10.1093/cercor/bhx202>
- Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45–57. <https://doi.org/10.1109/42.906424>
- Zhou, Y., Feinman, R., & Lake, B. M. (2024). Compositional diversity in visual concept learning. *Cognition*, 244, 105711. <https://doi.org/10.1016/j.cognition.2023.105711>

## Appendix A: Chapter 2 Supplement

### *Including prediction confidence to predict memory performance*

In our pre-registration, we planned to predict memory accuracy, reaction time, and confidence based on three regressors: in addition to the schema-consistency of the move (move probability) and prediction accuracy, which we reported in the paper, we also proposed to include another model-free measure of prediction: prediction confidence. Overall, we found that incorporating prediction confidence did not change the main findings of our study and we chose not to include this measure in the main text, but for completeness we report these analyses below.

Prediction confidence measured the extent to which a participant spent time looking at specific empty squares versus uniformly fixating across all empty squares pre-stimulus presentation. High values of prediction confidence indicate that a participant spent a large fraction of the trial looking at only a small number of empty squares, indicating a strong prediction about the upcoming move. As in Huang et al., (2023), we compute this as the expected information gain between a uniform distribution over all empty squares and the fixation distribution. Given the fixation time  $T(x_i)$  for each square  $x_i$ , we define  $P(empty)$  as the fraction of the 10 or 8-second window during the initial board phase spent fixating on empty squares, and  $P(x_i) = T(x_i)/P(empty)$  as the normalized fixation distribution over empty squares. The information gain from a fixation is 0 for fixations on occupied squares, and for fixations on empty squares reflects the entropy difference between a uniform distribution and the fixation distribution. Therefore, we define:

$$\text{Prediction confidence} = P(\text{empty}) \cdot (\log(N_{\text{empty}}) - \sum_i^N P(x_i) \log P(x_i))$$

Refer to figure 6 of Huang et al., (2023) for visualization of some example boards with high and low prediction confidence.

We conducted linear mixed effect logistic regression model, predicting memory (accuracy, reaction time, and confidence) from move probability, prediction accuracy, prediction confidence, and the interaction between prediction accuracy and move probability, with subject random intercept.

In study 1 and 2, both move probability and prediction accuracy significantly lead to better memory ( $p < .001$ ). In both studies, prediction confidence also led to better memory, although the effect is only marginally significant in study 2 (study 1:  $\beta = 0.29$ ,  $z = 2.87$ ,  $p = .004$ ; study 2:  $\beta = 0.085$ ,  $z = 1.71$ ,  $p = .09$ ). For the (memory) confidence measure, we found that move probability ( $t = 2.098$ ,  $p = 0.04$ ) and prediction accuracy ( $t = 2.767$ ,  $p = 0.006$ ), and prediction confidence ( $t = 2.51$ ,  $p = 0.01$ ) all led to more confident memory in study 1. Similarly, in study 2, all of the factors led to higher confidence (move probability:  $t = 8.05$ ,  $p < .001$ , prediction accuracy:  $t = 3.71$ ,  $p < .001$ , prediction confidence:  $t = 3.33$ ,  $p < .001$ ). None of the interactions were significant. For reaction time, prediction confidence was strongly associated with faster reaction time in study 1 ( $t = -3.524$ ,  $p < .001$ ) while the other effects were in the same direction but not significant (move probability:  $t = -1.37$ ,  $p = .171$ ; prediction accuracy:  $t = -1.75$ ,  $p = .079$ ). In study 2, move probability, prediction accuracy, and prediction confidence all predicted faster reaction time: move probability ( $t = -3.96$ ,  $p < .001$ ), prediction accuracy ( $t = -2.01$ ,  $p = .044$ ), prediction confidence ( $t = -7.27$ ,  $p < .001$ ). Overall, these results showed that making confident predictions (spending a lot of time focusing on a few empty squares during encoding) lead to better memory, higher confidence, and faster reaction time, while the effects of accurate prediction and move probability remained the same as the main paper.

### ***Including prediction confidence to predict retrieval strategy***

In the pre-registration, we also mentioned that we would predict eye-movement measures at retrieval based on the schema-consistency of the move, prediction confidence, and prediction accuracy. As a result, we also incorporated prediction confidence into the regression predicting  $w_{moveProb}$  (retrieval schematic eye movement) and  $w_{correctMove}$ . We conducted linear mixed effect regressions to predict  $w_{moveProb}$  from move probability, prediction accuracy, and prediction confidence, and the interaction between prediction accuracy and move probability. In study 1, higher move probability led to higher  $w_{moveProb}$  ( $t = 3.3, p = .001$ ), and higher prediction accuracy led to lower  $w_{moveProb}$  ( $t = -9.82, p < .001$ ). Prediction confidence did not have a significant effect on  $w_{moveProb}$  ( $t = 1.558, p = .119$ ). The interaction between move probability and prediction accuracy was not significant. In study 2, move probability did not lead to a change in  $w_{moveProb}$ , but prediction accuracy decreased  $w_{moveProb}$  ( $t = -15.57, p < .001$ ), and prediction confidence increased  $w_{moveProb}$  ( $t = 2.42, p = .016$ ). There is a significant interaction between move probability and prediction accuracy ( $t = 2.94, p = .003$ ).

### ***Analyses based on prediction confidence and accuracy at retrieval***

In our pre-registration, we planned to compute three statistics for retrieval eye movements, each corresponding to a measure of prediction in the previous study (prediction coefficient, prediction confidence, prediction accuracy). In the analysis reported in the main paper,  $w_{moveProb}$  in retrieval is obtained in a similar way as prediction coefficient during the encoding phase, which we reported in detail. It can be considered as a model-based measure of how much schema is being used during retrieval.

We also proposed to look at applying the “prediction confidence” (described in the first section of the supplementary material) and “prediction accuracy” (described in the main paper)

measures to retrieval, as model-free measure of retrieval strategy. A high “prediction accuracy” at retrieval would mean that a participant spent a large fraction of the time during recall looking at the correct move. A high “prediction confidence” at retrieval would mean that participants look at only a small number of empty squares during retrieval (low entropy), potentially indicating that they have the subjective experience of a precise episodic memory (whether or not that memory is accurate). To avoid confusion, here we refer to retrieval “prediction confidence” and “prediction accuracy” as “retrieval fixation confidence” and “retrieval fixation accuracy,” respectively.

$$\text{retrieval fixation confidence} = (\log(N_{\text{empty}}) - \sum_i^N P(x_i) \log P(x_i))$$

Note that, unlike the prediction confidence measure described in the first section of the supplementary materials, we do not weight the retrieval fixation confidence based on the time spent fixating on empty squares – retrieval is self-paced, and we would therefore expect the most confident responses to occur quickly (with relatively little time spent looking at empty squares). Retrieval fixation confidence can be considered a model-free measure of how much participants were debating between different response options, indicating how easy it was to recall the move.

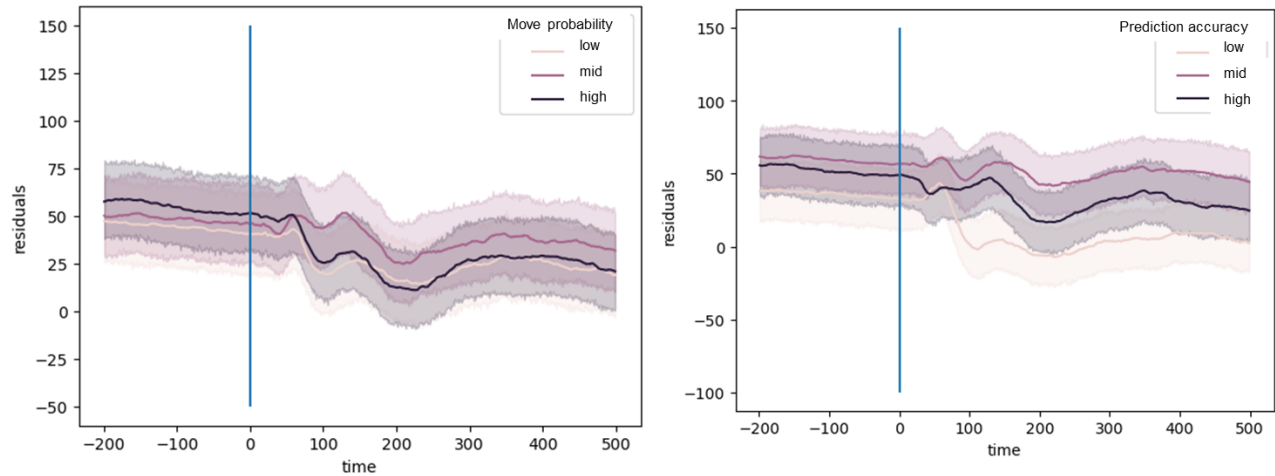
Using mixed-effects linear regression, we predicted memory accuracy, reaction time, and memory confidence from retrieval fixation confidence and accuracy, with random subject intercepts. We found that moves with high retrieval fixation confidence were more likely to be accurately recalled (study 1:  $z = 9.92$ ,  $p < .001$ ; study 2:  $z = 14.32$ ,  $p < .001$ ), with higher memory confidence (study 1:  $t = 15.04$ ,  $p < .001$ ; study 2:  $t = 18.45$ ,  $p < .001$ ), and faster reaction times (study 1:  $t = -22.98$ ,  $p < .001$ ; study 2:  $t = -40.93$ ,  $p < .001$ ). Similarly, moves with high retrieval fixation accuracy were better recalled, with higher memory confidence and faster reaction time (all  $p < .001$ ).

We then tried to predict retrieval fixation confidence from move probability and prediction accuracy (during encoding). We found that in both studies, accurately predicted moves showed higher retrieval fixation confidence (study 1:  $t = 7.55$ ,  $p < .001$ ; study 2:  $t = 7.62$ ,  $p < .001$ ). More probable moves also showed higher retrieval fixation confidence (study 1:  $t = 3.23$ ,  $p = .001$ ; study 2:  $t = 3.54$ ,  $p < .001$ ). These results suggest that for probable and/or predicted moves participants considered only a small number of possible responses before making their response. Based on our results in the main text, however, we hypothesize that these effects arise for two different reasons: accurate predictions facilitate precise episodic memory (allowing for quick retrieval of the correct square), while, for probable moves, the move will be one of the first that participants generate (and then recognize) using their schema knowledge.

### ***Pupillometry analyses***

We also planned in our pre-registration to examine effects of the predictors (prediction accuracy, move probability, and prediction confidence) on pupil size. This was unfortunately difficult with our paradigm – the participants were constantly moving their eyes to look at different pieces with different luminance at different angles from the center of the screen, which had strong effects on measured pupil size. We first interpolated the size of the pupil during blinking periods using the package MNE (Gramfort, 2013), and then temporally downsampled the pupil timecourse by a factor of 10 (to 100 datapoints per second). For all moves, we built a model to predict pupil size from the distance of the eye from the center of the screen, and the color of the move closest to the location of the eye. We then took the residual of the model and visualized the change as a function of time (Figure A.1). As can be seen from the figure, the confidence intervals were largely overlapping for moves with varying probability and prediction

accuracy. We have therefore chosen not to proceed with additional pupillometry analyses for this paradigm.



**Figure A.1. Residual of pupil size as a function of time (1/100 sec).** Different colors represent the tertiles of move probability and prediction accuracy. Blue marks represent the presentation of the stimulus. Error bars represent 95% confidence intervals.

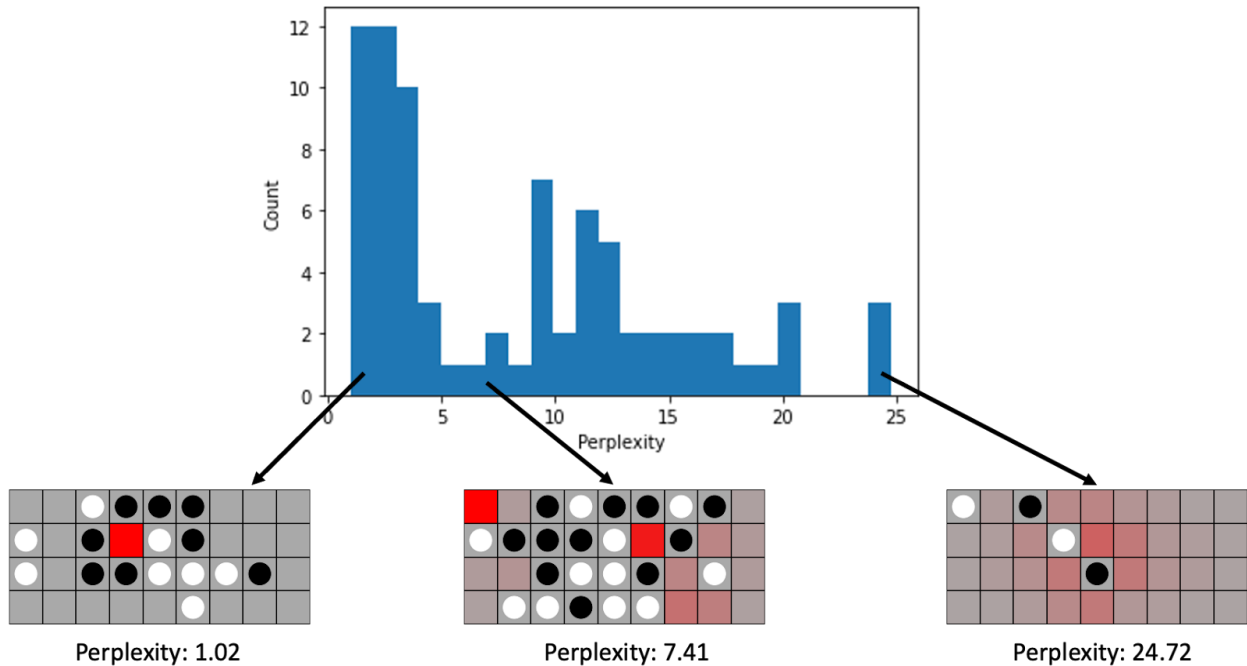
### *Amount of options on the board*

For our manipulation to work, it is important that there are multiple good options for participants to make predictions about and that a move that was not predicted is still relatively probable. We looked at the perplexity of the boards that were shown to the participants in the experiment, calculated as the exponential of the entropy of the probability distribution of the next move:

$$PP = 2^{H(p)}$$

Where  $H(p)$  is the entropy of the probability distribution. The perplexity measure could provide information about the level of uncertainty on the board. A perplexity of  $k$  means that the probability distribution has the same level of uncertainty as a  $k$ -sided dice. Thus, if there is one very dominant move, the perplexity would be close to 1.

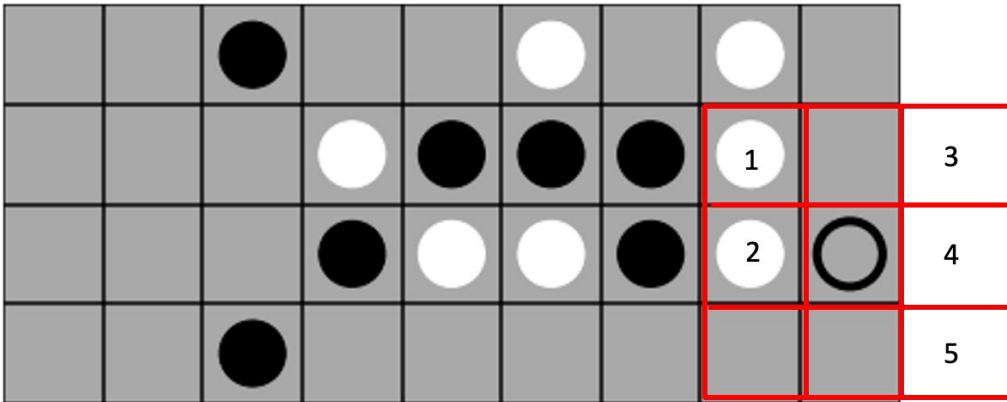
On average, the perplexity of the boards was 8.38 (SD = 6.53). Below we show the distribution of perplexity for the 80 boards participants saw, and examples with minimum, medium, and maximum amount of perplexity (with the color indicating the probability of the moves). Thus, while there are cases where the perplexity is close to 1, in most cases there are at least two moves on the board that are worth considering.



**Figure A.2. Distribution of perplexity in the boards shown to the participants.** The boards are example boards with high, median, and low perplexity, where the color represents the probability of each empty square.

***Number of “landmarks” for each move***

One potential way that participants could remember the moves, which were unrelated to their schema, is to focus on the visual feature of the move. If a move is surrounded by other moves or on the edge of the board, these could serve as “landmarks” for remembering the move. This might explain some of the effect, particularly of move probability, because a better move is more likely to be surrounded by other moves. To test this possibility, we looked at the number of landmarks surrounding a move in our manipulation conditions (Supplementary figure A.3).



**Figure A.3. Illustration of the number of “landmarks” measure.** We looked at the 8 tiles surrounding a move, and count the number of edges and occupied tiles in the 8 tiles.

We found that indeed there is a big difference in the number of landmarks for moves in each condition. In study 1, probable+predicted condition having an average of 5.01 (SD = 1.80) landmarks, followed by predicted but improbable with 4.67 (SD = 1.80); probable and unpredicted 4.05 (SD = 1.97); and improbable and unpredicted having least landmarks of 2.90 (SD = 1.56). Similarly, in study 2, predicted condition has a mean of 4.80 landmarks (SD = 1.78) while unpredicted condition has 4.06 landmarks (SD = 2.00).

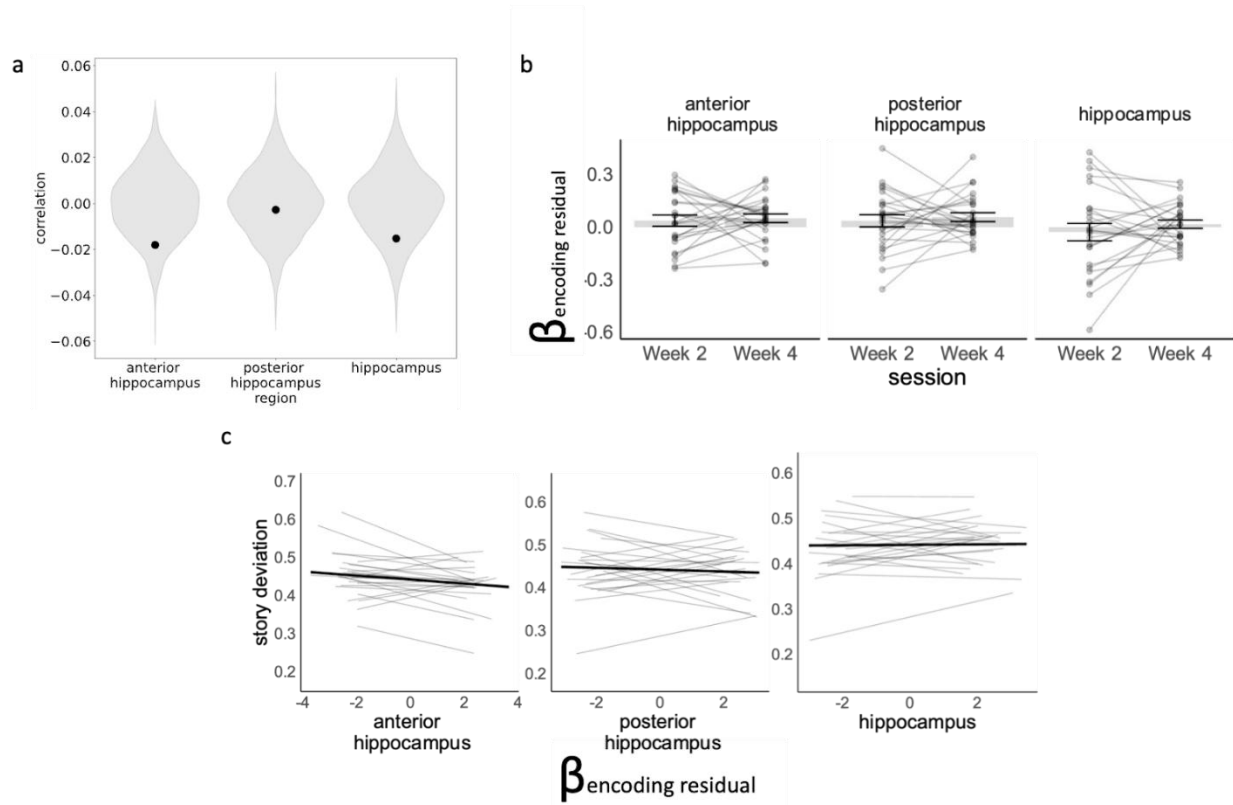
It is therefore a valid concern that the effects of these factors on memory were caused by differences in the number of landmarks. To see if that is the case, we next looked at whether the number of landmarks can account for the memory effect of prediction accuracy and move probability on memory, by including the term into the model predicting memory accuracy from move probability and prediction accuracy.

$$\text{memory\_accuracy} \sim \text{move\_probability} * \text{prediction\_accuracy} + \text{number\_of\_landmarks} + (1|\text{subject\_id})$$

Although the number of landmarks significantly predicted memory accuracy (study 1: beta = .09, z = 4.115, p < .001; study 2: beta = .055, z = 3.80, p < .001), the effects of move

probability and prediction accuracy were still highly significant (all  $p < .001$ ). Thus, even though the number of landmarks is related to both move probability and prediction accuracy, it cannot fully explain the benefit of these factors on memory accuracy.

## Appendix B: Chapter 4 Supplement



**Figure B.1. Results of the ROI analyses done in hippocampal ROIs (anterior, posterior, and whole).** **a.** There was no evidence that encoding residual in any of the hippocampal ROIs tracked semantic similarity in the story told by the participants. **b.** There was no evidence for changes in weights of encoding residual in hippocampus over the course of the training, despite the anterior and posterior hippocampus showing weights significantly above zero at week 4 (anterior:  $t = 2.07$ ,  $p = .05$ , posterior:  $t = 2.25$ ,  $p = .034$ ). **c.** There was no evidence that the amount of encoding residual is related to story deviation.