

Flexible representations of temporal structure guide multistep prediction

Hannah Tarder-Stoll

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

2023

© 2023
Hannah Tarder-Stoll
All Rights Reserved

Abstract

Flexible representations of temporal structure guide multistep prediction

Hannah Tarder-Stoll

Many experiences in our daily lives are temporally structured, enabling prediction of events that will occur in the future. We can anticipate upcoming subway stops during our daily commute, or plan multiple steps ahead when cooking a meal we have made many times.

Although sequences of events in daily life can be multiple steps long, like the stations along a subway line, it is unknown how extended temporal structure enables predictions over multiple timescales. The three studies reported in this dissertation investigate how extended temporal structure is flexibly represented in memory and in the brain to guide multistep prediction.

Chapter 1 demonstrates that memory for temporal structure is enhanced with memory consolidation, enabling more efficient judgements about predictable future events over time.

Chapter 2 shows that temporal structure is represented bidirectionally and hierarchically across the hippocampus and across visual regions during multistep anticipation. Finally, chapter 3 addresses how internal models are updated when regularities in our environment change: the hippocampus rapidly reconfigures memories of temporal structure in response to learning new information, which supports the planning of novel trajectories. Together, the studies presented in this dissertation shed light on how we represent internal models of the world that span multiple timescales to guide adaptive behavior.

Table of Contents

List of Figures.....	iii
List of Tables.....	iv
Acknowledgments	v
Dedication	vii
Introduction	1
Chapter 1: Consolidation enhances multistep anticipatory judgements but diminishes access to perceptual features	9
1.1 Introduction.....	10
1.2 Experiment 1	12
1.2.1 Methods	12
1.2.2 Results.....	23
1.3 Experiment 2	26
1.3.1 Methods	26
1.3.2 Results.....	29
1.4 Modeling Anticipation Strategies	33
1.4.1 Methods	34
1.4.2 Results.....	36
1.5 General Discussion.....	39
Chapter 2: The brain hierarchically represents the past and future during multistep anticipation	43
2.1 Introduction.....	44
2.2 Results	48
2.3 Discussion.....	63

2.4 Methods	68
Chapter 3: Rapid Reconfiguration of Sequence Representations in the Hippocampus.....	89
3.1 Introduction.....	90
3.2 Methods	94
3.3 Results	102
3.4 Discussion.....	108
Discussion	113
References	123
Appendix A: Supplementary Materials for Chapter 1	153
Appendix B: Supplementary Materials for Chapter 2	155

List of Figures

Figure 1.1: Sequence Structure and Stimuli.....	15
Figure 2.2: Task schematic.....	19
Figure 3.3: Anticipation Task performance in Experiment 1.....	25
Figure 4.4: Anticipation Task performance in Experiment 2.....	32
Figure 5.5: Link and Cue Model Schematic and Results.....	38
Figure 2.6: Sequence structure and behavioral training.....	47
Figure 2.2: Anticipation task and behavioral performance.....	49
Figure 2.3: Template brain activity patterns for each environment.....	51
Figure 2.4: Bidirectional and graded representations of temporal structure in hippocampus and visual cortex.....	54
Figure 2.5: Suppression of further environments in the hippocampus is related to response time costs.....	56
Figure 2.6: Bidirectional and graded representations of temporal structure across visual regions reveals within-region hierarchies.....	58
Figure 2.7: Context-dependent representations of multistep anticipation across the cortex.....	62
Figure 3.1: Sequence Structure and Integration Task.....	93
Figure 3.2: Integration Task Performance.....	103
Figure 3.3: Sequence Reconfiguration in Hippocampus.....	106
Figure 3.4: Rapid sequence configuration in the hippocampus is related to integration performance.....	108

List of Tables

Table 7.1: Demographic Information for Participants Across Experiments.....	22
---	----

Acknowledgments

I could easily write another dissertation expounding all the ways in which my colleagues, friends, and family have supported and shaped my graduate school experience, but I will keep this as brief as possible.

To Mariam and Chris, thank you for your exceptional mentorship over the years. Thank you for trusting me to start your labs as one of your first graduate students, for fearlessly leading us through a pandemic, for your patience, for your brilliance, and for your kindness.

Thank you to my committee members, Lila Davachi, Daphna Shohamy, and Russell Epstein, for your knowledge, time, and feedback. I feel honoured to be able to discuss my work and learn from you.

To the entire Alyssano lab, past and present: Halle, Sam, Serra, Eren, Manasi, Matt, Mike, Jiawen, Chey, Craig, Caroline, Taylor, Alex, Nick, and Zall. Thank you for creating a safe, supportive, and fun community where I could learn and grow as a scientist. You have all taught me so much and I am grateful to have had the opportunity to work with you. Thank you to my excellent RAs Edoardo, Maggie, Michelle, Laura, and Franck for making invaluable contributions to this research.

Thank you to Manasi and Halle: I would not have made it through graduate school without you. Thank you for listening and supporting me during the lows, celebrating with me during the highs, and always being there to offer wisdom and sometimes a yummy treat. Thank you to Wangjing, for always bringing a smile to my face, for being willing to explore New York with me, and for being the best food buddy. Thank you also to the wonderful community of graduate students and postdocs; I will miss you.

To the friends I've made in New York, thank you for making this big and overwhelming city feel like home. To my friends from Toronto, thank you for the years of long-distance friendships, for visiting me, and for making it impossible to not return home.

Thank you to my former advisors and continued mentors from the University of Toronto. To Morris Moscovitch, for showing me what it means to be a brilliant scientist and a deeply kind person, simultaneously. To Katherine Duncan, for taking a chance on me, for showing me how fun science can be, and for permanently shaping how I think. To Meg Schlichting, for always having insightful ideas and for your kind support. Morris, Katherine, and Meg, I would not be here without you.

To Tony and Michelle, thank you for welcoming me into your family, into your home at the height of the pandemic, and for supporting Peter and I throughout this journey.

To my brother Kyle, thank you for always putting things into perspective and for never failing to make me laugh.

To my mom and dad, my biggest cheerleaders and role models. Thank you for your unconditional love and unwavering support, regardless of whether I'm singing opera or scanning people's brains. Thank you for filling my life with safety, love, and laughter from the very beginning. Thank you for feeding me, housing me, clothing me, teaching me, protecting me, uplifting me, challenging me, trusting me, and encouraging me. You are my rocks and you are my best friends.

To Peter, thank you for trusting me to turn our lives upside down for five years and for your confidence that it would all work out. For selflessly encouraging me to go after this dream. For spending countless hours laughing and crying with me on FaceTime. For keeping everything in perspective. For your constant love and support. We did it.

Dedication

In memory of Auntie Mandy, whose drive, courage, and love for life shone through in everything she did.

To my mom, love you “less”.

To my dad, my person of the year, every year.

To Peter, for everything.

Introduction

The Adaptive Nature of Memory

Why do we remember? This question is central to understanding our memory system. We often think of episodic memory as the ability to bend “time’s arrow...into a loop” (Tulving, 2002) to reflect on our past experiences. But what is the adaptive purpose of retrospection? Tulving offers a potential answer to this question, noting that memories for the past may enable prediction or imagination about future scenarios (Tulving, 1985, 2005). From this perspective, time’s arrow may bend in a loop to determine an adaptive path forward. Thus, memory’s purpose may be “intrinsically prospective” (Buckner, 2010; S. B. Klein et al., 2002).

Decades of memory research have since confirmed that the episodic memory system is not uniquely used to reflect on past experiences, but is adaptively used to make inferences, generalizations, and predictions (Biderman et al., 2020; Buckner, 2010; Eichenbaum, 2000; Schlichting & Preston, 2015; Shohamy & Adcock, 2010; Shohamy & Wagner, 2008). The hippocampus, which has long been known to play a critical role in forming episodic memories (Scoville & Milner, 1957), is also necessary to imagine novel, future experiences (Hassabis et al., 2007; S. B. Klein et al., 2002; Tulving, 1985). Further demonstrations using fMRI have revealed that remembering the past and imagining the future activate overlapping brain regions (Addis et al., 2007). Together, this suggests that memory plays a critical role in the construction and prediction of future events (Bar, 2007; Buckner, 2010; Hassabis & Maguire, 2007; Schacter et al., 2007, 2012; Schacter & Addis, 2007).

While the past decades have made substantial progress in our understanding of the adaptive function of memory, many questions remain unanswered about how predictions are represented in the brain and used to guide adaptive behavior. The current body of work examines

and extends these seminal ideas by investigating how episodic memory for the temporal order of events enables flexible predictions about the future along a range of timescales. First, I review how temporal structure is represented in the hippocampus and across the brain to guide predictions about future events. Next, I briefly discuss recent empirical and theoretical work that describes how the brain might hierarchically organize experiences, and how this may be used to guide predictions over multiple timescales. Finally, I discuss open questions that I will investigate in the present work.

Temporal structure and prediction in the hippocampus

Our daily lives are filled with temporal structure. For example, your daily commute might consist of the same sequence of events: you might start your morning by getting coffee from a local coffee shop, followed by taking the subway, walking down the block, and eventually arriving at your office. Memory for this structure enables predictions about future events: while at the coffee shop, we can predict that next we will arrive at the subway station. Predicting upcoming events in such a sequence can help us navigate through the world efficiently to achieve our goals, such as arriving at work. Thus, memory for temporal structure can help us build internal models that capitalize on the regularities in our environments to enable prediction.

Indeed, this notion is supported by foundational research on the acquisition of structured knowledge. Tolman (1948) elegantly demonstrated the formation of structured knowledge about environments in rats: after learning specific routes through a maze, rats were able to flexibly use and combine knowledge of the entire maze structure to navigate along novel trajectories (Tolman et al., 1946; Tolman & Gleitman, 1949). Later, the concept of an internal model or cognitive map

was linked to the hippocampus through the discovery of place cells (O’Keefe & Dostrovsky, 1971), prompting suggestions that the hippocampus represents an internal model of the world that guides adaptive behavior (Epstein et al., 2017; O’Keefe & Nadel, 1979). In support of this theory, hippocampal place cells in rodents exhibit prospective firing of spatial locations while planning routes along learned sequential trajectories (Johnson & Redish, 2007; Wikenheiser & Redish, 2015). Notably, although foundational demonstrations of such cognitive maps relied on memory for a spatial environment (Epstein et al., 2017; O’Keefe & Nadel, 1979; Tolman, 1948), it is now appreciated that internal models of structure need not be entirely spatial. They can represent complex sequential structures (Buckner, 2010; Lisman & Redish, 2009), community structure (Schapiro et al., 2013, 2016), attentional states (Aly & Turk-Browne, 2016b, 2016a), and abstract rules (Behrens et al., 2018; Constantinescu et al., 2016; Y. Liu et al., 2019; Schuck et al., 2016) to govern flexible and adaptive behavior in complex environments.

In humans, the hippocampus has been extensively shown to represent the temporal structure of experience (Bellmund et al., 2020; Brown et al., 2016; Davachi & DuBrow, 2015; Deuker et al., 2016; Schapiro et al., 2012; Thavabalasingam et al., 2019). With repeated exposure to a sequence, the hippocampus can rapidly extract temporal regularities to make predictions about upcoming events in the learned order (Davachi & DuBrow, 2015). This extraction of temporal regularities enables prediction: after extensively learning sequence pairs with embedded temporal regularities, multivariate patterns of activity in the hippocampus for the first item in the pair became more similar to the second item, suggesting anticipation (Schapiro et al., 2012). Indeed, anticipatory signals in the hippocampus can be decoded for predicted items in a sequence, even when the predicted items are omitted during sequence viewing (Hindy et al., 2016; Kok & Turk-Browne, 2018) and during the explicit planning of trajectories in a learned

environment, in parallel to rodents navigating along a linear track (Brown et al., 2016). Together, this prior work suggests that the hippocampus forms an internal model of the world that represents the structure of temporal regularities, triggering predictions about upcoming events during temporally structured experience (Lisman & Redish, 2009; Stachenfeld et al., 2017).

The organization of temporally extended experience in the brain

In the previous section, I discussed how the hippocampus represents temporal structure to predict future events. However, prediction is not the sole territory of the hippocampus. Prediction is ubiquitous in the brain, spanning many different cognitive systems (den Ouden et al., 2010), including the visual system (de Lange et al., 2018; Summerfield & Egner, 2009), leading to influential theories suggesting that prediction is a fundamental function of the brain itself (Clark, 2013; Friston, 2005). Indeed, prediction has been shown across the brain, from sensory regions such as visual cortex (Clarke et al., 2022; Ekman et al., 2017, 2023; Hindy et al., 2016; Kok et al., 2012, 2013, 2014), to higher order regions such as the insula, medial prefrontal cortex (mPFC), and orbitofrontal cortex (Wimmer & Büchel, 2019; Lee et al., 2021; Long et al., 2016). For example, in visual cortex, decoding of expected (vs. unexpected) stimuli is enhanced (Kok et al., 2012), even when the visual stimulus is omitted (Hindy et al., 2016). Patterns of activity in medial prefrontal cortex also represent expected stimulus outcomes after learning sequential pairs (Long et al., 2016). Thus, temporal structure is represented across the brain in support of prediction. However, this begs the question of why predictions are represented in such a diverse set of regions. Are whole brain predictions organized in a meaningful way that might be useful for adaptive behavior?

One possibility is that predictions across the brain are organized by a temporal hierarchy. This hierarchy could meet the behavioral demand to keep track of information on multiple timescales to behave adaptively, from minutes to hours to days. For example, we can make predictions about upcoming events multiple steps into the future along our commute to work: while at the coffee shop, we can anticipate that we will arrive at the subway station within minutes, but will arrive at our office potentially up to an hour later. Keeping track of temporal regularities at multiple timescales is important for behaving adaptively, especially in complex, temporally extended environments in which we often have to plan multiple steps ahead to achieve our goals.

How does the brain organize information that exists at multiple timescales? One possible organizing principle is that the brain keeps track of information along a posterior to anterior hierarchy. Recent theories have proposed such a hierarchy to describe how information is integrated over the past, with short timescales of lingering past information represented in posterior regions of the brain and successively longer timescales of information in successively more anterior regions of the brain (Aly et al., 2018; Baldassano et al., 2017; Hasson et al., 2008, 2015; Honey et al., 2012; Lerner et al., 2011). For example, when viewing a film or listening to a narrative in intact vs scrambled order, posterior regions integrate information over short timescales while anterior regions integrate information over long timescales of up to minutes into the past (Aly et al., 2018; Hasson et al., 2008; Lerner et al., 2011). The brain also segments ongoing temporal experiences into events along a cortical hierarchy: event segmentation is more fine-grained in posterior regions and successively coarser in successively more anterior regions, suggesting that the brain organizes continuous perception hierarchically (Baldassano et al., 2017). Posterior to anterior hierarchies have further been proposed to organize mnemonic

representations within the hippocampus (Brunec et al., 2018; Poppenk et al., 2013; Strange et al., 2014).

The brain therefore represents multiple timescales of past information along a posterior to anterior hierarchy. Does the same hierarchy exist in the forward direction? Indeed, recent work has also shown that prediction of upcoming events are organized hierarchically, with shorter timescales of prediction in posterior regions and longer predictions in more anterior brain regions (Brunec et al., 2018; Wimmer & Büchel, 2019; Lee et al., 2021). For example, during repeated viewing of a movie, visual cortex primarily represented the current moment while insula anticipated many seconds into the future (Lee et al., 2021). These findings of hierarchically organized predictions are generally consistent with computational theories that the brain builds models of the world that cache temporal information about successive events, with different predictive timescales in different brain regions (i.e., multi-scale successor representations; Brunec & Momennejad, 2022; Momennejad & Howard, 2018).

Thus, the posterior to anterior hierarchical organization of the brain appears to offer a solution to the problem of how multiple timescales of information are simultaneously represented. Temporal regularities may be represented on multiple timescales across varying brain regions, allowing the brain to predict imminent events as well as those in the far future to achieve both short and long-term goals in complex environments.

Overview of the Current Work

The research reviewed so far has suggested that temporal structure is represented across the brain to guide predictions about future events over multiple timescales, organized along a posterior to anterior hierarchy. However, many questions remain unanswered about the

representations of multistep prediction. Specifically, because multistep predictions capitalize on regularities in our environments over time, they should be maintained, or strengthened, with consolidation and experience to continue to guide adaptive behavior (Hobson & Friston, 2012). However, it is unknown how memory for temporal structure is consolidated to guide multistep predictions over time. Further, for predictions to be useful in guiding adaptive behavior they should be flexibly represented: only those relevant for the current context should be prioritized, and they should be adaptively updated when our environments change. Yet, there is sparse evidence for context-dependent and flexible updating of internal models in service of multistep prediction.

In the current dissertation, I use behavior and fMRI to investigate how temporal structure is flexibly represented, consolidated, and updated to guide predictions about events multiple steps into the future. Across all three chapters, participants learned temporally extended sequences of scenes. The order of the scenes was constant and looped in a circle. This enabled individuals to make predictions about scenes coming up multiple steps into the future, regardless of their location in the sequence. Importantly, pairs of sequences had the same scenes in a different order, allowing us to test how participants flexibly make predictions multiple steps into the future in a context dependent manner.

Our experiments in Chapter 1 explored how consolidation influences multistep anticipatory judgements. We found that consolidation enhances memory for temporal structure to guide multistep anticipation, with an especially large post-consolidation improvement for predictions that extend further into the future. Consolidation, however, came at the cost of diminished access to perceptual features. Further, response time modeling revealed that maintaining a link-based representation of temporal structure, akin to model-based frameworks

(Daw & Dayan, 2014; Doll et al., 2012), improved prediction accuracy. In Chapter 2, we used multivariate fMRI analyses to show that extended temporal structure in both the future and the past was represented in a graded manner. Bidirectional representations of temporal structure were represented hierarchically within and across regions in the visual system and hippocampus, with representations gradually incorporating less information from perception and more information from memory in progressively anterior regions. In Chapter 3, we explored how internal models of extended temporal structure are updated when our environment changes. We showed that the hippocampus rapidly reconfigured sequence memories after they became linked. The degree to which sequences became integrated was related to the ability to generate multi-step predictions along novel trajectories.

Overall, this dissertation uses novel experiment design and analysis approaches to investigate how temporal structure is represented in memory and in the brain to support predictions at multiple timescales. The work presented here advances our knowledge of how internal models spanning multiple timescales are represented and updated to guide adaptive behavior.

Chapter 1: Consolidation enhances multistep anticipatory judgements but diminishes access to perceptual features

Hannah Tarder-Soll, Christopher Baldassano*, & Mariam Aly*

*These authors contributed equally to this work

Please note, chapter published as a preprint as:

Tarder-Stoll, H., Baldassano, C., & Aly, M. (2023). *Consolidation enhances multistep anticipatory judgements but diminishes access to perceptual features*. PsyArXiv.

<https://doi.org/10.31234/osf.io/x2f7s>

1.1 Introduction

Our daily lives are filled with temporal structure, from our commute to work to the steps in cooking our favorite dinner. Memories for temporal regularities allow us to build internal models that are useful for planning and anticipating future events (Behrens et al., 2018; Momennejad, 2020; O’Keefe & Nadel, 1979; Tolman, 1948). Accordingly, for these models to continue to be useful over a lifetime, they must persist over delays of days, weeks, or years. How do we build durable internal models of temporal structure and how do these models transform over time?

Past work has investigated how individuals use temporal structure to anticipate events at a relatively short timescale, such as one step into the future (Davachi & DuBrow, 2015; de Lange et al., 2018; Hindy et al., 2016; Kok et al., 2012, 2014; Kok & Turk-Browne, 2018; Schapiro et al., 2012). Recent work has further shown that sequence memories guide anticipation of events multiple steps in the future (Brunec & Momennejad, 2022; Caucheteux et al., 2023; Elliott Wimmer & Büchel, 2019; Lee et al., 2021), which may help individuals efficiently plan trajectories through learned environments (Bonasia et al., 2016; Brown et al., 2016).

Because we often repeatedly visit an environment whose temporal structure is stable over time, our internal models must be maintained — and potentially transformed — over periods of memory consolidation (Rauss & Born, 2017). Indeed, consolidation improves memory for temporal order (Drosopoulos et al., 2007), including temporal structure acquired through statistical learning (Arciuli & Simpson, 2012; Durrant et al., 2011, 2013; Kim et al., 2009; H. Liu et al., 2023). Consolidation can also enhance integration of sequences with overlapping items (Tomary & Davachi, 2022) and anticipation of upcoming items (Lutz et al., 2018). However, these studies overwhelmingly investigate how motor sequences are consolidated (Galea et al.,

2010; Janacsek & Nemeth, 2012; Kóbor et al., 2017; Lutz et al., 2018; Romano et al., 2010; Sanchez et al., 2010; Walker et al., 2002, 2003), and the processes underlying learning and consolidation of procedural vs. episodic memories likely differ in important ways (Stickgold, 2005). Studies that have investigated episodic learning of temporal structure have only investigated consolidation of shorter sequences (e.g., three items) and did not test anticipatory judgements (Drosopoulos et al., 2007; Tompary & Davachi, 2022). Thus, we have a limited understanding of how episodic memories of extended sequences are learned, maintained, or transformed over time to enable anticipation multiple steps in the future.

How might memory for temporal structure change with consolidation? Consolidation promotes schematization (Dudai et al., 2015; McClelland et al., 1995; Tompary & Davachi, 2017), such that events are remembered at a high level, but often at the expense of perceptual details (Audrain & McAndrews, 2022; Lifanov et al., 2021; Robin & Moscovitch, 2017; Sekeres et al., 2016, 2018; Winocur et al., 2010) with concomitant changes in neural representations that promote structured, or schematic, knowledge (Audrain & McAndrews, 2022; Tompary & Davachi, 2017). If memories for temporal structure are likewise schematized, anticipatory judgments may become more efficient after a period of consolidation — particularly for events in the far future. In tandem, such memories may at first contain visual detail, but lose this detail over time in favor of representing schematized sequential structure.

Thus, our first aim was to test whether consolidation of sequence memories promotes efficient far-reaching anticipation at the expense of maintaining incidental perceptual details. Our second aim was to determine what types of internal models are most useful for multi-step anticipatory judgements (Momennejad, 2020), both soon after learning and after a consolidation delay. One possibility, inspired by model-based frameworks of decision making (Daw & Dayan,

2014), is that individuals may form an internal model of a sequence that contains representations of each link between items. An alternative possibility, inspired by successor representation frameworks (Dayan, 1993; Momennejad et al., 2017; Momennejad & Howard, 2018; Stachenfeld et al., 2017), is that information about future items in a sequence becomes cached into the current item representation.

Across two experiments, we therefore asked how internal models of temporal structure are (1) used to make anticipatory judgements at multiple timescales, (2) transformed with consolidation, and (3) represented in memory to guide accurate anticipation. Participants learned two sequences of ten images that were predictable at the scene category level. Participants were then cued with a scene category and a sequence and made anticipatory judgements about upcoming scene categories up to four steps into the future. We applied sequence-specific perceptual filters to each image; this allowed us to investigate whether incidental perceptual details were represented in participants' internal models by occasionally swapping the perceptual filter from the cued sequence to the one associated with the uncued sequence. In Experiment 1, participants learned the sequences and made anticipatory judgements immediately after. In Experiment 2, participants learned the sequences and made anticipatory judgements both immediately and after a period of consolidation. Finally, we modeled response times from Experiment 1 and 2 to determine the strategies that promoted better retention after consolidation.

1.2 Experiment 1

1.2.1 Methods

Participants

Our target sample size was determined from a priori power analyses. We had two primary measures of interest: the effect of steps into the future and the effect of trial validity. A pilot study ($n = 15$) indicated that the effect of steps into the future was substantially more robust than that of trial validity. We therefore powered our studies to detect the smaller effect of trial validity ($dz = 0.59$). A power analysis conducted with the pwr package in R (Champely et al., 2020) determined that 100 participants would achieve 98% power (at an alpha of 0.05). We opted to use a higher power threshold than the traditional 80% because pilot samples can overestimate effect sizes (Gelman & Carlin, 2014).

We recruited participants through Prolific (www.prolific.co), an online participant management tool, and the Columbia University Psychology Department Participant Pool until our target sample size was met. We recruited a total of 162 participants (116 through Prolific and 46 through Columbia University). Two participants were excluded from data analysis because they failed to respond on more than 50% of trials and 60 participants were excluded from data analysis because they did not perform statistically above chance on the Anticipation Task (56.875% as determined by a binomial test; see Procedure). The data from the remaining 100 participants are analyzed here ($M_{age} = 26.74 \pm 7.33$, $M_{education} = 14.88 \pm 1.91$; see Table 1.1 for demographic information). To be eligible for the experiment, participants had to report that they were between the ages of 18 and 40, fluent English speakers, and resided in the US. Participants were compensated \$7 per hour (Prolific) or received course credit (Columbia Psychology) for participating. All participants provided informed consent and all procedures were in accordance with the Institutional Review Board at Columbia University.

Stimuli

Stimuli consisted of images of ten different scene categories. The categories were: airplane cabins, beaches, bedrooms, castles, city skylines, forests, kitchens, lecture halls, restaurants, and ski slopes. Thirty exemplars of each scene category were used in the experiment for a total of 300 unique images. Images were obtained through the SUN database (Xiao et al., 2010) and through google image searches.

The ten scene categories were used to form both Sequence A (i.e., the first learned sequence) and Sequence B (i.e., the second learned sequence), which had the same ten scene categories in a different order. We used overlapping sequences to encourage rich episodic encoding of context-specific (here, sequence specific) memories (Chanales et al., 2017) and to make the task difficult enough to avoid ceiling performance. The sequences were circular such that the final category in the sequence connected back to the first category. The order of the scene categories across Sequence A and Sequence B were designed to be as distinct as possible: for any given category, the two preceding and two succeeding categories were different across sequences (Figure 1.1A). The order of the scene categories in the sequences was randomized across participants with the constraint that Sequence A was shuffled in the same way across participants to create Sequence B as described above.

We created two versions of each image of each scene category, using Photoshop's "sponge" and "mosaic" filters (Figure 1.1B). This resulted in 600 total images that varied by scene category (ten categories), unique exemplar (thirty of each category), and perceptual filter (two filters). For each participant, each filter was assigned to either Sequence A or Sequence B (i.e., for half of the participants, all images in Sequence A had the sponge filter applied to them and all images in Sequence B had the mosaic filter applied to them, and the opposite was true for

the remaining participants; Figure 1.1C). Participants were not informed about this sequence to filter mapping.

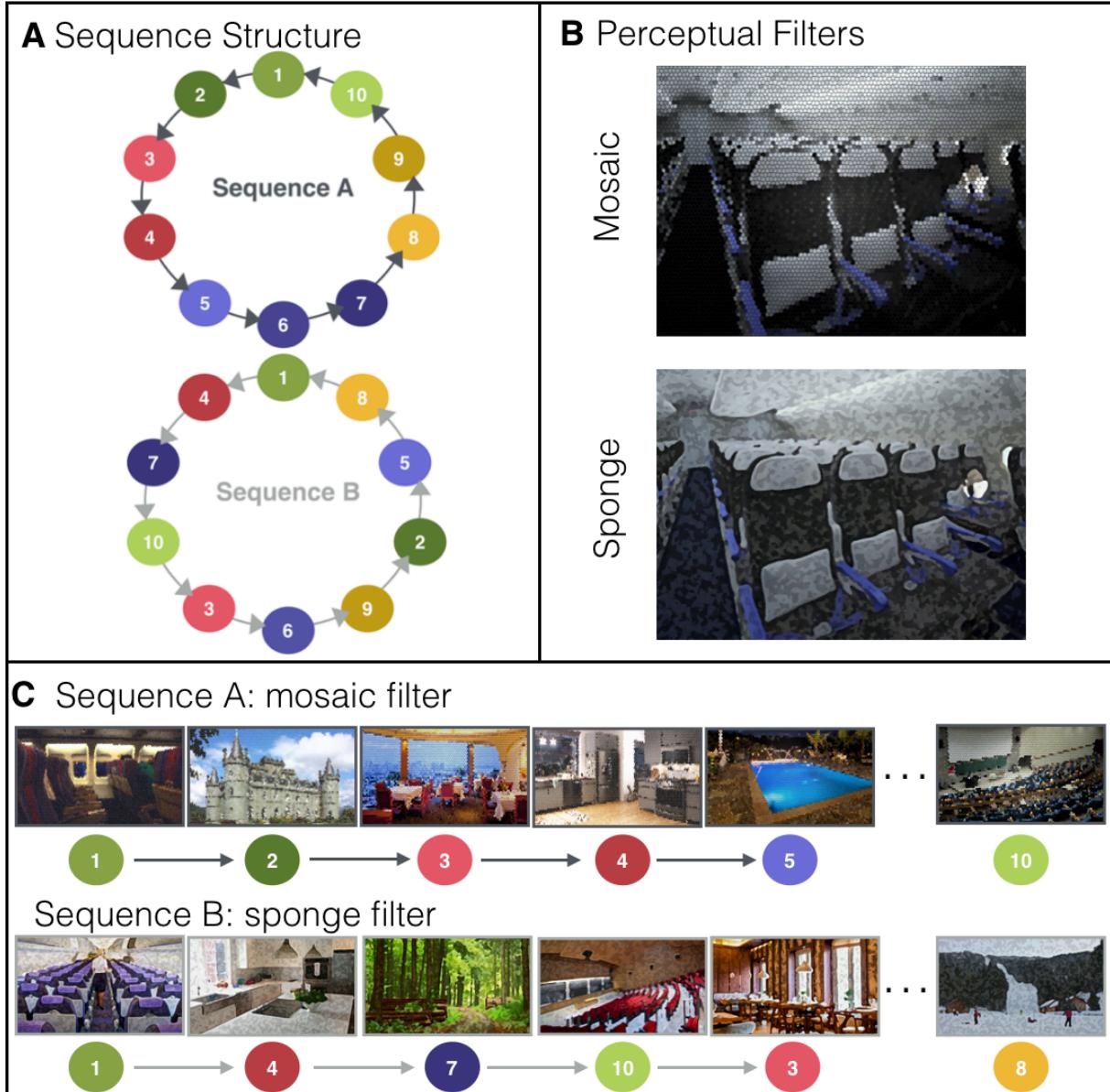


Figure 8.1: Sequence Structure and Stimuli. (A) Sequence structure. The sequences consisted of 10 scene categories (e.g., airplane cabins, swimming pools, etc.), indicated by the colored nodes. Sequence A (dark arrows, top) and Sequence B (light arrows, bottom) consisted of the same scene categories in a different order. The sequences were constructed to be as distinct as possible: for a given category, the two preceding and two succeeding categories were different across the sequences. (B) Perceptual Filters. Each image had a perceptual filter applied to it in either the mosaic style (top) or the sponge style (bottom). Each filter was assigned to one of the two sequences (counterbalanced across participants), but participants were not informed of this

sequence-filter mapping. (C) Example sequences. Examples of the sequence order with perceptual filters applied to each scene exemplar for sample sequence A (top) and sample sequence B (bottom). In this example, sequence A has the mosaic filter and sequence B has the sponge filter. Only six of the ten categories are shown here for illustrative purposes. Sequence A and Sequence B were defined by a fixed order of scene categories, but different exemplars of each category were shown on each trial with minimal exemplar repetition. The sequence order of the categories was randomized across participants.

Procedure

The experiment was conducted on the Gorilla platform (www.gorilla.sc; Anwyl-Irvine et al., 2020) and was composed of two tasks: Sequence Learning and Anticipation (Figure 1.2A).

Sequence Learning. During the Sequence Learning task, participants were instructed to learn the order of two sequences (A and B, see Stimuli). First, participants were shown all ten scene categories in the Sequence A order on the screen and told to generate a story to link the ten categories in order. Participants indicated that they were finished generating a story by pressing a button. Then, they were shown the sequence as pairs of adjacent scene categories with a text box displayed underneath (e.g., images #1 and #2, then images #2 and #3, etc). Participants were told to write down the story that they had generated (Figure 1.2B; see Appendix A for story examples). During the story generation and writing phases, perceptual filters were not applied to the scene images (see Stimuli) so that participants could not incorporate information about the perceptual filters into their stories; this ensured that these perceptual features were incidental to the main Anticipation Task. Participants had unlimited time to generate and write down their stories.

After writing down their Sequence A stories, participants then viewed exemplars of the scene categories in the same sequence order, with the perceptual filter for Sequence A (either mosaic or sponge, counterbalanced across participants) applied to each image (Figure 1.2B).

Participants were told to rehearse their stories. Each image was displayed for five seconds with a 0.5 second inter-trial interval (ITI) between images. After 20% of trials (“test trials”), instead of seeing the next image in the sequence, participants were shown two images of upcoming scene categories and were told to indicate which of those two categories was coming up sooner in the sequence, relative to the preceding category. Participants had eight seconds to respond and were given feedback about whether their answer was correct or incorrect. The entire Sequence A order was shown three times, with the exemplars of each scene category changing on each repetition.

Following Sequence A learning, the story generation, writing, and rehearsal phases were repeated for Sequence B. In the rehearsal phase, each exemplar had the Sequence B filter (either mosaic or sponge, counterbalanced across participants) applied to it.

After Sequence A and Sequence B learning, the rehearsal phase (including test trials) was repeated for each sequence in an interleaved fashion. Participants saw all categories from Sequence A in order, with the sequence-specific filter applied, and then all categories from Sequence B in order, with the sequence-specific filter applied. This procedure occurred three times for a total of six presentations for each sequence. The exemplar of a given scene category was different on each presentation of a given sequence. Finally, participants were prompted to recall the order of Sequence A and Sequence B by writing the order of the categories in provided text boxes. If a participant could not remember a category, they were instructed to write “don’t know” in the text box. In total, the Sequence Learning task took approximately 30 minutes to complete.

Anticipation Task. During the Anticipation Task, participants were cued with an exemplar of a scene category on the screen, along with a sequence cue (A or B) for four seconds (Figure 1.2C). The cue image always had the assigned sequence-specific filter applied to it.

Participants were then probed with two exemplars of upcoming categories and were told to indicate which of the two categories was coming up sooner in the cued sequence, relative to the cue image (Figure 1.2C). The correct probe could be one to four steps away from the cue image (“steps into the future” variable), and the correct and incorrect probes could be one to four steps away from each other in the sequence (“granularity” variable). As the sequences were circular, participants could be cued with any of the scene categories and probed with successors up to eight steps away. Participants had four seconds to respond. On 80% of trials, the probe images had the correct sequence-specific filter applied to them (e.g., a mosaic filter on a mosaic sequence trial; Valid Trials). However, on 20% of trials, the probe images had the filter from the other sequence applied to them (e.g., a sponge filter on a mosaic sequence trial; Invalid Trials; Figure 1.2C). Together, the cue and the probe screens comprised a single trial, with a 1 second ITI between trials.

Participants performed 160 trials of the Anticipation Task, with 128 valid trials and 32 invalid trials. The correct answer was equally distributed across steps into the future (one to four) and granularity (one to four), and valid and invalid trials were equally distributed across each step into the future and granularity condition. Participants performed the Anticipation Task for both Sequence A and B in alternating blocks (with the starting sequence counterbalanced across participants). There were eight blocks (10 trials each) of Sequence A anticipation and eight blocks (10 trials each) of Sequence B anticipation. In each block, participants were cued with unique exemplars of each of the ten scene categories in the sequences in a randomized order. Participants were never cued with the same image (same scene category, exemplar, and filter) more than once. Images were not reused from the Sequence Learning phase and were, at most, shown once as a cue and once as a probe during the Anticipation Task. Participants were given

three 60-second breaks spaced evenly throughout the task. The Anticipation Task took 27 minutes to complete.

Following the Anticipation Task, participants completed a short post-task questionnaire in which we asked them about their strategies during the task and whether or not they noticed the sequence-to-perceptual-feature mapping.

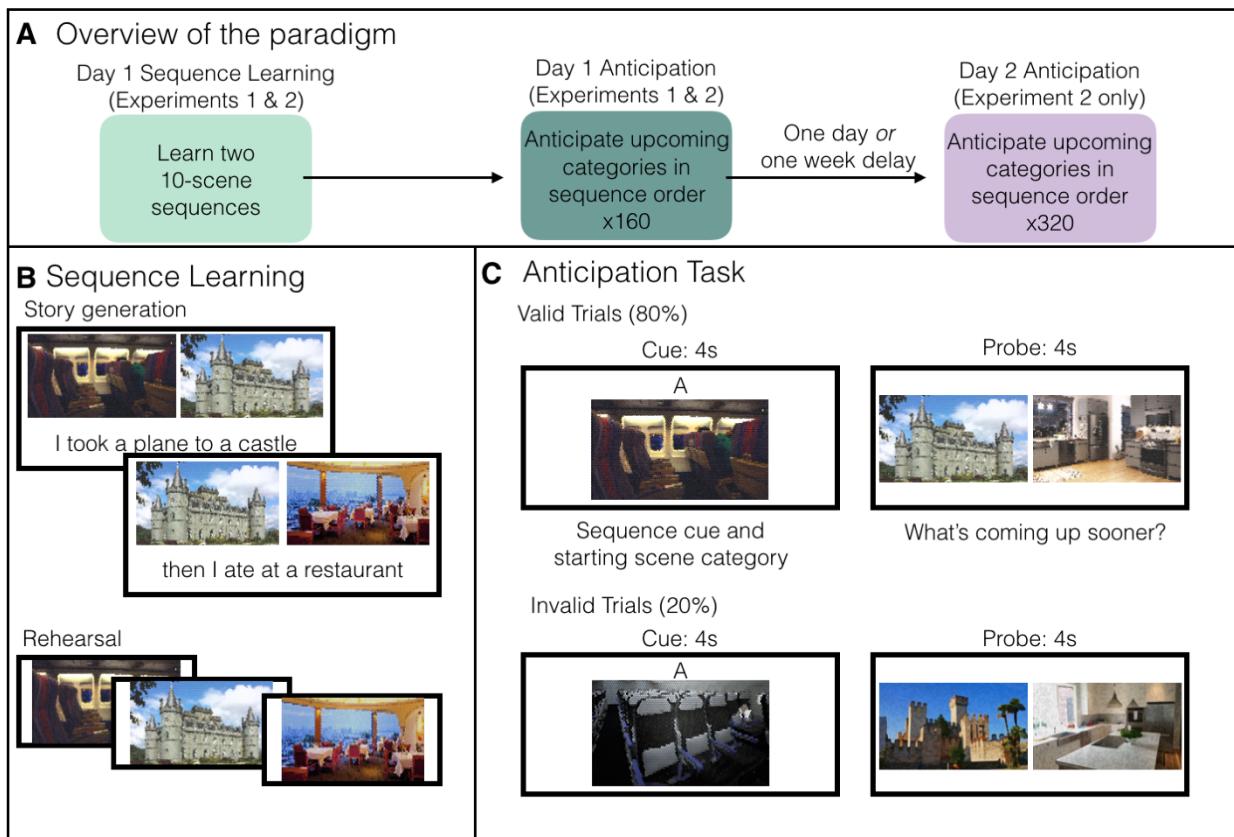


Figure 9.2: Task schematic. (A) Overview of the paradigm. All participants learned the two category sequences (see Figure 1.1) and then immediately completed 160 trials of the anticipation task. In Experiment 2 only, participants returned either one day or one week later and completed another 320 trials of the anticipation task. (B) Sequence learning. Participants learned the two sequences by generating stories that linked adjacent scene categories. After story generation, participants rehearsed their stories (see text for details). Finally, we tested sequence learning with a recall test for each sequence (not shown). (C) Anticipation task. In the anticipation task, participants were cued with an image from one of the learned categories along with a sequence cue (A or B) for 4 seconds. They were then probed with two images and had 4 seconds to indicate which of the two scene categories was coming up sooner in the cued sequence, relative to the cue image. The correct answer could be 1 to 4 steps away from the cue image. On valid trials (80% of trials), images in both the cue and probe screen had the filter from

the cued sequence. In this example, based on the sample sequence in Figure 1.1, both the cue and probe filters are mosaic, the filter assigned to sequence A. On invalid trials (20% of trials), images in the cue screen had the filter from the cued sequence, but images in the probe screen had the filter from the uncued sequence. In this example, the cue filter is mosaic (assigned to sequence A in our example), but the probe filter is sponge (assigned to sequence B in our example). Participants performed this anticipation task for both sequence A and B in interleaved blocks, with the order of the cue images randomized within each block.

Analyses

All analyses were conducted in the R programming language using generalized linear and linear mixed effects models (GLMMs and LMMs, *glmer* and *lmer* function in the *lme4* package; Bates et al., 2014). For analyses that modeled multiple observations per participant, such as accuracy or response time on a given trial, models included random intercepts and slopes for all within-participant effects. All response time models examined responses on correct trials only. For analyses that modeled summary statistics, such as inverse efficiency, models included random intercepts and slopes for all within-participant main effects, but not interactions.

We first checked whether accuracy and response time differed across the sequences (A and B) and the perceptual filters (mosaic and sponge). To examine sequence effects, we fit separate models for accuracy (a GLMM) and response time (an LMM) as a function of sequence (A = -0.5, B = 0.5). We used the following R-based formulas (where “participant” indicates participant number):

```
glmer(correct~sequence+(1+sequence|participant), family = "binomial", data)
```

```
lmer(RT~sequence+(1+sequence|participant), data, subset = (correct == 1))
```

To examine effects of the perceptual filters, we fit separate models for accuracy (a GLMM) and response time (an LMM) as a function of perceptual filter (mosaic = -0.5, sponge = 0.5). We used the following R-based formulas:

```
glmer(correct~filter+(1+filter|participant), family = "binomial", data)
```

```
lmer(RT~filter+(1+filter|participant), data, subset = (correct == 1))
```

For our primary analyses, we used inverse efficiency as our dependent variable (average response time divided by accuracy; Townsend & Ashby, 1978). We opted to use inverse efficiency because performance generally got worse, as measured by both response times and accuracy, with further steps into the future (i.e., slower response times and less accurate responses with further steps in the future). Thus, inverse efficiency effectively captured both aspects of performance.

Because we had a priori hypotheses about the interaction between steps into the future and trial type (valid vs. invalid), we calculated this inverse efficiency score for each participant in each trial type and steps into the future bin. Finally, we conducted an LMM predicting inverse efficiency as a function of steps into the future (-1.5 = 1 step, -0.5 = 2 steps, 0.5 = 3 steps, 1.5 = 4 steps), trial type (-0.5 = valid trials, 0.5 = invalid trials), and their interaction. We used the following R-based formula:

```
lmer(inverseEfficiency~trialType*steps+(1+trialType+steps|participant), data)
```

Finally, we used an LMM to model inverse efficiency as a function of granularity (-1.5 = 1 step, -0.5 = 2 steps, 0.5 = 3 steps, 1.5 = 4 steps). We used the following R-based formula:

```
lmer(inverseEfficiency~granularity+(1+granularity|participant), data)
```

We did not include granularity or sequence (A vs. B) in our main inverse efficiency model because (1) we did not have enough trials to separately estimate the effects of steps into the future, trial type, granularity, and sequence in the same model and (2) there were no significant interactions between granularity and any other variable or between sequence and any

other variable. However, because steps into the future, granularity, and sequence were orthogonalized, our design effectively controls for these variables at each step in the future.

	Experiment 1	Experiment 2 (1 day delay)	Experiment 2 (1 week delay)
Sample Size	100	99	105
Age (M \pm SD)	26.74 ± 7.33	31.65 ± 5.03	30.65 ± 6.73
Years of Education (M \pm SD)	14.88 ± 1.91	15.18 ± 1.85	14.99 ± 2.12
Recruitment Method	116 Prolific 46 CU	99 Prolific	105 Prolific
Gender	61 F 38 M 1 NB	51 F 47 M 1 NB	55 F 45 M 4 NB
Race	63 W 19 A 10 BR 6 B/AA 1 AI/AN 1 O	77 W 7 A 7 B/AA 7 BR 1 O	80 W 8 B/AA 8 BR 6 A 1 AI/AN 1 O
Ethnicity	97 NH/L 3 H/L	91 NH/L 8 H/L	88 NH/L 16 H/L

Table 1.1: Demographic Information for Participants Across Experiments. Abbreviations used in the table are as follows: for recruitment method, CU = Columbia University Psychology Department Participant Pool; for gender, F = female, M = male, and NB = non-binary; for race, W = white, A = Asian, BR = bi-racial, B/AA = Black or African American, AI/AN = American Indian or Alaskan Native, and O = other; for ethnicity, NH/L = not Hispanic or Latino, and H/L = Hispanic or Latino. Note that one participant failed to report demographic information for Experiment 2 (one-week delay).

1.2.2 Results

Sequence Learning

Participants successfully learned the category sequences, achieving an accuracy of 84.33% on test trials from the rehearsal phase of the Sequence Learning task (greater than chance level of 50%, $t(99) = 25.042$, $p < 0.001$). Accuracy was high across both sequences (Sequence A: 85.3%, Sequence B: 81.5%), but was significantly higher for Sequence A (i.e., the first learned sequence) than Sequence B (i.e., the second learned sequence; $\beta = -0.313$, 95% CI [-0.537, -0.089], $p = 0.006$). When asked to explicitly recall the category orders, participants' average accuracy was 97%, and recall accuracy was not significantly different between sequences (Sequence A: 96.9%; Sequence B: 97%; $V = 78.5$, $p = 0.488$; because most participants scored 9/10 or 10/10 on recall, we used a Wilcoxon signed rank test to account for non-normality). Overall, these results verify that participants learned the order of both sequences.

Anticipation Task

Overall Task Performance

We first conducted analyses to ensure that participants performed effectively on the Anticipation Task. Participants successfully used their memory to anticipate upcoming events (mean accuracy = 75.22%; $t(99) = 72.19$, $p < 0.00001$ compared to chance performance). Participants were overall more accurate for Sequence B compared to Sequence A ($\beta = 0.149$, 95% CI [0.044, 0.255], $p = 0.005$), although accuracy was high and significantly above chance for both sequences (Sequence A: 74%, $t(99) = 21.117$, $p < 0.00001$; Sequence B: 76.4%, $t(99) = 23.22$, $p < 0.00001$). Response times did not significantly differ between the two sequences ($\beta =$

-23.43, 95% CI [-56.718, 9.806], $p = 0.137$). Furthermore, neither accuracy ($\beta = -0.063$, 95% CI [-0.161, 0.035], $p = 0.206$) nor response times ($\beta = 26.12$, 95% CI [-3.777, 56.027], $p = 0.09$) differed between trials with a mosaic vs. sponge filter.

Primary Analyses

Turning to our hypothesized effects of interest, we calculated each participant's inverse efficiency (average response time divided by accuracy; Townsend & Ashby, 1978) for each step into the future and conducted an LMM (see Figure 1.3A for model coefficients). Steps into the future robustly impacted inverse efficiency, with less efficient responses for progressively further steps into the future ($\beta = 310.41$, 95% CI [243.846, 376.974], $p < 0.000001$; Figure 1.3B).

Granularity also influenced inverse efficiency, with less efficient responses when the number of steps between the probes was smaller ($\beta = -64.60$, 95% CI [-99.822, -29.37], $p = 0.0005$).

Participants therefore accurately anticipated scene categories multiple steps into the future. How visually detailed were their predictions? To determine whether participants were only anticipating the upcoming category or were also representing sequence-specific perceptual features, we investigated whether trial type affected performance, and whether this interacted with steps into the future. If participants anticipate visually-specific information, performance should decrease on invalid trials, in which participants are probed with an incorrect sequence filter, compared to valid trials, in which the perceptual filter matches learned expectations.

Consistent with our hypothesis, participants responded less efficiently to invalid vs. valid trials ($\beta = 224.01$, 95% CI [87.069, 360.950], $p = 0.002$; Figure 1.3C). This cost to performance is striking because the sequence-specific filters were not relevant to the correct answer and, when asked in the post-task questionnaire, most participants reported not noticing the sequence-to-filter mapping ($n = 68$ did not notice, $n = 32$ noticed). Importantly, the effect of trial type on

inverse efficiency was not different between participants who noticed the manipulation compared to those who did not ($\beta = -203.44$, 95% CI [-495.686, 88.798], $p = 0.175$).

Finally, we tested whether the effect of valid vs. invalid trials was larger for closer vs. further predictions. If anticipated information declines in vividness or detail the further in the future it is, then trial validity should have a larger effect on nearby vs. farther away predictions. Contrary to our hypothesis, however, there was no interaction between steps into the future and trial type on inverse efficiency ($\beta = 51.12$, 95% CI [-51.542, 153.786], $p = 0.330$).

Together, our results show that individuals accurately anticipate events multiple steps in the future, but do so less efficiently for further steps, and that anticipated information at multiple timescales contains task-irrelevant perceptual features, regardless of whether participants explicitly noticed these features and how far into the future participants anticipated.

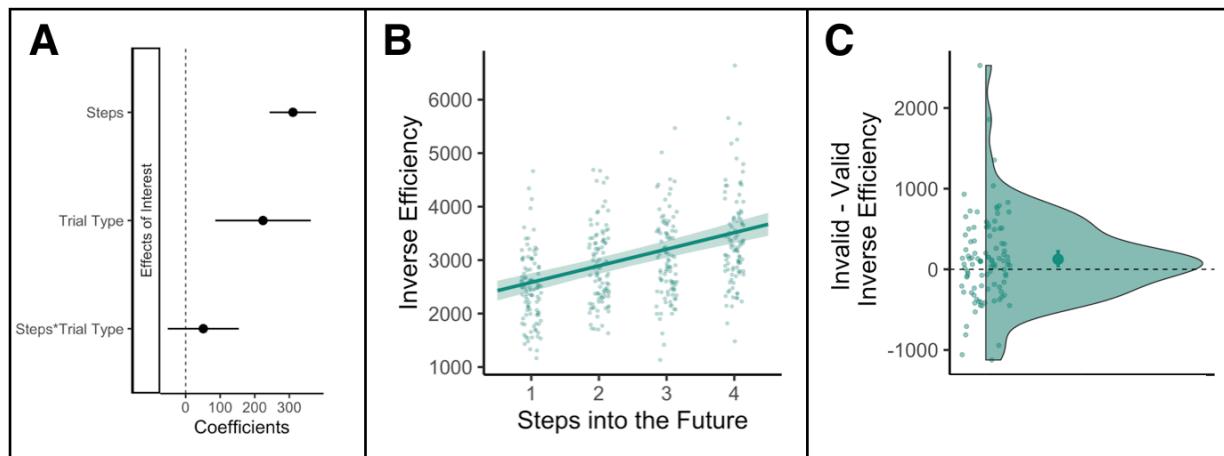


Figure 10.3: Anticipation Task performance in Experiment 1. (A) Coefficient estimates with 95% confidence intervals for the inverse efficiency model. The black points indicate effects of interest and the gray points indicate all other effects in the model. Predictors in the model were effect coded. (B) Inverse efficiency as a function of the number of steps between the cue and the correct probe (“steps into the future”). Higher inverse efficiency values indicate less efficient (slower and/or less accurate) responses. Responses were less efficient when the correct probe was more (vs. fewer) steps away from the cue. Green lines and error ribbons indicate model predictions with 95% confidence intervals; green points indicate each participant’s average inverse efficiency for each step into the future. (C) Inverse efficiency differences between valid

and invalid trials (“trial type”). The dashed line at 0 indicates equally efficient responses on valid and invalid trials. Responses on invalid trials were overall less efficient compared to valid trials, as indicated by a positive difference score. Small green points indicate the average inverse efficiency difference for each participant. The large green point indicates the average inverse efficiency difference across participants with 95% confidence intervals.

1.3 Experiment 2

Having shown that individuals anticipate events multiple steps into the future with perceptual detail in Experiment 1, we next investigated how these effects change with consolidation. We ran a second experiment in which participants completed the Anticipation Task both immediately after Sequence Learning, to replicate Experiment 1, and either one day or one week later, to determine how varying lengths of consolidation influence multistep anticipation.

1.3.1 Methods

Participants

Our target was to double the sample size of Experiment 1, with participants split between two delay conditions. We recruited 355 participants through Prolific (www.prolific.co) to meet our target sample size. Sixty-two participants did not complete the second session of the experiment (21 participants in the one-day condition and 41 participants in the one-week condition), leaving 293 participants that completed the full experiment. An additional six participants were excluded from data analysis because they failed to respond on more than 50% of trials and 83 participants were excluded from data analysis because they did not perform statistically above chance on the Anticipation Task during the first session (56.875% as determined by a binomial test; see Procedure). Applying these exclusions resulted in 204

participants (One-day condition: $n = 99$, $M_{age} = 31.65 \pm 5.03$, $M_{education} = 15.18 \pm 1.85$; One-week condition: $n = 105$, $M_{age} = 30.65 \pm 6.73$, $M_{education} = 14.99 \pm 2.12$; see Table 1.1 for demographic information).

To be eligible for the experiment, participants had to report that they were between the ages of 18 and 40, fluent English speakers, and resided in the US. Participants were compensated \$8 per hour for participating in session 1 and \$8.50 per hour for participating in session 2. All participants provided informed consent and all procedures were in accordance with the Institutional Review Board at Columbia University.

Stimuli

Stimuli were identical to Experiment 1.

Procedure

The experiment was conducted on the Gorilla platform (www.gorilla.sc; Anwyl-Irvine et al., 2020) and was composed of three tasks: Sequence Learning, Session 1 Anticipation, and Session 2 Anticipation (Figure 1.2A).

Sequence Learning. The Sequence Learning task was identical to Experiment 1.

Session 1 Anticipation Task. The Session 1 Anticipation Task was identical to Experiment 1, except that participants did not complete the post-task questionnaire following the task.

Session 2 Anticipation Task. Half of the participants were invited to return one day later and half were invited to return one week later to take part in the Session 2 Anticipation Task. Participants were informed about the opportunity to return for a follow-up experiment the evening before they could complete the Session 2 test. The task was identical to the Session 1

Anticipation Task, except that the total number of trials was increased to 320, with 256 valid trials and 64 invalid trials. As such, there were 16 blocks of Sequence A anticipation (ten trials each) and 16 blocks of Sequence B anticipation (ten trials each). Sequence A and Sequence B blocks alternated, with the starting sequence counterbalanced across participants. Participants were given six 60-second breaks spaced evenly throughout the task. The Session 2 Anticipation task took 60 minutes to complete. Following the task, participants completed the post-task questionnaire.

Analyses

All analyses were identical to Experiment 1, with the following exceptions. In our primary inverse efficiency model, we included (1) delay (immediate test = -0.5, delayed test = 0.5) as a main effect and a random effect, and (2) the interaction between delay, steps into the future, and trial type. Additionally, to control for the varying delay lengths (one day or one week) and to determine whether they differentially influenced any consolidation-dependent effects, we included (1) delay length (-0.5 = one-day delay, 0.5 = one-week delay) as a main effect, and (2) the interaction between delay length, delay, steps into the future, and trial type. We report significant interactions with delay length in Results. However, because there were no major differences between the one-day and one-week delay, we report the main results collapsed across delay length. As in Experiment 1, we report the main effect of granularity on inverse efficiency, but we did not include granularity or sequence in our main inverse efficiency model after confirming there were no significant interactions between granularity and other variables or between sequence and other variables.

1.3.2 Results

Sequence Learning

To verify participants learned both sequences, we calculated accuracy on test trials from the rehearsal phase of the Sequence Learning Task (see Methods). As in Experiment 1, accuracy was significantly higher than chance performance (50%) on the test trials (82.62%, $t(203) = 30.457$, $p < 0.000001$). Accuracy was high across both sequences (Sequence A: 79.4%, Sequence B: 83.7%), but was significantly higher for Sequence B (i.e., the second learned sequence) than Sequence A (i.e., the first learned sequence; $\beta = 0.334$, 95% CI [0.183, 0.485], $p = 0.00001$).

Next, we calculated recall accuracy for Sequences A and B. Participants' average recall accuracy was 96.5%. Recall accuracy was not significantly different between sequences (Sequence A: 96.0%; Sequence B: 97.1%; $V = 324$, $p = 0.675$; because most participants scored 9/10 or 10/10 on recall, we used a Wilcoxon signed rank test to account for non-normality). Overall, these results verify that participants learned the order of both sequences.

Anticipation Task

Overall Task Performance

We first ensured participants were performing effectively on the Anticipation Task for both the immediate and delayed tests in both delay length conditions. Accuracy was significantly higher than chance performance (50% accuracy) in the Session 1 (i.e., the immediate test) and Session 2 (i.e., the delayed test) Anticipation Tasks in both the one-day and one-week conditions (all $p < 0.00001$).

Accuracy was not significantly different between Sequences A and B ($\beta = -0.016$, 95% CI [-0.082, 0.05], $p = 0.636$). Response times, however, were slower for Sequence B compared

to Sequence A (i.e., the second vs. first learned sequence; $\beta = 22.264$, 95% CI [8.387, 35.918], $p = 0.001$), with this effect becoming larger at the delayed test compared to the immediate test ($\beta = 30.203$, 95% CI [12.764, 47.643], $p = 0.007$). Neither accuracy ($\beta = -0.038$, 95% CI [-0.084, 0.008], $p = 0.111$) nor response time ($\beta = 0.305$, 95% CI [-10.74, 11.35], $p = 0.957$) differed between trials with a mosaic vs. sponge filter.

Primary Analyses

Turning to our hypothesized effects of interest, we calculated inverse efficiency scores to capture both accuracy and response time in a single measure. We then conducted a linear mixed-effects model (LMM) assessing whether delay, delay length, steps into the future, trial type, and their interactions influenced inverse efficiency on the Anticipation Task (see Figure 1.4A for model coefficients).

Delay robustly influenced inverse efficiency, with participants' responses becoming more efficient in the delayed vs. immediate test ($\beta = -318.413$, 95% CI [-403.126, -233.699], $p < 0.000001$; Figure 1.4B). There was a significant interaction between delay and delay length such that efficiency in the delayed (vs. immediate) test improved more at the one-day (vs. one-week) delay ($\beta = 161.470$, 95% CI [7.206, 315.735], $p = 0.041$), but there were no other interactions with delay length. This suggests that delay-dependent effects of other variables (e.g., steps into the future, trial type) on memory were similar for the one-day and one-week delays.

Replicating Experiment 1, both granularity and steps into the future affected inverse efficiency: there were less efficient responses when the number of steps between the probes was smaller ($\beta = -64.254$, 95% CI [-76.584, -51.924] $p < 0.000001$) and when the correct answer was further into the future ($\beta = 231.175$, 95% CI [202.401, 259.950], $p < 0.000001$; Figure 1.4B).

As hypothesized, there was a significant steps into the future by delay interaction, such that inverse efficiency improved the most at the delayed test (vs. immediate test) for further steps into the future ($\beta = -62.635$, 95% CI [-105.992, -19.279], $p = 0.005$; Figure 1.4B). There was no three-way interaction between steps into the future, delay, and delay length; this suggests that delay-dependent changes in the effect of steps into the future on inverse efficiency were similar at the one-day and one-week tests ($\beta = -15.525$, 95% CI [-140.943, 32.483], $p = 0.726$).

Therefore, participants became more efficient in their responses following consolidation, and this efficiency improvement was larger for further steps into the future. Was this improvement in anticipation for further steps accompanied by forgetting of sequence-specific perceptual details? To answer this question, we again investigated whether trial type (valid vs. invalid) affected performance, and whether this interacted with delay, steps into the future, and delay length (see Figure 1.4A for model coefficients). Replicating Experiment 1, participants responded less efficiently to invalid vs. valid trials ($\beta = 129.683$, 95% CI [71.837, 187.529], $p = 0.00001$; Figure 1.4C). Critically, there was also a trial type by delay interaction such that the difference in efficiency between invalid and valid trials was smaller in the delayed (vs. immediate) test ($\beta = -139.373$, 95% CI [-236.321, -42.424], $p = 0.05$; Figure 1.4C). Contrary to our hypothesis but consistent with Experiment 1, there was no interaction between trial type and steps into the future ($\beta = -6.171$, 95% CI [-49.528, 37.185], $p = 0.780$).

Together, our results show that anticipation became more efficient following consolidation, with a larger efficiency improvement for further steps into the future, compared to closer ones. Consolidation also reduced the extent to which sequence-specific perceptual features were incorporated into anticipated information, as indicated by the more limited impact of the perceptual filter applied to the probe images.

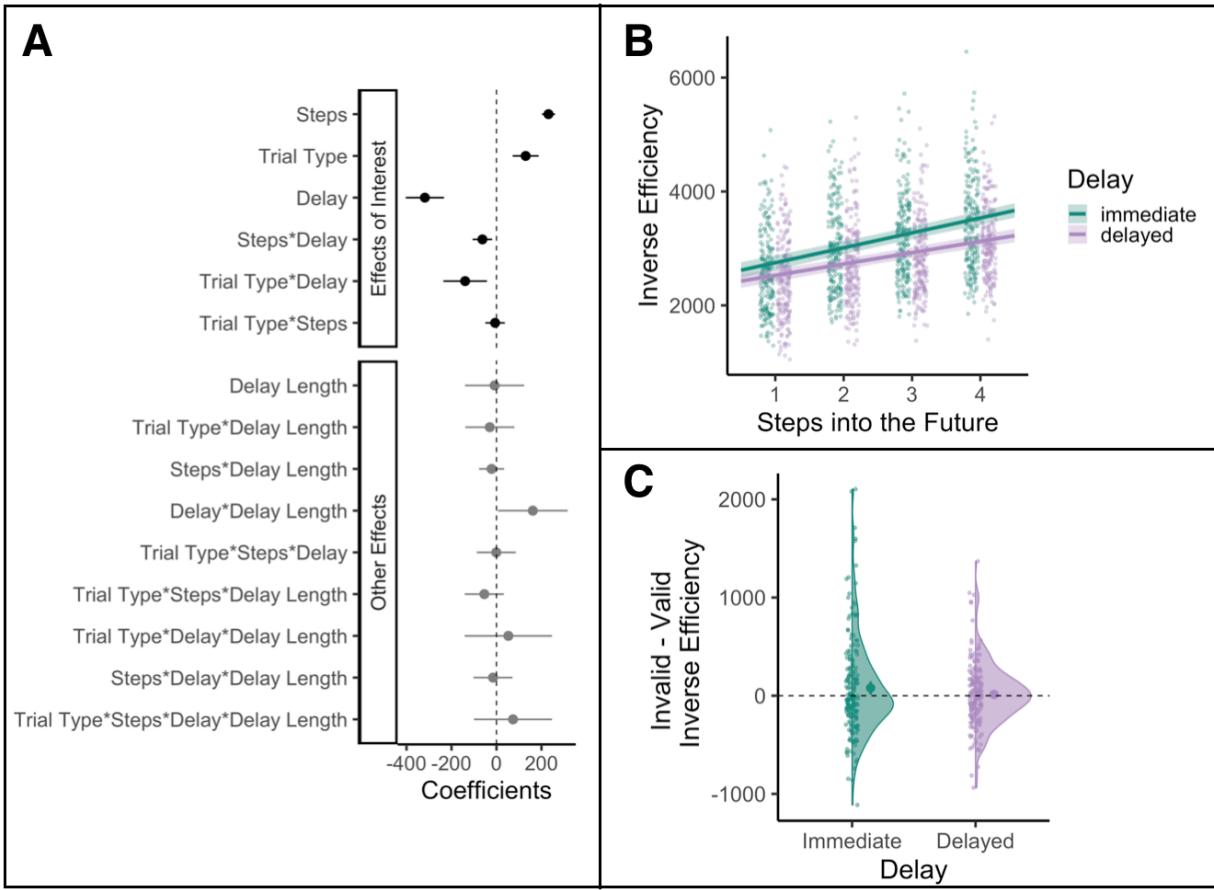


Figure 11.4: Anticipation Task performance in Experiment 2. (A) Coefficient estimates with 95% confidence intervals for the inverse efficiency model. The black points indicate effects of interest and the gray points indicate all other effects in the model. Predictors in all the models were effect coded. (B) Inverse efficiency as a function of steps into the future and delay. Higher inverse efficiency values indicate less efficient responses. Responses were more efficient in the delayed (vs. immediate) test, with the inverse efficiency benefit for the delayed (vs. immediate) test increasing with steps into the future. Green lines and error ribbons indicate model predictions with 95% confidence intervals for the immediate test; green points indicate each participant's inverse efficiency score for each step in the future for the immediate test. The same conventions are used for the delayed test, plotted in purple. (C) Inverse efficiency differences between valid and invalid trials as a function of delay. Participants were less efficient for invalid vs. valid trials at the immediate test (replicating the results of Experiment 1), but the efficiency difference between valid and invalid trials decreased after a delay. The dashed line at 0 indicates equally efficient responses on valid and invalid trials. Small points indicate the average inverse efficiency difference for each participant at each delay. The large points indicate the average inverse efficiency difference across participants with 95% confidence intervals.

1.4 Modeling Anticipation Strategies

Memory for temporal structure therefore changes over time to support multistep anticipation. But what strategies do participants use to anticipate upcoming information? Which strategies are most beneficial for behavior, and do they change with consolidation? One class of models predicts that a participant maintains an internal representation of the entire sequence in memory and explicitly rolls out the sequence, link by link, to anticipate upcoming events (Daw & Dayan, 2014). Another class predicts that a participant will build a representation for each item that incorporates cached information about future items, with stronger cached representations for events that are coming up sooner in the sequence (Dayan, 1993).

Drawing inspiration from these models, we tested whether participants were using a link-based strategy or a cue-based strategy. In a link-based strategy, each link between items in a sequence is represented in memory and participants sequentially “traverse” the sequence of links beginning at the cue item until they reach one of the probe items. A weak (vs. strong) link would be more difficult to traverse (increasing response time), and this slowdown would occur whenever the rollout from the cue to the nearest probe includes this link. In a cue-based strategy, information about future items becomes embedded in the representation of each cue, with closer upcoming items being more strongly represented. Identifying which of two probes is coming up sooner would be accomplished by directly comparing the probe items to this learned cue representation to determine which is more similar. Here, response times should depend on the quality of the cached representation at the given cue item, and trials for neighboring cues in the sequence could yield very different response times.

To adjudicate between these two strategies, we modeled response times using both a Link Model and a Cue Model, and tested which model better explained anticipation behavior.

1.4.1 Methods

Analyses

Using the data reported in Experiment 1 and Experiment 2, we created two different kinds of linear models in the R programming language. We modeled response times, as opposed to inverse efficiency, in these models because they required trial-by-trial measures and inverse efficiency is a summary statistic. We only modeled response times on correct trials to ensure that participants were successfully using learned information.

For each participant, we first modeled response times as a function of the links that participants crossed to get from the cue to the correct answer in the probe during each Anticipation Task trial. In this Link Model (Figure 1.5A), each link between scene categories in a given sequence was modeled separately, binarized by whether or not that link would be “traversed” from the cue to the correct answer on a given trial (0 = not used, 1 = used). Because the links between scene categories were different in each sequence, we created separate models for Sequence A and Sequence B. For Experiment 2, we also modeled response times separately for the immediate and delayed sessions. We used the following R based formula, with separate regressors for each link between adjacent scene categories in the circular sequence:

```
lm(RT~link1+link2+link3+link4+link5+link6+link7+link8+link9+link10, data, subset =  
(correct == 1))
```

We next modeled each participant’s response times as a function of the scene category that they were cued with during each trial in the Anticipation Task. This model assumed that increasingly distant future states would be increasingly difficult to access (consistent with temporal discounting; Gershman et al., 2012) but that the degree of difficulty could be different

for different cues. In this Cue Model (Figure 1.5A), each scene category in a given sequence was modeled separately, coded by whether or not that category was the cue on a given trial and, if it was cued, how many steps into the future the correct answer was (0 = not cued, 1 = cued on a 1-step trial, 2 = cued on a 2-step trial, 3 = cued on a 3-step trial, 4 = cued on a 4-step trial). We again modeled response times separately for Sequence A and Sequence B (because their future states differed) and, in Experiment 2, separately for the immediate and delayed test. We used the following R based formula, with separate regressors for each scene category:

```
lm(RT~cue1+cue2+cue3+cue4+cue5+cue6+cue7+cue8+cue9+cue10, data, subset = (correct == 1))
```

We assessed goodness of fit for both the Link and the Cue models by creating null models for each participant, in which the data were fit to a model in which the cue identities were randomly shuffled. We created 100 null models for each participant by repeating this procedure 100 times. Next, we calculated the R² from the real model and each of the 100 null models for each participant. We then created a model fit score for each participant and each sequence by calculating the difference between the R² of the real model and the average R² of the null models. To test whether the Link and Cue model fits were significantly better than the permuted model fits, we conducted one-sample t-tests comparing the model fit score for each model, averaged across sequences, to 0.

Next, to determine whether the Link or the Cue model provided a better fit to response times and whether this differed by delay, we created a model difference score by subtracting each participants' Cue Model fit score from their Link Model fit score, separately for each sequence. We then predicted this model difference score as a function of sequence (Experiment 1), or delay, delay length, their interaction, and sequence (Experiment 2). In these models, the

intercept term would provide evidence for better Link or Cue model fits overall across our sample, whereas the delay term would provide evidence for a shift in model fit with consolidation. We used the following R based formula:

*lmer(modDiff~1+delay*delayLength+sequence+(1|participant), data)*

Finally, to determine which strategy yielded superior behavioral performance, we investigated whether individual differences in participants' model difference scores predicted accuracy on the anticipation task. We focused on accuracy to have a dependent measure that is independent from response times, which were the models' outcome variable. We predicted participants' average accuracy as a function of their model difference score and sequence (Experiment 1) or model difference score, delay, delay length, their interactions, and sequence (Experiment 2). We used the following R based formula:

*lmer(accuracy~modDiff*delay*delayLength+sequence+(1|participant), data)*

1.4.2 Results

We first tested whether our Link Model and Cue Model provided good fits to participants' response times in the Anticipation Task by comparing their model fit scores (model R² vs. null models' R², see Methods) to 0. In Experiment 1, the Link Model provided a better fit to response times than the null models (mean R² difference: 0.011; t(99) = 2.226, p = 0.028). The Cue Model, however, was not better than the null models (mean R² difference: -0.002; t(99) = -0.53, p = 0.594). In Experiment 2, the Link Model provided a better fit to response times than the null models in both Session 1 (mean R² difference: 0.0179; t(203) = 5.106, p = 0.0000007) and Session 2 (mean R² difference: 0.0182; t(203) = 7.865, p < 0.000000001). Contrary to

Experiment 1, the Cue Model also provided a better fit to response times than the null models for both Session 1 (mean R2 difference: 0.0104; $t(203) = 3.936$, $p = 0.0001138$) and Session 2 (mean R2 difference: 0.0105; $t(203) = 5.470$, $p = 0.0000001$).

We next investigated whether the Link or Cue model performed better and whether the superior model changed with delay. The Link model provided a better fit to response times than the cue model in both Experiment 1 ($\beta = 0.014$, 95% CI [0.006, 0.0215], $p = 0.0006$; Figure 1.5B) and Experiment 2 ($\beta = 0.008$, 95% CI [0.004, 0.012], $p = 0.00003$; Figure 1.5B). This superiority of the Link model did not change across sequences (Experiment 1: $\beta = -0.01$, 95% CI [-0.025, 0.006], $p = 0.21$; Experiment 2: $\beta = -0.0001$, 95% CI [-0.005, 0.005], $p = 0.948$), nor (in Experiment 2), across delay ($\beta = -0.0008$, 95% CI [-0.008, 0.006], $p = 0.796$), delay length ($\beta = 0.0005$, 95% CI [-0.007, 0.008], $p = 0.883$), or their interaction ($\beta = -0.009$, 95% CI [-0.022, 0.004], $p = 0.172$).

Finally, we assessed whether individual differences in strategy use, indexed by superior fits for the Link vs. Cue Model, predicted average accuracy on the Anticipation Task, and whether this effect changed with delay. In Experiment 1, differences in the Link vs. Cue model difference score did not predict accuracy ($\beta = 0.026$, 95% CI [-0.166, 0.219], $p = 0.787$; Figure 1.5C). In contrast, model difference scores did predict accuracy in Experiment 2 ($\beta = 0.165$, 95% CI [0.006, 0.323], $p = 0.042$; Figure 1.5C). Importantly, there was a model difference score by delay interaction ($\beta = 0.366$, 95% CI [0.066, 0.666], $p = 0.017$; Figure 1.5C), such that higher Link vs. Cue model difference scores did not predict accuracy in the immediate test session, replicating Experiment 1 ($\beta = 0.074$, 95% CI [-0.09, 0.239], $p = 0.378$), but did predict accuracy in the delayed test session ($\beta = 0.299$, 95% CI [0.052, 0.547], $p = 0.018$). Delay length did not interact with model difference score ($\beta = -0.006$, 95% CI [-0.313, 0.300], $p = 0.968$) nor was

there a three-way interaction between delay length, model difference score, and delay ($\beta = 0.2333$, 95% CI [-0.368, 0.833], $p = 0.448$).

Thus, the Link Model more effectively captured participants' response times than the Cue Model, regardless of delay. Higher Link vs. Cue Model fits also predicted accuracy in the Anticipation Task, but only after consolidation.

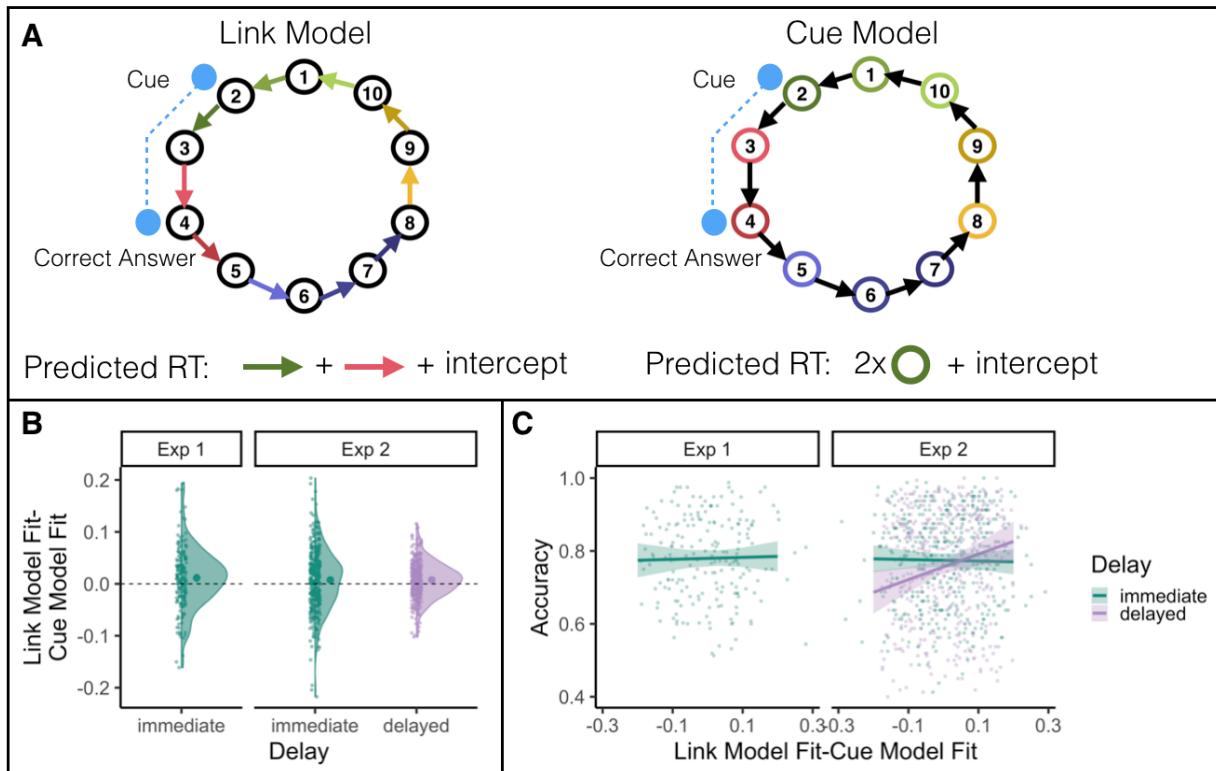


Figure 12.5: Link and Cue Model Schematic and Results. (A) Model schematic for the link and cue models. Open circles represent scene categories and lines represent the link between the categories in the sequence. In the link model (left), trial-wise response times were predicted as a function of the links between the cue image and the correct answer. In this example, the predicted response time depends on link-specific values for the green link (from category 2 to 3) and the red link (from category 3 to 4). In the cue model (right), trial-wise response times were instead assumed to vary as a function of the cued category, with linearly increasing response times based on the distance to the correct answer. In this example, the predicted response time depends on a cue-specific value for the green cue (category 2) modulated by the number of steps to the correct answer (here, two steps). (B) Link model fits were better than cue model fits, as indicated by a positive difference score; this difference was significant across both experiments and both test delays in Experiment 2 (immediate and delayed). The dashed line at 0 indicates equal cue and link model fits. Small points indicate each participant's model fit difference for each experiment and delay. The large points indicate the average model fit difference across

participants with 95% confidence intervals. (C) Higher link (vs. cue) model fits predict accuracy on the anticipation task, but only in the delayed (vs. immediate) test. Lines and error ribbons indicate model predictions with 95% confidence intervals. Small points indicate each participant's model fit difference at each delay.

1.5 General Discussion

Memory for temporal structure is adaptive because it allows us to anticipate what is likely to happen in the future. Indeed, prediction is thought to be a primary cognitive and neural function (Clark, 2013; Friston, 2005), and offline periods of memory consolidation may play a crucial role in extracting regularities that will be useful for prediction (Hobson & Friston, 2012). Our findings show how memories for such regularities are shaped by consolidation, and provide an important link in understanding the adaptive function of memory.

Across two experiments, we found that multistep anticipation became more efficient with consolidation, particularly for further events. Anticipated events contained representations of task-irrelevant perceptual features, but these perceptual features had less influence on behavior after consolidation. Finally, maintaining a link-based, rather than a cue-based, model of the sequence after consolidation benefitted multistep anticipation. Overall, these results shed light on how memories adaptively shift to prioritize temporal structure at the cost of perceptual details.

Our findings are consistent with, and build upon, influential theories of memory consolidation, which posit that memories shift from detailed to schematic over time (Robin & Moscovitch, 2017; Sekeres et al., 2018; Winocur et al., 2010). Schematized representations should support memory for the high-level structure of experiences (McClelland et al., 1995) but may lack perceptual detail. Indeed, we found that, with consolidation, perceptual details became more weakly represented in memory. They nevertheless still exerted an influence on behavior, suggesting that perceptual details were not entirely forgotten after consolidation (Gilboa & Moscovitch, 2021; Robin & Moscovitch, 2017). However, it is important to note that, in our

experiments, temporal structure was task-relevant whereas perceptual features were incidental to the task. Perceptual details may have been more strongly maintained in memory if they were task-relevant (Schapiro, McDevitt, et al., 2017).

In contrast to perceptual details, information about temporal structure was more efficiently accessed after a delay. Our findings of efficient access of multistep sequences after consolidation extends prior work showing memory improvements after consolidation for category structure (Schapiro, McDevitt, et al., 2017) and short timescale statistical regularities (Durrant et al., 2011). Extending consolidation-related benefits to multistep anticipation allows insights into how behavior may become increasingly adaptive over time, by allowing fast access to anticipated future events — particularly those that are further in the future, suggesting increasingly compressed memory for temporal structure.

To further probe how temporal structure is represented in memory, we drew inspiration from reinforcement learning models to determine what types of internal models were most useful for anticipation, both immediately and at a delay. Individuals tended to use a link-based strategy (akin to a model-based representation) rather than a cue-based strategy (akin to a successor representation), and this preference for the link versus cue model was present in both the immediate session and after consolidation. Critically, maintaining a link (versus cue) model after consolidation predicted accuracy on the anticipation task. This finding dovetails with predictions of model-based learning: such models are thought to lead to highly accurate representations post-consolidation because they store a fully connected internal model of an environment (Y. Liu et al., 2021; Wimmer et al., 2023). However, such models are computationally expensive because they require an individual to traverse individual links between states (Daw & Dayan, 2014). Thus, we propose that consolidation processes may be particularly adaptive because they allow

the brain to maintain and stabilize — and perhaps make more efficient — computationally intensive representations.

Although we found that the link model was the optimal internal model in our task, other tasks might be better solved with a cue model. Past studies have shown a preference for successor representation strategies (conceptually similar to our cue model) when short sequences end in monetary reward (Momennejad et al., 2017). A cue model preference may therefore emerge if sequences end in a goal location or a reward — whereas in our task, the sequences were circular without a salient end-point. Further, including other features in the sequence, such as event boundaries, could lead to a model in which participants cache future states within an event and skip between event boundaries, rather than traversing each link separately (Michelmann et al., 2023).

Despite finding robust differences in memory for temporal structure between the immediate and delayed sessions, we failed to find differences after a 1-day versus 1-week delay. This may have occurred because participants repeatedly anticipated the same sequences across the experiment. Repeated retrieval may quicken memory consolidation (Antony et al., 2017). Indeed, repeated retrieval, compared to restudy, increases behavioral markers of semanticization after consolidation (Lifanov et al., 2021), and repeatedly testing regularities in a statistical learning task reduces forgetting of explicit memories over the course of consolidation (H. Liu et al., 2023). Repeated anticipation may have likewise hastened schematization processes, thus reducing differences between our 1-day and 1-week conditions. Another possibility is that participants who forgot the sequences over the 1-week delay were likely to drop out of the study, leading to the higher attrition rates observed in the 1-week delay and similar performance between the two delay conditions.

In summary, we showed that consolidation leads to efficient access of temporal structure in the service of multistep anticipatory judgements, but at the cost of perceptual details. Furthermore, post-consolidation maintenance of internal models that linked experienced events predicted accurate anticipation, raising the intriguing possibility that consolidation may increase the efficiency, accuracy, or stability of computationally intensive strategies. Together, our work shows how memories are maintained and transformed over time to prioritize representations of temporal structure at the expense of incidental perceptual features — in turn allowing us to anticipate likely upcoming events and behave adaptively in a complex world.

Chapter 2: The brain hierarchically represents the past and future during multistep anticipation

Hannah Tarder-Soll, Christopher Baldassano*, & Mariam Aly*

*These authors contributed equally to this work

2.1 Introduction

Memory allows us to use past experience to generate expectations about the future.

Integration of past information to predict future events enables efficient planning and flexible behavior in complex environments (Behrens et al., 2018; Epstein et al., 2017; Momennejad, 2020; Stachenfeld et al., 2017; Tolman, 1948) and has been proposed to be a primary function of memory systems (Buckner, 2010; Stachenfeld et al., 2017) and of the brain itself (Clark, 2013; Friston, 2005). For predictions to usefully impact behavior, they should be represented on multiple timescales, allowing us to anticipate not just immediately upcoming events but also events further in the future. Furthermore, predictions that are relevant for the current context should be flexibly prioritized over those that are less relevant. For example, when riding the subway, it would be useful to anticipate multiple stations ahead on the relevant line, but we need not anticipate upcoming stops on other lines passing through the same stations. Such context-specific prediction may be supported by leveraging memories of past stops, which contextualize where we are in the present. How does contextual relevance modulate prediction of future events? Here, we aimed to test three central hypotheses. First, that the brain will flexibly anticipate events at multiple timescales in the future; second, that the future and the past will be represented simultaneously in the same brain regions; and third, that anticipatory representations will be prioritized for events that are contextually relevant.

To address these questions, we drew on prior research showing anticipatory signals across the brain, particularly in memory and sensory systems (Buckner, 2010; de Lange et al., 2018; Lee et al., 2021; Lisman & Redish, 2009; Ouden et al., 2010; Summerfield & Egner, 2009). For example, predictions about upcoming items or locations in a sequence are represented in visual cortex (Gavornik & Bear, 2014; Hindy et al., 2016; Kok et al., 2012, 2014) and

hippocampus (Aitken & Kok, 2022; Brown et al., 2016; Hindy et al., 2016; Kok & Turk-Browne, 2018; Schapiro et al., 2012), suggesting coordination between these regions in memory-based prediction of visual stimuli (Hindy et al., 2016). Although earlier research on prediction typically focused on one or a few brain regions (Aitken & Kok, 2022; Gavornik & Bear, 2014; Hindy et al., 2016; Kok & Turk-Browne, 2018; Schapiro et al., 2012) and predictions about immediately upcoming events (Hindy et al., 2016; Kok et al., 2012; Kok & Turk-Browne, 2018), more recent work has shown that the brain represents anticipatory signals at multiple timescales simultaneously, with shorter timescales of prediction in more posterior regions and successively longer anticipatory timescales in progressively more anterior regions (Brunec & Momennejad, 2022; Lee et al., 2021). For example, during repeated viewing of a movie clip, posterior regions like visual cortex primarily represent the current moment, while anterior regions such as the insula represent upcoming events multiple seconds into the future (Lee et al., 2021). These findings of multistep anticipatory signals are generally consistent with computational theories that the brain builds models of the world that cache temporal information about successive events, with different predictive timescales in different brain regions (i.e., multi-scale successor representations; Brunec & Momennejad, 2022; Momennejad & Howard, 2018).

This research on multiscale anticipation in the brain complements earlier work showing hierarchical representations of past states (Hasson et al., 2015). Mirroring the predictive hierarchy for future states (Lee et al., 2021), information from the past lingers in the brain during ongoing experience, with shorter timescales of past information represented in posterior regions and longer timescales in anterior regions (Aly et al., 2018; Baldassano et al., 2017; Hasson et al., 2008; Lerner et al., 2011). In the hippocampus specifically, temporal coding in the form of sequence reactivation extends both into the past and the future (Diba & Buzsáki, 2007; Foster &

Wilson, 2006; Ólafsdóttir et al., 2018). Furthermore, the brain's representations of the past and future can be flexibly modulated based on task demands (Wimmer et al., 2020). Although this work suggests that the brain may represent both anticipated and past events, these prior studies did not test whether forward and backward representations of temporally extended structure existed simultaneously in the same brain regions. We therefore examined whether the brain contains bidirectional representations of the past and future, with the scale of these representations varying systematically across the brain.

For the brain's representations of temporal structure to be adaptive for behavior, they should flexibly change depending on context. Recent work in humans has shown context-specific patterns of activity in the hippocampus during goal-directed planning of future trajectories, suggesting that anticipation of temporally structured experience is specific to the upcoming items in a given context (Crivelli-Decker et al., 2023). However, it remains unknown whether contextual modulation of temporal structure representations is specific to planning trajectories in the forward direction or if contextual relevance also modulates representations of the past.

In the present study, we investigated how context-specific temporal structure is represented in the brain during a novel multistep anticipation task. Participants learned, in immersive virtual reality, four temporally extended sequences of eight environments each (Figure 2.1). Critically, pairs of sequences contained the same environments in a different order, requiring individuals to flexibly anticipate environments based on the current sequence context. Sequences were circular, such that environments were temporally predictable multiple steps into the future and the past regardless of location in the sequence. This allowed us to test whether temporal structure in both the prospective and retrospective direction is automatically

represented in the brain even if only future states are task-relevant. Following sequence learning, participants were scanned with fMRI as they anticipated upcoming environments one to four steps into the future in a given (cued) sequence (Anticipation Task; Figure 2.2). Using multivariate pattern similarity analyses in visual cortex, hippocampus, insula, and across the brain, we determined the extent to which temporal structure was (1) hierarchically represented along a posterior to anterior gradient, with further-reaching representations in more anterior regions; (2) represented in a bidirectional manner, with simultaneous representations of future and past environments within a context; and (3) modulated by context, with prioritized representations of nearby environments in the cued vs uncued sequence.

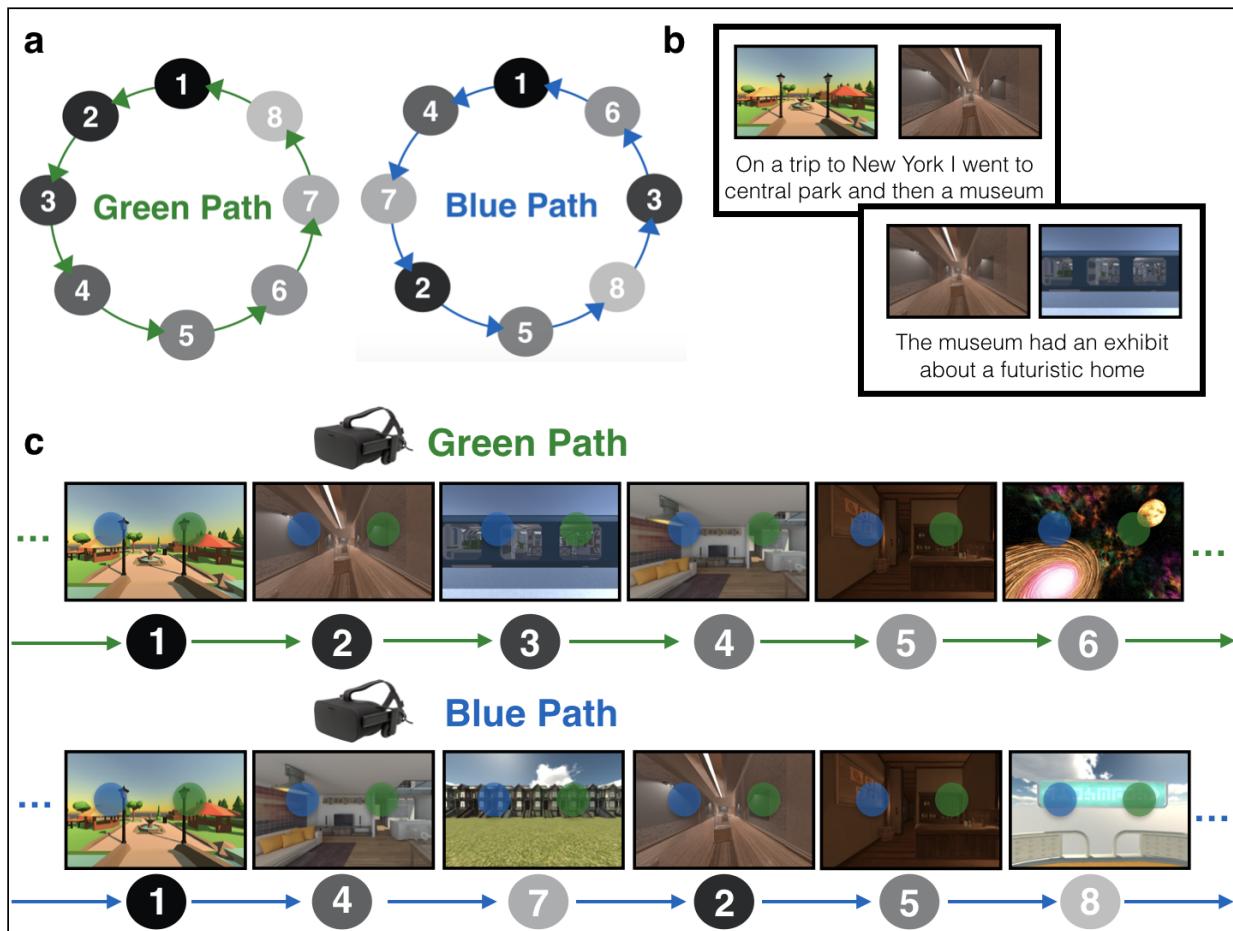


Figure 2.1: Sequence structure and behavioral training. (a) Sequence structure. Participants learned sequences of eight environments, indicated by the gray nodes. The green path and the blue

path consisted of the same environments in a different order. The sequences were constructed to be as distinct as possible: for a given environment the two preceding and two succeeding environments were different across the sequences. Participants learned four sequences in total: one green and blue path with a set of eight environments, and another green and blue path with a different set of eight environments. Only one green and blue path is depicted here for illustrative purposes. **(b) Story Generation.** To learn the sequence of environments, participants generated stories for each path to link the environments in order. Participants were told to link the final environment back to the first environment to create a loop. **(c) Virtual Reality Training.** Participants then explored the environments in immersive virtual reality in the green path order and the blue path order while rehearsing their stories. In a given environment, a green and blue sphere would appear. These spheres, when touched, teleported that participant to the next environment in the corresponding (green or blue) sequence. Participants then recalled the order of each of the four sequences (not shown).

2.2 Results

Anticipation Task Performance

Participants performed effectively on the Anticipation Task (**Figure 2.2a**), correctly choosing the closer of the two probe images, relative to the cued image and path, 86.86% of the time ($sd = 0.08$), which was significantly higher than chance performance of 50% ($t(31) = 61.08$, $p = < 0.00001$). There was a trend toward higher accuracy in Map A (i.e. the first learned map) compared to Map B ($\beta = -0.283$, 95% CI = [-0.576, 0.001], $p = 0.058$). Accuracy did not vary by path (Green vs Blue, $\beta = -0.124$, 95% CI = [-0.458, 0.209], $p = 0.464$), nor was there a map by path interaction ($\beta = 0.268$, 95% CI = [-0.274, 0.81], $p = 0.333$). Response times on the Anticipation Task were not significantly influenced by map ($\beta = 0.029$, 95% CI = [-0.014, 0.072], $p = 0.185$), path ($\beta = 0.038$, 95% CI = [-0.009, 0.085], $p = 0.127$), or their interaction ($\beta = -0.052$, 95% CI = [-0.134, 0.029], $p = 0.209$).

We next determined how performance on the Anticipation Task varied by steps into the future (i.e. how many steps the correct probe was from the cue image on the cued path; also see Tarder-Stoll et al., 2023). Steps into the future had a trending impact on accuracy ($\beta = -0.15$, 95% CI = [-0.33, 0.027], $p = 0.096$; **Figure 2.2b**), with an average difference in accuracy of

3.9% between one-step and four-step trials. Steps into the future robustly impacted response time ($\beta = 0.13$, 95% CI = [-0.103 0.149], $p < 0.000001$; **Figure 2.2b**). Responses were on average 126 ms slower for each step into the future, with an average difference of 380 ms for one-step and four-step trials. Together, this suggests that participants performed accurately on the Anticipation Task, but were slower to anticipate upcoming environments that were further into the future.

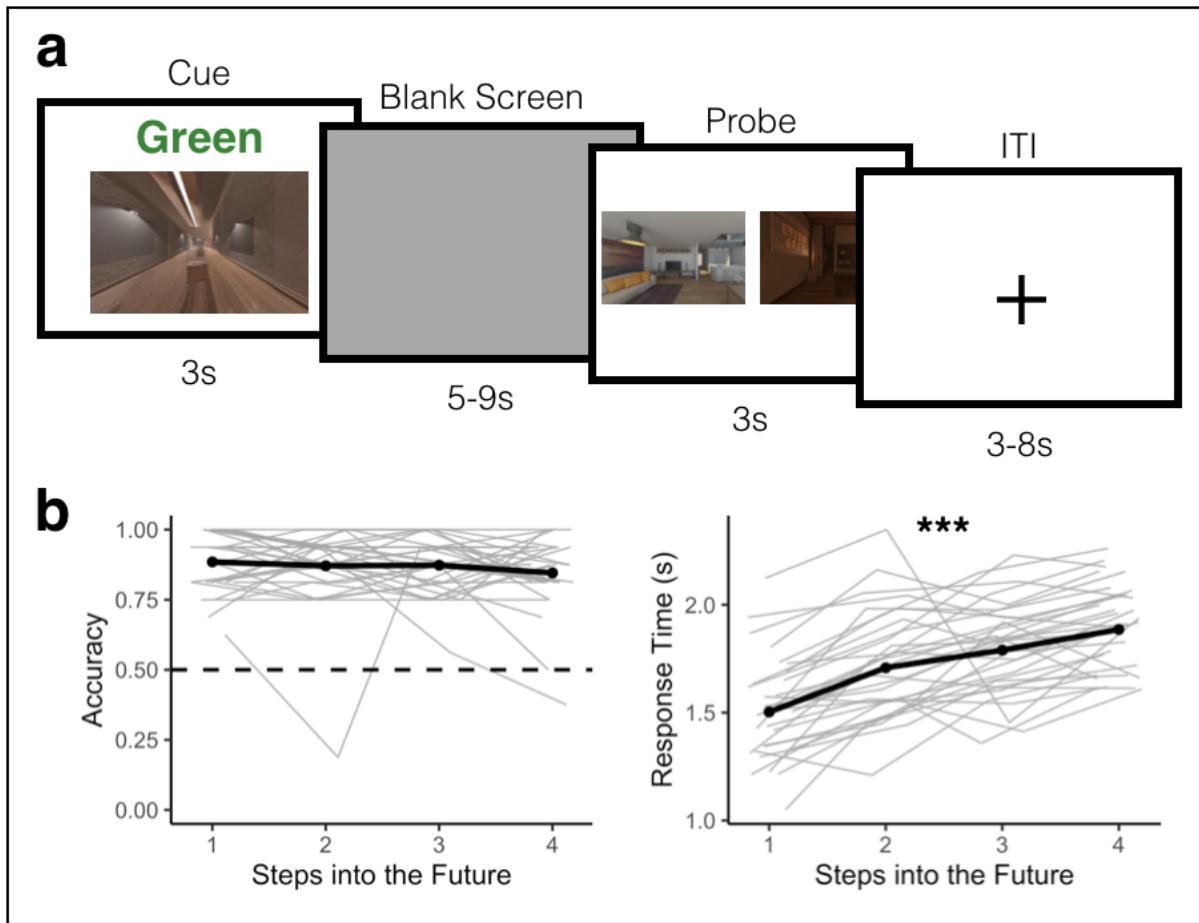


Figure 2.2: Anticipation task and behavioral performance. (a) Anticipation Task. Participants returned one day after behavioral training and completed the anticipation task inside the MRI scanner. Participants were cued with a 2D image of an environment from one of the sequences along with a path cue (Green or Blue) for 3 seconds. They then saw a blank screen for a variable duration of 5 to 9 seconds during which they were told to prepare their response. Participants were then probed with two images of upcoming environments and had 3 seconds to indicate which of the two environments was coming up sooner in the cued sequence, relative to the cue image. The

correct answer could be 1 to 4 steps away from the cue image. **(b)** Behavioral Performance. Participants accurately anticipated upcoming environments in the cued sequence. Accuracy did not significantly differ across steps into the future (left). Response time, however, was significantly slower for further steps into the future (right). Pale gray lines indicate data for individual participants; the black line is the group mean. *** p < .0001

Bidirectional Representations of Temporal Structure in Hippocampus and Visual Cortex

For our MRI analyses, we first created conjunction ROIs by selecting voxels within visual cortex, hippocampus, and insula that reliably responded to distinct environments in the Localizer Task (Figure 2.3a-c). Next, we obtained the across-participant multivariate pattern of brain activity for each environment within each of our ROIs (Figure 2.3d). To investigate neural representations of temporal structure during multi-step anticipation, we calculated pattern similarity between (1) multivariate patterns of brain activity evoked during the Anticipation Task for each trial type (cue and path combination) for each participant and (2) the multivariate patterns of brain activity evoked during Localizer Task, averaged across the remaining participants, for each environment on the same map (Figure 2.4a; see Methods). We then ordered the resulting correlation values in the order of the cued path with the cued environment in the center, successors following the cue to the right of the center, and predecessors to the left of the center (Figure 2.4a; see Methods). Importantly, because the order of the sequences was randomized across participants, the across-participant multivariate pattern of activity during the Localizer Task cannot include information about successors for individual participants' sequences, resulting in a relatively pure measure of environment representations. Thus, this analysis allows us to determine the extent to which our regions of interest represented upcoming or preceding environments during the Anticipation Task, using activity pattern "templates" for each environment that were constructed to remove information about sequence structure.

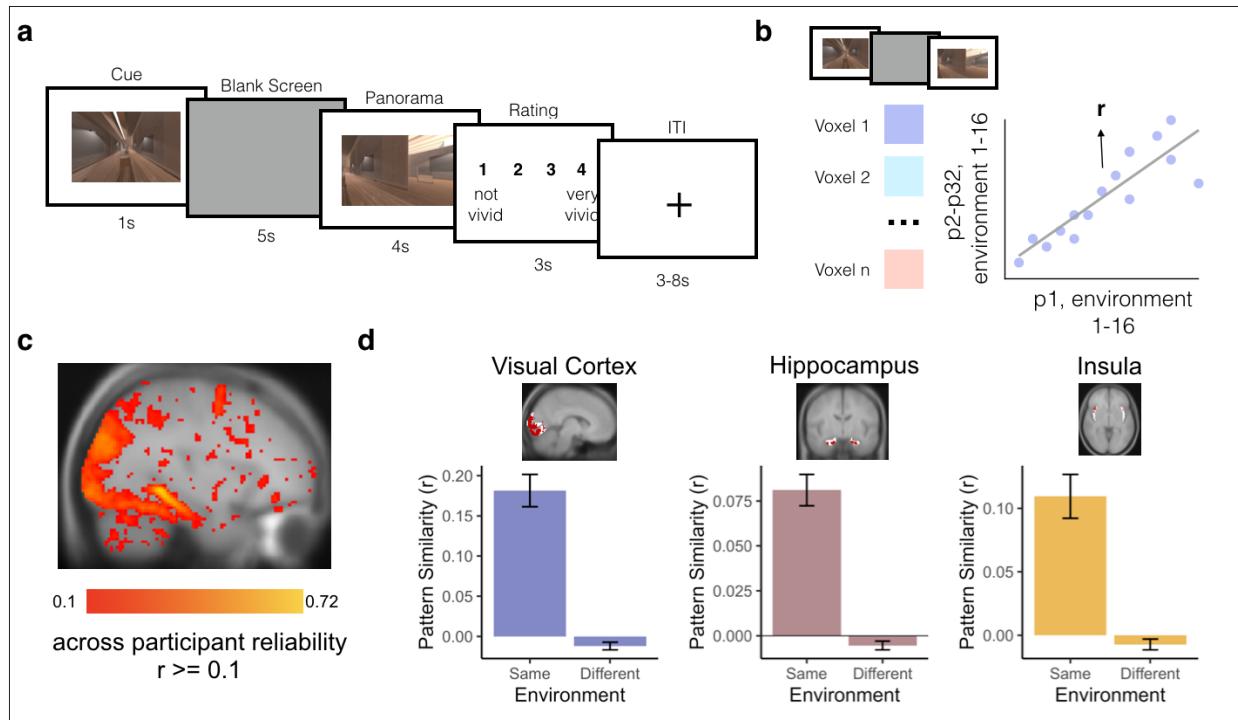


Figure 2.3: Template brain activity patterns for each environment. **(a)** Localizer Task. Participants completed a localizer task inside the MRI scanner at the end of the session. Participants were cued with a 2D image of an environment from the experiment for 1 second. They then saw a blank screen for 5 seconds during which they were told to imagine being inside the environment in VR. Next, they saw images of the environment from different angles for 4 seconds and were given 3 seconds to rate how well their imagination matched the actual images of the environment. **(b)** Across-participant analysis for identifying voxels that reliably discriminate between environments. We measured the activity of each voxel in each participant during the Localizer Task (combining the cue, blank screen, and panorama phases) for each of the 16 environments. Next, we obtained the Pearson correlation (r) in each voxel between a participant's responses to the 16 environments (e.g., P1) and the 16 average responses in the remaining participants (e.g., P2-P32). Averaging across all choices of the left-out participant, this yielded an across-participant reliability score for each voxel. **(c)** Whole brain map of voxels that reliably discriminate between environments. We only included voxels that had an environment reliability value of 0.1 or greater and were part of a cluster of at least 10 voxels in our subsequent analyses. **(d)** Environment reliability in ROIs. In visual cortex (left), hippocampus (middle), and insula (right), we selected the environment-reliable voxels (red) within each anatomically or functionally defined ROI (white). We then confirmed that the analysis successfully identified across-participant patterns of activity within the conjunction ROIs that were more correlated for the same environment than for different environments. Error bars indicate standard error of the mean.

We first examined whether the sequence structure was reflected in neural representations during the Anticipation Task. We tested two main questions: first, whether these representations were biased in the forward direction, suggesting stronger anticipation of future environments

rather than representations of past environments; and, second, whether anticipatory representations were further reaching in some brain regions relative to others. To do so, we fit an asymmetrical Gaussian curve to the ordered pattern similarity values (see Methods). The Gaussian similarity model has four parameters: amplitude, asymptote, forward and backward width (σ) (Figure 2.4b). The amplitude of the curve indicates the degree to which a brain region is representing the cue environment while it is on the screen. The asymptote is an indication of the representations of the environments furthest from the cue; if the asymptote is lower than baseline (defined as pattern similarity between the cue and all environments from the other map, henceforth referred to as different-map baseline), this suggests that the representations of further environments are suppressed. The forward and backward widths (σ) control the width of the curve, indicating how similarity to neighboring environments falls off with the number of steps in the forward and backward directions. Wider (vs. narrower) widths (σ) indicate that the brain region represents environments that are further away. If a brain region has a wide forward width (σ) but a narrow backward widths (σ), this indicates a bias towards representing upcoming environments, indicating anticipation, over retrospective representations of preceding environments.

In visual cortex, the Gaussian model fit significantly outperformed the fits of null models in which the order of the pattern similarity values was shuffled ($p < 0.001$). The amplitudes of participants' Gaussian fits were significantly higher than the different-map baseline, indicating that the cue environment was represented while it was on the screen (mean = 0.091, standard deviation = 0.029; $t(31) = 18.01$, $p < 0.000001$; Figure 2.4c). The asymptote was significantly lower than the different-map baseline, suggesting that other environments surrounding the cue were suppressed (mean = -0.014, standard deviation = 0.009; $t(31) = -5.97$, $p = 0.000001$; Figure

2.4c). The backward and forward widths (σ) were 0.712 steps and 0.634 steps, respectively, and were not significantly different from each other ($V(31) = 195.00$, $p = 0.203$), suggesting that representations were not biased toward one direction over the other.

Turning to representations in hippocampus, we also found that the asymmetric Gaussian provided better fits than a permuted model ($p = 0.029$). In the hippocampus, similar to visual cortex, the amplitude of the Gaussian fit was significantly higher than the different-map baseline (mean = 0.0185, standard deviation = 0.030; $t(31) = 2.959$, $p = 0.005$; Figure 2.4d) and the asymptote was significantly lower than the different-map baseline (mean = -0.007, standard deviation = 0.014; $t(31) = -3.476$, $p = 0.001$; Figure 2.4d). The backward and forward widths (σ) were 2.313 steps and 1.782 steps, respectively, and were again not significantly different from each other ($V(31) = 231.00$, $p = 0.548$), suggesting that the hippocampus represented forward and backward environments similarly. In contrast to hippocampus and visual cortex, Gaussian representations in insula were not significantly different from chance ($p = 0.139$).

We statistically compared the width parameters in both regions, to determine whether hippocampus represented more distant environments surrounding the cue and whether this differed in the forward vs. backward direction. Hippocampus had significantly wider widths than visual cortex ($\beta = 1.379$, 95% CI = [0.559, 2.199], $p = 0.0023$; Figure 2.4e), indicating that hippocampus represented more distant environments. There was no effect of direction ($\beta = -0.246$, 95% CI = [-0.935, 0.443], $p = 0.489$), nor a region by direction interaction ($\beta = -0.569$, 95% CI = [-1.746, 0.606], $p = 0.346$). Together, this suggests that hippocampus had further reaching representations of temporal structure than visual cortex. However, representations were not biased toward the forward, compared to the backward, direction in either region nor were there differential directional biases across regions.

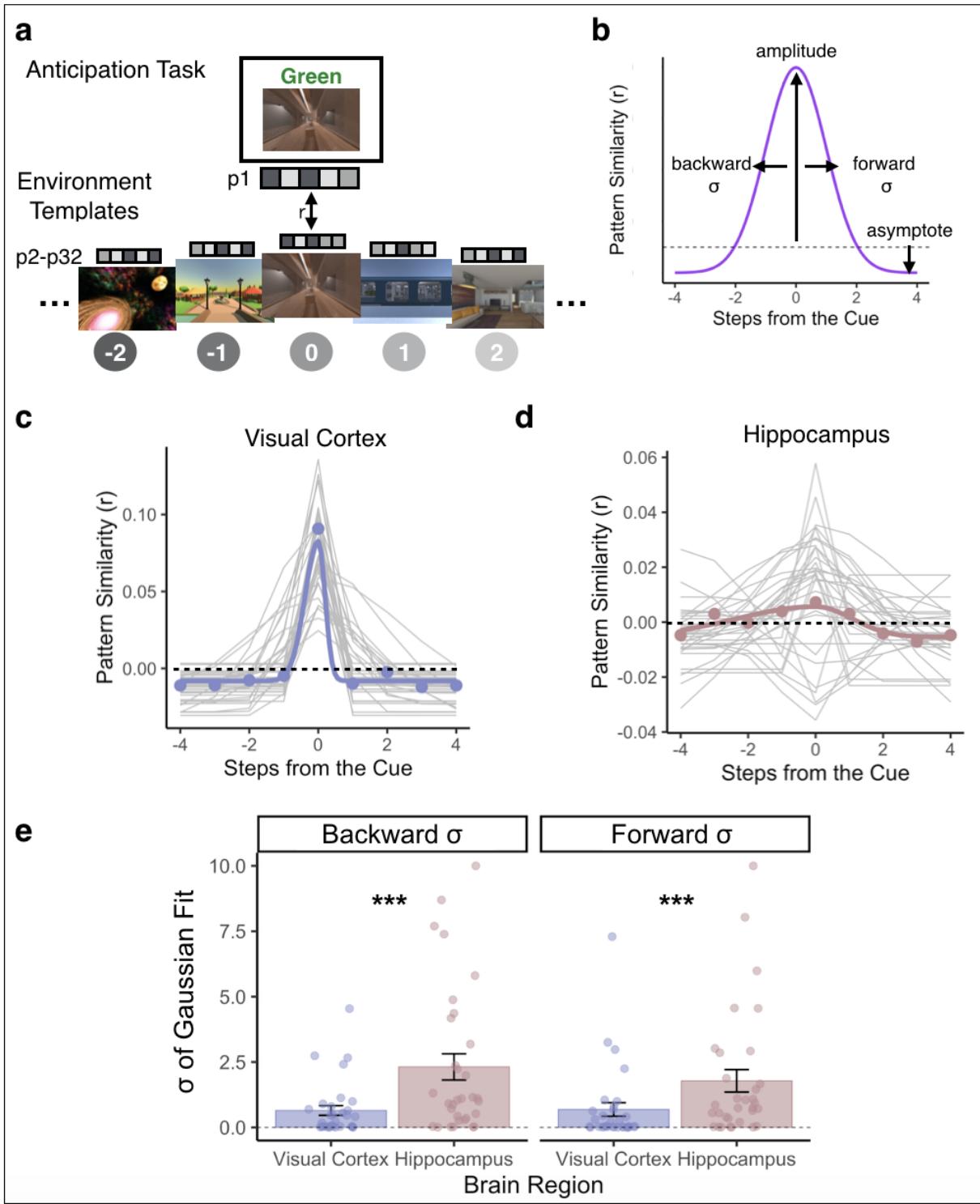


Figure 2.4: Bidirectional and graded representations of temporal structure in hippocampus and visual cortex. (a) Schematic depiction of Gaussian analysis. We obtained the correlation between a given participant's (e.g., P1) cue screen activity pattern for each trial of the Anticipation Task and the remaining participants' (e.g., P2-P32) averaged patterns of activity for each of the environment templates on the cued path. We then ordered the resulting pattern similarity values

with the cue in the center and fit an asymmetrical Gaussian curve. **(b)** Gaussian similarity model. The amplitude of the curve is an indication of the degree to which a brain region is representing the cue environment while it is on the screen. The widths (σ) control the width of the curve, indicating how similarity to neighboring environments falls off with the number of steps in the forward and backward directions. Wider (vs. narrower) widths (σ) indicate that the brain region represents environments that are further away. The asymptote quantifies the representations of environments that are not captured by the width of the Gaussian; if the asymptote is lower than the dashed line (other-map baseline) this suggests that these environments are suppressed. **(c)** Gaussian curve in visual cortex. Visual cortex strongly represented the cue environment while it was on the screen (above-baseline amplitude) and did not strongly represent nearby environments (narrow forward and backward widths (σ)), instead showing suppression of environments other than the cue (below-baseline asymptote). Purple lines and points indicate the group-average pattern similarity values and Gaussian curve. Gray lines indicate each participant's Gaussian curve. **(d)** Gaussian curve in the hippocampus. The hippocampus represented the cue environment while it was on the screen (above-baseline amplitude), represented nearby environments in a graded manner, in both the forward and backward direction (wide forward and backward widths (σ)), and suppressed environments that were furthest away (below-baseline asymptote). Pink lines and points indicate the group-averaged pattern similarity values and Gaussian curve. Gray lines indicate each participant's Gaussian curve.* p < .05 **(e)** Widths (σ) were wider in the hippocampus than in visual cortex, indicating representations of environments further away. Widths (σ) did not significantly differ between the forward vs. backward directions in either visual cortex or the hippocampus. Bars indicate average width (σ) across participants, error bars indicate standard error of the mean, and small, transparent points indicate each participant's width (σ) estimates. *** p < .001

Hippocampal suppression of environment representations predicts response time costs

We next sought to examine whether neural representations of temporal structure were related to behavioral performance on the Anticipation Task. We reasoned that suppression of environments surrounding the cue, indicated by the asymptote parameters of the model fits, should interfere with the generation of long timescale predictions: more suppression should be associated with more response time costs for accessing future environments. To test this, we obtained the Spearman rank-order correlation between participants' Gaussian asymptote, separately for visual cortex and hippocampus, and the slope of their response times across steps into the future. We hypothesized that a lower (more suppressed) asymptote would be related to steeper response time slopes across steps into the future, indicating a larger response time cost when making judgements about further environments. We expected this relationship to be

stronger in hippocampus vs visual cortex, because visual cortex showed suppression of even the most nearby environments (Figure 2.4c).

In visual cortex, there was no relationship between the asymptote of the Gaussian curve and the response time slope across steps into the future ($\rho = -0.050$, $p = 0.784$, Figure 2.5a). However, in the hippocampus there was a significant negative correlation between the asymptote of the Gaussian curve and response time slope ($\rho = -0.362$, $p = 0.042$, Figure 2.5b), suggesting that suppression of environments surrounding the cue was related to response time costs for anticipating further environments.

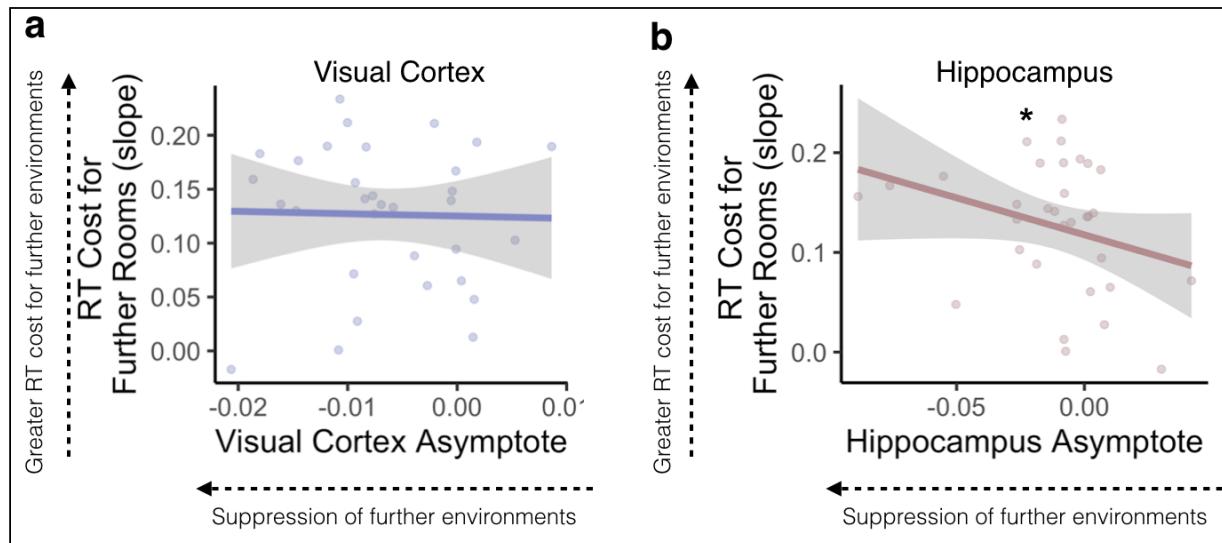


Figure 2.5: Suppression of further environments in the hippocampus is related to response time costs. In visual cortex (a), lower asymptotes, indicating suppression of non-cued environments, were not related to the slope of response times across steps into the future. In the hippocampus (b), lower asymptotes were related to steep response time slopes, suggesting that participants were slower to respond to further environments when those environments were relatively suppressed. Lines and gray error ribbons indicate the correlation with 95% confidence intervals; points indicate each participant's asymptote and response time slope. * $p < .05$

Temporal structure is hierarchically organized within visual regions

We next conducted an exploratory searchlight analysis to determine which brain regions outside visual cortex and hippocampus exhibited Gaussian representations (see Methods). Our

searchlight analysis revealed significant Gaussian representations across voxels in the visual system (Figure 2.6a), including regions that code for scene information such as parahippocampal place area (PPA) and the retrosplenial cortex (RSC) (Epstein, 2008). There were no differences in backward vs forward widths (σ) in any voxel in the searchlight, suggesting bidirectional representations of temporal structure across the visual system.

Prior work has shown within-region functional differences in posterior vs anterior ventral visual regions, including PPA and RSC (Baldassano et al., 2013; Steel et al., 2021). Posterior aspects of ventral visual regions may play a larger role in scene perception while anterior aspects may represent scene memories. Based on these differences, we hypothesized that there may be hierarchical representations of temporal structure within PPA and RSC, with further reaching representations (as indicated by wider vs. narrower widths (σ)) in successively more anterior aspects of regions. To test for hierarchical organization of temporal structure, we obtained the correlation for each participant between (1) the averaged forward and backward widths (σ) of the Gaussian curve in each voxel and (2) that voxel's y-coordinate, indicating its position along the posterior-anterior axis. We then tested whether these correlations were different from 0 across participants. We conducted the same analysis for the amplitude and asymptote, to determine if the representation of the cued environment and suppression of nearby environments also changed along the posterior-anterior axis.

There was a significantly positive correlation between width (σ) and y-coordinate, indicating that widths became progressively wider in progressively more anterior aspects, in both PPA ($t(31) = 2.424, p = 0.021$) and RSC ($t(31) = 2.638, p = 0.013$). There was a negative correlation between amplitude and y-coordinate in PPA ($t(31) = -2.636, p = 0.013$), but not RSC ($t(31) = -0.550, p = 0.586$). Finally, there was no correlation between asymptote and y-coordinate

in either PPA ($t(31) = 1.721$, $p = 0.095$) or RSC ($t(31) = 1.047$, $p = 0.303$). This suggests a within-region hierarchical organization of representations in the visual system, such that more anterior (vs posterior) aspects of PPA and RSC represent further reaching predictions. Importantly, because amplitude was not correlated with y-coordinate in RSC, this suggests that further reaching anticipation is not necessarily a consequence of reduced processing of the present.

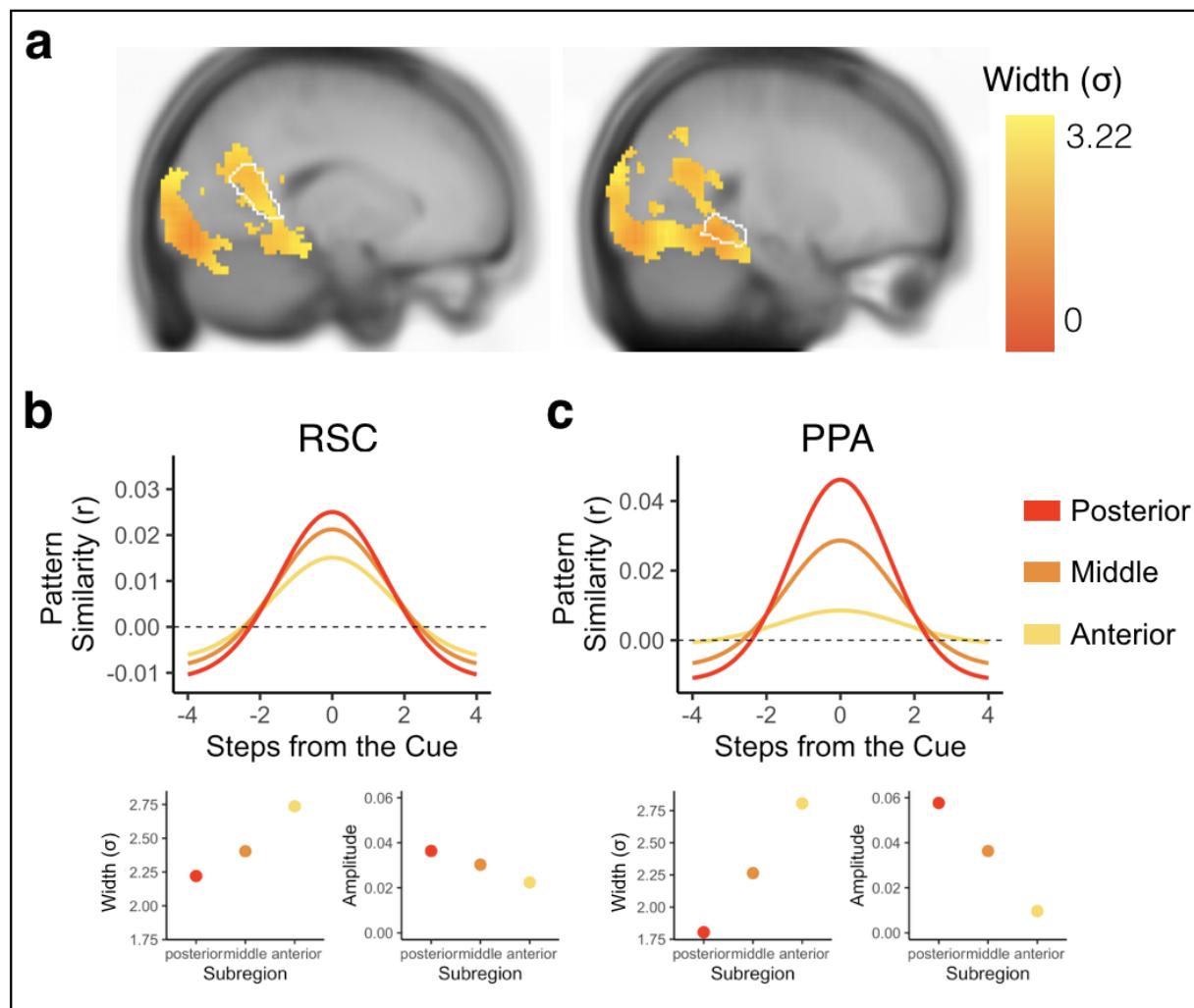


Figure 2.6: Bidirectional and graded representations of temporal structure across visual regions reveals within-region hierarchies. (a) Searchlight results revealed statistically reliable Gaussian representations in voxels across visual regions. Forward and backward widths (σ) of the Gaussian curves were hierarchically organized within visual regions (e.g., RSC and PPA), with narrow widths (indicated in red) in more posterior aspects of the region and progressively wider

widths (indicated in yellow) in progressively more anterior aspects of the region. Gaussian fits of sample voxels are shown from RSC (**b**) and PPA (**c**). Voxels in progressively more anterior (indicated in yellow) compared to posterior (indicated in red) aspects of RSC and PPA had progressively wider widths (bottom left in **(b)** and **(c)**) and progressively lower amplitudes (bottom right in **(b)** and **(c)**).

Context-dependent representations of temporal structure

Having established how the brain represents temporal structure within a given sequence, we next sought to establish whether sequence representations are context dependent. We addressed this question by leveraging pattern similarity differences between the Green and the Blue path, which contained the same environments in a different order. We conducted pattern similarity analyses by obtaining the correlation between the multivariate activity pattern evoked during a specific trial type (cue and path) for a given participant and the averaged multivariate activity patterns from the Localizer Task from remaining participants for environments (1) one and two steps away, (2) in the forward and backward direction, and (3) on the cued and the uncued path (Figure 2.7a). We focused on one and two steps because our sequences were specifically designed so that environments one and two steps away were not shared between the two paths. We then obtained the difference in pattern similarity values on the cued and uncued path for each step and direction. We then tested whether the cued vs uncued difference in pattern similarity was influenced by step (one vs. two), direction (forward vs. backward), and their interaction in each of our three ROIs (visual cortex, hippocampus, and insula). Finally, we assessed whether the cued vs uncued difference in pattern similarity was different from zero, separately by trial type and ROI.

In the insula, there was a main effect of direction ($\beta = 0.033$, 95% CI = [0.011 0.054], $p = 0.003$; Figure 2.7b), with higher cued vs uncued pattern similarity in the forward compared

to the backward direction. There was no main effect of step ($\beta = 0.016$, 95% CI = [-0.005 0.038], $p = 0.129$), nor a step by direction interaction ($\beta = -0.010$, 95% CI = [-0.052 0.032], $p = 0.644$). Given the main effect of direction, we conducted follow-up t-tests to assess whether context-dependent representations were significantly different from zero in the forward or backward directions. We found context-dependent prioritization of upcoming environments ($t(31) = 2.443$, $p = 0.017$), but not past environments ($t(31) = -1.804$, $p = 0.076$). This is consistent with a forward bias in the insula, with context-dependent prioritization of upcoming but not past environments.

In contrast, neither the hippocampus nor visual cortex showed modulation of context-dependent pattern similarity values by direction (hippocampus: $\beta = 0.004$, 95% CI = [-0.011, 0.019], $p = 0.589$; visual cortex: $\beta = 0.009$, 95% CI = [-0.002 0.021], $p = 0.111$; Figure 2.7b), step (hippocampus: $\beta = -0.011$, 95% CI = [-0.024, 0.001], $p = 0.095$; visual cortex: $\beta = 1.235e-03$, 95% CI = [-0.010, 0.013], $p = 0.835$), or their interaction (hippocampus: $\beta = -0.010$, 95% CI = [-0.031, 0.011], $p = 0.342$; visual cortex: $\beta = 0.003$, 95% CI = [-0.019 0.027], $p = 0.745$). Furthermore, the differences in pattern similarity values for the cued vs. uncued paths were not different from zero for either step in either direction for either region (all $p > 0.08$), indicating that these regions did not show context-dependent representations during the cue screen.

Although we failed to find evidence for context-dependent representations in visual cortex during the cue screen of the Anticipation Task, this may have been because a visual stimulus was presented during the cue screen and visual cortex activity is strongly modulated by visual input. Thus, we tested whether visual cortex representations during the Blank Screen period of the Anticipation Task (Figure 2.2a) showed evidence for anticipation (Kok et al.,

2014). We focused on context-dependent anticipation of the immediately upcoming environment because of past work showing one-step anticipatory signals in visual cortex (Gavornik & Bear, 2014; Hindy et al., 2016). We obtained the correlation between (1) multivariate patterns from the blank screen of the Anticipation task for each trial type (cue and path) for a single participant and (2) the averaged multivariate patterns for the remaining participants for the environment templates coming up one step in the forward direction on the cued and uncued paths (Figure 2.7a). We then obtained the difference in the cued and uncued pattern similarity values. As hypothesized, we found that visual cortex represented the upcoming environment one step into the future on the cued path more than the uncued path ($t(31) = 2.079$, $p=0.046$; Figure 2.7c). Thus, visual cortex represented one step context-dependent predictions, but only in the absence of visual stimulation.

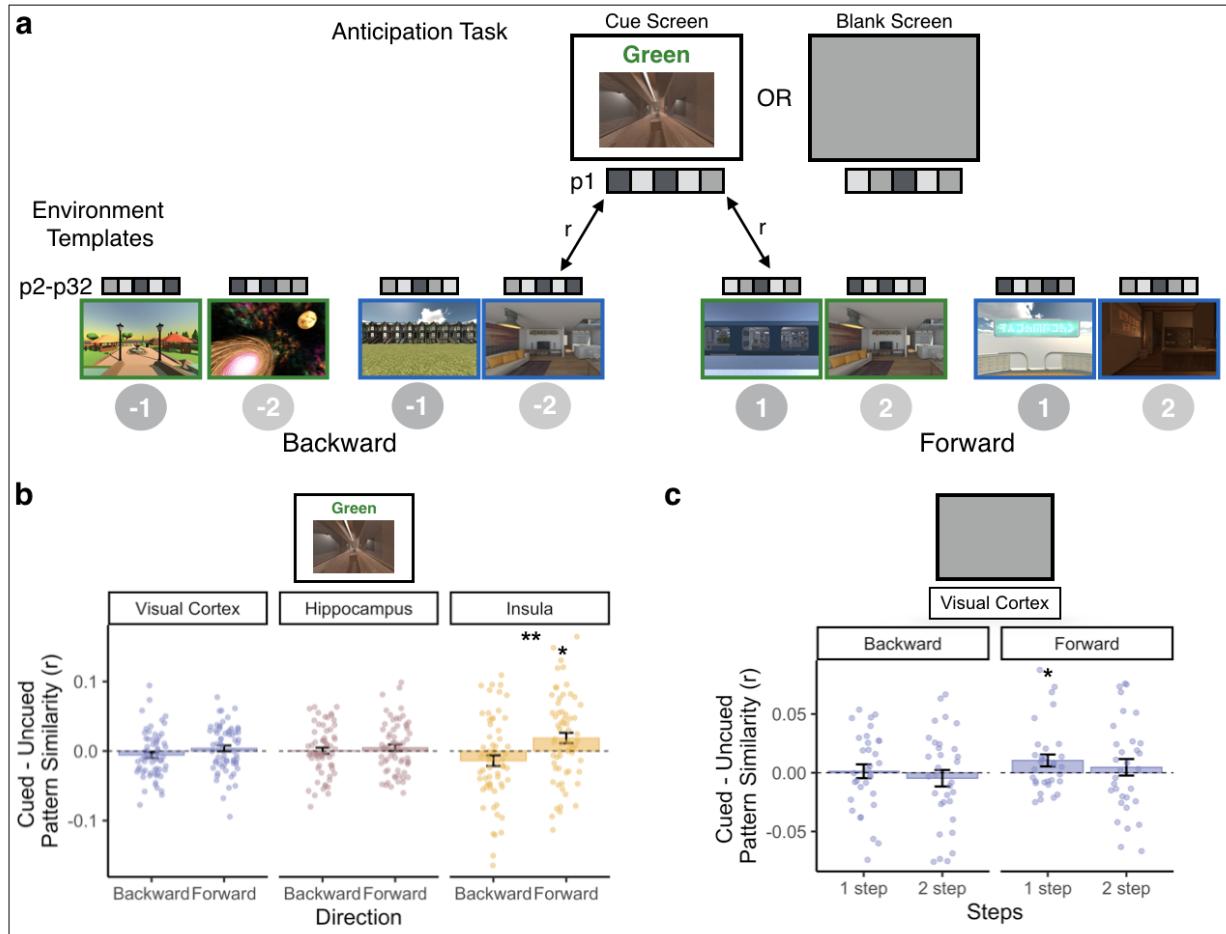


Figure 2.7: Context-dependent representations of multistep anticipation across the cortex.

(a) Schematic depiction of context-dependence analysis. We obtained the correlation between a given participant's (e.g., P1) cue screen (or blank screen) activity patterns from the Anticipation Task and the remaining participants' (e.g., P2-P32) averaged patterns of activity for environment templates (1) one and two steps away, (2) in the forward and backward direction, and (3) on the cued and uncued path. We then subtracted the correlations on the cued and uncued path (e.g., one step in the forward direction on the cued path minus one step in the forward direction on the uncued path) to obtain a measure of context-dependent pattern similarity. Positive values indicate prioritized representations for environments on the cued path, whereas negative values indicate suppressed representations for environments on the cued path. **(b)** Cue screen analysis. During the cue screen, the insula represented environments one and two steps in the forward direction in a context-dependent manner, with stronger representations for upcoming environments on the cued vs. the uncued path. The insula also suppressed environments one and two steps in the backward direction on the cued vs. the uncued path. Yellow bars indicate group-average pattern similarity values and yellow points indicate individual participants' pattern similarity values. Error bars indicate standard error of the mean. The hippocampus (indicated in pink) and visual cortex (in purple) did not represent nearby environments in either direction in a context dependent manner. Graphs are collapsed across steps for visualization purposes. **(c)** Blanks screen analysis. During the blank screen, visual cortex more strongly represented one step (but not two steps) in the forward direction on the cued vs. the uncued path. Visual cortex did not show context-dependent representations in the backward direction for either one or two steps. Purple bars indicate group-

average pattern similarity values and blue points indicate individual participants' pattern similarity values. Error bars indicate standard error of the mean.

2.3 Discussion

We examined how extended temporal structure is represented in the brain during context-dependent anticipation of future events. Participants anticipated multiple steps into the future accurately, but were slower to anticipate far vs near events. Multivariate fMRI analyses revealed bidirectional representations of temporal structure within a context, with graded representations of environments in the forward and backward direction. Visual cortex (V1-4) primarily represented the current environment, whereas hippocampus represented further environments in a graded fashion. Hippocampal representations of temporal structure were relevant for behavior: suppression of environment representations was linked to response time costs for anticipating further events. Beyond visual cortex and hippocampus, a hierarchy of temporal structure was also apparent within individual visual regions: successively more anterior aspects of PPA and RSC represented further environments into the past and future. Finally, visual cortex and insula exhibited context-dependent representations in the forward but not backward direction, suggesting prioritization of context-relevant information about future states.

Our results are generally consistent with influential theories of prediction in the brain. Graded coding of upcoming events is consistent with successor representation models (Dayan, 1993; Momennejad et al., 2017; Momennejad & Howard, 2018; Stachenfeld et al., 2017), which suggest that information about future states becomes cached into the representation of the current state in a temporally discounted manner. These models have been extended to account for multiple timescales of prediction by incorporating different scales of temporal discounting (Momennejad & Howard, 2018). In line with these theories, recent work has shown that multiple timescales of prediction are represented simultaneously in the brain (Lee et al., 2021), with less

evidence for further predictions (Brunec & Momennejad, 2022; Ekman et al., 2023). Strikingly, although our asymmetric Gaussian analysis was designed to allow differential coding of the future vs the past, representations were not uniquely biased toward future states. Instead, the hippocampus and visual system represented temporal structure bidirectionally, with graded representations into the past and future. Taken together with prior work showing that hippocampal representations of temporal distance (Deuker et al., 2016; Ekman et al., 2023; Fernandez et al., 2023) can be flexibly biased in either the forward or backward direction based on task demands (Wimmer et al., 2020), our findings suggest that representations of the past and future can exist simultaneously within the hippocampus, even though the task demands were to anticipate future states. Thus, our work extends theories of prediction across the brain, suggesting that graded retrospection of past states can occur alongside prediction of future ones.

An important distinction between our experiment and past studies of prediction is that our sequences were circular and temporally extended, whereas sequences in prior studies tended to have a clear end point (i.e. were linear instead of circular) (Brunec & Momennejad, 2022; Crivelli-Decker et al., 2023; Elliott Wimmer & Büchel, 2019; Lee et al., 2021) or were shorter (Brown et al., 2016). Our novel design therefore allowed us to detect temporally extended representations in both the forward and backward directions. Thus, it is possible that prior work emphasizing anticipatory coding in brain regions missed simultaneous representations of the past because the tasks were not designed to detect bidirectional temporal coding.

In addition to representing nearby environments in the past and future, we also found that the hippocampus suppressed more distant environments, showing deactivation of these environments' patterns relative to an unrelated-environment baseline. Although suppressing distant environments can be beneficial for responding to imminent events, it can also lead to

behavioral costs. Indeed, hippocampal suppression of distant environments was related to response times costs for anticipating further events. This highlights a trade-off between prioritizing nearby events and being able to quickly respond to upcoming events further in the future.

Representations of temporal structure extended beyond hippocampus and visual cortex. In an exploratory whole-brain searchlight analysis, we found representations of temporal structure across the visual system, including PPA and RSC, regions that play an important role in spatial cognition (Epstein, 2008). Both PPA and RSC represented the cued environment but also represented the temporal structure of surrounding environments in the sequence in both the forward and backward direction. Our findings therefore extend prior work showing that PPA responses can be modulated by temporal context (Turk-Browne et al., 2012) and prior contextual associations more generally (Aminoff et al., 2007; Aminoff et al., 2013; Bar et al., 2008; Marchette et al., 2015). Notably, our findings go beyond this prior work by showing a gradual progression of sequence coding within PPA and RSC, with progressively more anterior regions representing more of the future and past and less of the present. This is broadly consistent with prior work suggesting a posterior vs. anterior division within PPA, with posterior aspects playing a larger role in scene perception and anterior aspects playing a larger role in scene memory (Baldassano et al., 2013; Steel et al., 2021). Thus, we show that, within a context, visual regions may balance representations of perception and memory, gradually incorporating less information from perception and more information about learned temporal structure along a posterior to anterior hierarchy.

To investigate bidirectional context-dependent sequence representations, we carefully manipulated overlapping sequences. Pairs of sequences contained the same environments in a

different order, with no overlap in environments one and two steps into the future and the past.

Prior work has found context-specific prediction of outcomes in visual cortex (Clarke et al., 2022; Hindy et al., 2016), which dovetails with our finding that visual cortex represented one-step into the future more strongly on the cued compared to the uncued sequence in the absence of visual stimulation. We also found that the insula represented two steps into the future in a context-dependent manner, furthering recent work from our lab showing long-timescale predictions in this region (Lee et al., 2021). These findings are also consistent with theories that anterior brain regions should predict further into the future than posterior ones (Brunec & Momennejad, 2022; Momennejad & Howard, 2018). Interestingly, we only observed context-dependent prioritization into the future, but not the past. This suggests that prediction of future states may be context-specific, whereas retrospection may be context-independent. Another possibility is that the context-specific prioritization of future or past states may depend on task demands: the demand to predict future states may have elicited context-dependent prediction but not retrospection. Future work could further examine the circumstances under which context-dependent prioritization of future and past states emerge.

Unlike visual cortex and insula, we were unable to detect context-specific representations in the hippocampus. This result is seemingly in contrast to past work showing context-dependent prediction in the hippocampus multiple steps into the future (Crivelli-Decker et al., 2023), and a large literature showing contextual representations across the hippocampus more generally (for reviews see Davachi, 2006; Eichenbaum et al., 2007). One possibility for this discrepancy is that some hippocampal subregions emphasize integration across contexts whereas others emphasize differentiation (Brunec et al., 2020; Dimsdale-Zucker et al., 2018; K. D. Duncan & Schlichting, 2018; Molitor et al., 2021; Schlichting et al., 2015). We focused our analysis on voxels that show

discriminable environment-specific activity patterns when no context cue is presented. These voxels may have been those that emphasize contextual integration in the hippocampus, and may therefore automatically activate associated events whether or not they are relevant in the current context. Future work can separately examine anterior vs posterior hippocampus, or the CA3 vs dentate gyrus subfields, to determine whether context-dependent anticipation in the hippocampus is more likely in the former vs latter regions.

Broadly, it may be advantageous to represent temporal structure bidirectionally, rather than only prioritizing future states. For example, representing past states and future states could be a useful strategy when events surrounding ongoing experience differ based on context. Activating links toward past states as well as future ones may allow individuals to contextualize their current location within the sequence (Hasson et al., 2015). This possibility is consistent with our prior work showing that individuals represent sequences in terms of context-specific links between environments (Tarder-Stoll et al., 2023); when an environment is cued, its associated links in both directions may be brought to mind so that the entire context is prioritized. An alternative possibility is that representing temporal structure into the past and future happens automatically: activating a particular moment within a temporally extended experience could cause activation to spread to the entire event representation (Clewett & Davachi, 2017; DuBrow & Davachi, 2014; Manning et al., 2011), which may comprise both past and future. Future work could disentangle these possibilities and further investigate the circumstances under which future and past states are simultaneously represented.

Overall, the results presented here show that temporal structure is represented bidirectionally in the hippocampus and visual system. Future and past representations of temporal structure were graded, with less evidence for further environments in the forward and

backward direction, and were organized along a posterior to anterior hierarchy within and across regions. Our results further our understanding of how temporal structure is represented in the brain: such bidirectional representations could allow integration of past events from memory alongside anticipation of future ones, which could support adaptive behavior during complex, temporally extended experiences.

2.4 Methods

Participants

Thirty-five healthy younger adults from the Columbia University community participated in the experiment. All participants gave written, informed consent in accordance with the Institutional Review Board at Columbia University. Participants were compensated \$15 per hour for the behavioral training session and \$20 per hour for the fMRI session (approximately \$80 combined across both sessions). Three participants were excluded for technical issues with data collection, excessive motion (10% of TRs across all runs of the experiment marked as motion outliers by fMRIprep output), and dizziness inside the fMRI scanner. Applying these exclusions resulted in a final sample of 32 participants (21 female, 19-35 years old, mean = 24.17, sd = 4.11, 13-29 years of education, mean = 16.85, sd = 3.76).

Overview

Participants learned two sequences (“Green Path” and “Blue Path”) within each of two maps (Map A and Map B; Figure 2.1a). Map A and Map B contained eight distinct environments each. Within each map, the Green Path and Blue Path contained the same environments in a different order.

Participants first learned the order of the four sequences of environments by generating stories (Figure 2.1b) and then experiencing the environment sequences in immersive virtual reality using an Oculus Rift (Figure 2.1c). Participants then returned 1 day later and completed the Anticipation Task in the MRI scanner (two runs, 32 trials per run). In the Anticipation Task, participants used their memory for the four sequences to anticipate upcoming environments. They then completed a localizer task to obtain multivariate patterns of brain activity for each environment (four runs, 16 trials per run).

Stimuli and Sequence Structure

Stimuli consisted of 16 3D virtual reality environments in the Unity game engine. Environments were obtained from asset collections in the Unity Asset Store. Half of the environments were indoor and half of the environments were outdoor. Using Unity, we created 2D images of each environment by rotating a virtual camera to eight different angles, 45 degrees apart. One angle was selected to be used as the cue and probe images throughout the task and the other angles were used for the panorama phase of the Localizer Task.

The 16 environments were used to form four sequences (Map A Green Path, Map A Blue Path, Map B Green Path, Map B Blue Path). Eight of the environments were assigned to Map A (i.e., the first learned set of environments) and the other eight environments were assigned to Map B (i.e., the second learned set of environments). Then, within each map (A or B), the Green Path and the Blue Path consisted of the same eight environments in a different order. The final environment in each sequence connected back to the first environment, forming a circle. The Green and Blue Paths were designed to be as distinct as possible: for a given environment the two preceding and two succeeding environments were different across the paths (Figure 2.1a). The environment-to-map assignment and the order of the environments within a sequence was

randomized across participants, although the Green Path was always shuffled in the same way to create the Blue Path, as described above.

Procedure

Participants first completed a training phase outside the MRI scanner. They then returned one day later and completed a sequence refresher task outside the MRI scanner before taking part in the fMRI session. During fMRI, they completed an Anticipation Task (two runs), an Integration Task (four runs, data not analyzed or reported in the current manuscript) and a Localizer Task (four runs). In the training session, stimuli were presented on a computer screen with PsychoPy (Peirce et al., 2019) and in virtual reality with an Oculus Rift and Unity, using a mixture of custom code and OpenMaze (Alsbury-Nealy et al., 2022). In the fMRI session, PsychoPy was used to present the stimuli, which were projected onto a screen in the scanner bore and viewed via a mirror mounted on the head coil.

Training Phase

In the training phase (one day before the fMRI scan), participants were instructed to learn the order of the four sequences (Map A Green, Map A Blue, Map B Green, Map B Blue; see *Stimuli and Sequence Structure*). Participants always began by learning the Map A Green Path, because Map A was defined as the first set of environments that participants learned and the Green Path was defined as the first sequence within each map.

Participants were instructed to learn the sequences by generating a story to link the environments in order. They first saw 2D renderings of all the environments in the Map A Green Path order displayed on a computer screen. They were told to generate a detailed story to link the environments in order, and that the final environment should loop back to the first environment in the sequence to create a circle. Participants indicated that they were finished generating a story

by pressing a button. Then, they were shown the sequence as pairs of adjacent environments with an empty text entry box displayed underneath (e.g., environments #1 and #2, then environments #2 and #3, etc). Participants were told to write down the story that they had generated (Figure 2.1b; see Appendix B for story examples). Participants were given unlimited time to generate and write down their story. Once they had finished, participants verbally repeated the story back to the experimenter.

Following story generation, participants then experienced the Map A Green Path in virtual reality using an Oculus Rift (Figure 2.1c). Participants were initially placed in the first environment in the sequence. After five seconds, a floating green sphere and blue sphere appeared in a random location within reaching distance of the participant. Participants were told that touching the spheres would teleport them to the next environment in the correspondingly colored sequence: they were told to touch the green sphere on the Green Path and the blue sphere on the Blue Path. After being teleported to the next environment in the corresponding sequence, participants were again given five seconds to explore the environment before the spheres would appear. After 20% of trials (“test trials”), instead of teleporting to the next environment in the sequence, participants were teleported to a black environment in which they were shown two images of upcoming environments and were told to indicate which of those two environments was coming up sooner in the sequence they were currently “traversing”, relative to the preceding environment. Participants had ten seconds to respond using the Y and B buttons on the left and right Oculus Rift controllers. They were given feedback about whether their answer was correct or incorrect. As participants were exploring the environments in virtual reality, they were also told to rehearse their stories to ensure the sequence was learned. Participants rehearsed the Map A Green Path sequence in virtual reality three times following this procedure.

Participants then repeated the exact same procedure, but learned the Map A Blue Path, which consisted of the same environments as the Map A Green Path in a different order. Participants were told to make their Blue Path story distinct from their Green Path story to avoid confusing the two paths. They then followed the same virtual reality procedure as noted above, but were instructed to touch the blue spheres instead of the green spheres to teleport between environments.

Following Map A Green and Blue Path learning, participants were exposed to each sequence three more times (including test trials) in virtual reality in an interleaved fashion (i.e., one repetition of Green Path, one repetition of Blue path). Participants then recalled the order of the Map A Green and Blue Paths. The above procedure was then repeated for the Map B Green and Blue Paths. In total, the training phase took between one and a half and two hours to complete. All participants performed at ceiling by the end of the training phase.

Sequence Refresher Task

Participants then returned one day later. Before the fMRI scan, they completed a sequence refresher task to ensure they maintained memory for all four sequences learned during the Training Phase. Participants viewed 2D renderings of all the environments from virtual reality, one at a time, in the order of each of the four sequences (Map A Green, Map A Blue, Map B Green, Map B Blue). Participants saw each sequence in order three times. In the first presentation, participants were told to verbally repeat the stories they had generated for each sequence. In the subsequent two presentations, participants were told to verbally recall the environment that came after the currently presented environment in the current sequence.

Anticipation Task

During the fMRI scan, participants first completed the Anticipation Task, for which there were two runs with 32 trials each (Figure 2.2a). On each trial in the Anticipation task, participants were cued with an environment and a path cue (“Green” or “Blue”) for 3 seconds. This cue indicated the starting point and sequence on that trial. Participants then viewed a blank (gray) screen for a variable duration (five to nine seconds). Then, participants were presented with two images of upcoming environments and were told to judge which of the two environments was coming up sooner in the cued sequence, relative to the cue image. Participants were given three seconds to make this judgment. This relatively short response deadline was implemented to encourage participants to use the blank screen period to generate predictions along the cued path in preparation for the forced choice decision. The correct answer could be one to four steps away from the cue image. The incorrect answer could be a maximum of five steps away from the cue image. Because the sequences were circular, every environment could be used as a cue with successors up to five steps away. There was a uniformly sampled three to eight second jittered inter-trial interval (ITI), during which participants viewed a fixation cross. At the end of each run, there was a 60 second rest period during which participants viewed a blank screen.

In each run, participants were cued with every environment from Map A and B on the Blue and Green Paths (eight environments per sequence) for a total of 32 trials per run (64 trials total). In the probe phase, the correct answer was equally distributed across steps into the future (one to four). The incorrect answer was randomly sampled to be one to four steps away from the correct answer (two to five steps away from the cue). Within a run, sequences were presented in blocks (i.e., participants completed the Anticipation Task for all the environments in the Map A

Green Path in one block), but the order of the cues was randomized within a block. The order of the sequence blocks was also randomized across runs and participants. A single run of the Anticipation Task was approximately 11 minutes, for a total of 22 minutes across both runs.

Integration Task

Following the Anticipation Task, participants completed an Integration Task, in which they were told that one of the environments in Map A was now connected to one of the environments in Map B on either the Green or Blue Path. One of the environments in Map B also connected back to Map A, creating a single integrated path encompassing all the environments in both maps. The integrated path (Green or Blue) was counterbalanced across participants, with the other path serving as a control non-integrated path. For example, if a participant learned that Map A and Map B were integrated on the Green Path, the Blue Path would be the non-integrated path. The environments that connected Map A to Map B were randomly selected, while the environments that connected Map B back to Map A were always the preceding environments in the sequence, allowing the integrated path to form a circle. Participants then completed a version of the Anticipation Task (see above) in which they anticipated upcoming environments in the non-integrated and integrated paths (four runs, 24 trials per run). The Integration Task is not analyzed or reported in the current manuscript.

Localizer Task

Participants then completed four runs of a Localizer Task used to obtain environment-specific patterns of brain activity across participants (Figure 2.3a, see *Environment Templates*). In the Localizer Task, participants were cued with an environment from Map A or B on the screen for one second. Importantly, the cue in the Localizer Task did not include a path cue (Green or Blue), allowing us to obtain a context-independent pattern of brain activity for each

environment. Following the cue, participants saw a blank gray screen for five seconds, during which they were told to imagine being inside the environment in virtual reality. Participants then viewed images of the cued environment from different angles, 45 degrees apart, for four seconds. They were then given three seconds to rate how well their imagination matched the actual images of the environment, on a scale from one to four (one = not well, four = very well). There was a three to eight second jittered ITI, during which participants viewed a fixation cross.

In each run, participants were cued with every environment from Map A and B for a total of 16 trials per run (64 trials total across all four runs). The order of the environments was randomized across runs and participants. A single run of the Localizer Task was approximately five and a half minutes, for a total of 22 minutes across all four runs.

Behavioral Analysis

We conducted analyses on the behavioral data in the R programming language using generalized linear and linear mixed effects models (GLMMs and LMMs, *glmer* and *lmer* function in the *lme4* package; Bates et al., 2014). For analyses that modeled multiple observations per participant, such as accuracy or response time on a given trial, models included random intercepts and slopes for all within-participant effects. All response time models examined responses on correct trials only.

To ensure that participants performed effectively during the Anticipation Task, we first tested that accuracy during the Probe screen (see Figure 2.2a) was better than chance performance (50%) using a one-sample t-test.

We next determined whether accuracy and response time differed across the Maps (A and B) and Paths (Green and Blue). To examine sequence effects, we fit separate models for accuracy (a GLMM) and response time (an LMM) as a function of Map (A = -0.5, B = 0.5), Path

(Green = -0.5, Blue = 0.5), and their interaction. We used the following R-based formulas (where “participant” indicates participant number):

*glmer(correct ~ map*path + (1+map*path|participant), family = “binomial”, data)*

*lmer(RT ~ map*path + (1+map*path|participant), data, subset = (correct == 1))*

Next, we determined whether accuracy and response time differed across steps into the future. We fit separate models for accuracy (a GLMM) and response time (an LMM) as a function of steps into the future (-0.75 = 1 step, -0.25 = 2 steps, 0.25 = 3 steps, 0.75 = 4 steps). We used the following R-based formulas (where “participant” indicates participant number):

glmer(correct ~ steps + (1+steps|participant), family = “binomial”, data)

lmer(RT ~ steps + (1+steps|participant), data, subset = (correct == 1))

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. To improve signal to noise ratio in mPFC, all slices were tilted approximately -20 degrees relative to the AC-PC line. There were ten functional runs in total: two runs of the Anticipation Task, four runs of an Integration Task (not analyzed or reported here), and four runs of the

Localizer Task. Field maps were collected after the final functional scan to aid registration (TR = 679 ms, TE = 4.92 ms/7.38 ms, flip angle = 60°, 69 slices, 2 mm isotropic).

Preprocessing

Results included in this manuscript come from preprocessing performed using *fMRIprep* 1.5.2 (Esteban et al., 2019, 2022); RRID:SCR_016216), which is based on *Nipype* 1.3.1 ((Esteban et al., 2022; Gorgolewski et al., 2011); RRID:SCR_002502).

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.2.0 (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 5.0.9, RRID:SCR_002823, Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, 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, A. Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al., 2009], RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

For each of the 10 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*. A deformation field to correct for susceptibility distortions was estimated based on a field map that was co-registered to the BOLD reference, using a custom workflow of *fMRIPrep* derived from D. Greve's epidewarp.fsl script and further improvements of HCP Pipelines (Glasser et al., 2013). Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. 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. 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 5.0.9, Jenkinson et al., 2002). The BOLD time-series, were resampled to surfaces on the following spaces: *fsaverage6*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. 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*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. 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 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). 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 standardised DVARS were annotated as motion outliers. 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* 0.5.2 (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.

Copyright Waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the CC0 license.

fMRI Analysis

After preprocessing, all fMRI analyses were performed in Python and R. Pattern similarity analyses were performed using custom code in Python 3. Statistical analysis comparing pattern similarity values across conditions, correlations between fMRI results and behavior, and visualizations were performed using custom code in R.

Localizer Task Analyses

We conducted GLMs predicting whole-brain univariate BOLD activity from task and nuisance regressors from the Localizer Task using custom scripts in Python. For each participant, we first concatenated the fMRI data across runs of the Localizer Task and modeled BOLD activity for each environment (1 to 16) with a boxcar regressor combined across the cue, blank screen, and panorama periods. We also included nuisance regressors in the same model (translation and rotation along the X, Y, and Z axes and their derivatives, motion outliers as

determined by fMRIprep, CSF, white matter, framewise displacement, and discrete cosine-basis regressors for periods up to 125 seconds).

We next looked across the whole brain for voxels that showed reliable, environment-specific patterns of activity during the Localizer Task. We used an approach that identifies voxels that respond reliably to different conditions across runs of an experiment (Tarhan & Konkle, 2020), here measuring reliability across different participants (Hasson et al., 2004). For each voxel, we obtained a 16-element vector of beta weights from the whole-brain GLM, reflecting the beta weight for each of the 16 environments for each participant (e.g., Participant #1 or P1). Next, we obtained the Pearson correlation (r) between each participant's 16-element vector in each voxel and the averaged 16-element vector from the remaining participants (e.g., P2-P32). Finally, we calculated an environment reliability score by averaging the r values across all iterations of the held-out participant (Figure 2.3b). Voxels that had an r value of 0.1 or greater ("environment-reliable voxels") were then included in subsequent steps (Figure 2.3c). We selected 0.1 as our cutoff because it resulted in reasonable spatial coverage while maintaining voxel reliability, including in hippocampus and prefrontal regions (Tarhan & Konkle, 2020).

Conjunction ROI Definition

Three *a priori* regions of interest (ROIs) were defined using environment-reliable voxels (see above) within anatomical or functional areas of interest. The V1-4 ROI was obtained from the probabilistic human visual cortex atlas provided in Wang et al. (2015) (threshold: $p = 0.50$). The hippocampus and insular cortex ROIs were both defined from the Harvard-Oxford probabilistic atlas in FSL (threshold: $p = 0.50$). We resampled the three ROIs onto the same MNI grid as the functional data (MNI152NLin2009cAsym), and then intersected them with our map

of environment-reliable voxels ($r > 0.1$, see description above) to create conjunction ROIs in visual cortex, hippocampus, and insula (Figure 2.3d).

We then obtained the spatial pattern of activity across voxels in each conjunction ROI; these served as environment-specific template activity patterns for each participant. Because the Localizer Task did not include a path cue (Green or Blue), participants should not have been differentially and consistently activating one path as they viewed each environment; thus, the pattern of activity obtained for each environment should be context-independent. Importantly, this approach yielded the expected result of producing ROIs with environment-specific patterns of activity: activity patterns for the same environment were more correlated than activity patterns for different environments within each conjunction ROI (Figure 2.3d). These environment-specific patterns (hereafter referred to as “environment templates”) are a necessary precursor for investigating prediction along each sequence (see below).

Anticipation Task Analyses

We conducted GLMs predicting whole-brain univariate BOLD activity from behavioral and nuisance regressors from the Anticipation Task using Python. For each participant, we modeled BOLD activity concatenated across both runs of the Anticipation Task with separate regressors for the cue, blank screen, and probe periods for each environment in Map A and B (1 to 16) and for each path (Green Path and Blue Path). This resulted in a total of 32 task regressors for each phase (cue, blank screen, probe) of the Anticipation Task (16 environments across Map A and Map B, with each environment modeled separately for the Green Path and the Blue Path). We also included nuisance regressors in the same model (the same as those used for the Localizer Task Analyses). For all subsequent analyses (except the searchlight analysis), the resulting beta weights were examined within our conjunction ROIs.

Cue Period - Asymmetrical Gaussian Analysis

To assess evidence for multivariate representations of temporal structure, we obtained the correlation between (1) a given participant's (e.g., P1) cue screen activity pattern for each trial type (a given environment cued on a given path) in the Anticipation Task and (2) the remaining participants' (e.g., P2-P32) averaged patterns of activity for each of the environment templates from the corresponding map (Figure 2.4a). For example, if Participant 1 was cued with environment one from Map A on the Green Path, we obtained the correlation between: (1) Participant 1's cue screen activity pattern for that environment and path and (2) each environment template (averaged across Participants 2-32) from Map A. This yielded eight separate pattern similarity values (because there are eight environments per map) for each trial type (a given environment cued on a given path). Because each participant had a different ordering of environments, these across-participant templates cannot contain any reliable information about the successors or predecessors of a given environment for a given participant. We then ordered the resulting pattern similarity values according to the cued Map and Path (in this example, the Map A Green Path order), with the cue in the center (Figure 2.4a). Thus, successors following the cue would be to the right of the center and predecessors would be to the left of the center. Because, in an eight-environment map, four steps away is an equal distance from the cue in both the past and the future, we included the pattern similarity value four steps away from the cue in both the forward and the backward direction. We also obtained the correlation between (1) a given participant's (e.g., P1) cue screen activity pattern for each trial type (a given environment cued on a given path) in the Anticipation Task and (2) the remaining participants' (e.g., P2-P32) averaged patterns of activity across all environment templates in the different map (in this example, Map B). This single value served as the different map baseline.

Next, we fit an asymmetrical Gaussian curve to the resulting ordered pattern similarity values. We chose to use a Gaussian curve because we hypothesized that brain regions would represent upcoming (or past) environments in a graded manner, with stronger representations for nearby environments (Momennejad & Howard, 2018). The asymmetrical Gaussian has four parameters: amplitude, asymptote, and forward and backward widths (σ) (Figure 2.4b). The amplitude controls the height of the peak of the Gaussian curve, and indicates the extent to which a brain region is representing the cue environment presented on the screen. The asymptote controls the vertical shift of the Gaussian curve; a negative asymptote reflects suppression of some environments (i.e. the cue pattern is anticorrelated with some environment templates). The widths (σ) control the slope of the fall off from the amplitude to the asymptote. Wider widths (σ) indicate activation of environment patterns further away from the cue. Because we fit an asymmetrical Gaussian, we obtained different widths (σ) in the forward and backward direction; this allows brain regions to potentially represent more environments in one direction (e.g., upcoming environments) than another (e.g., past environments). This in turn enables us to detect if some brain areas anticipate the future but do not represent the past. We constrained the widths (σ) to be a maximum of 10 and applied L2 regularization to the amplitude and intercept (with strength = 0.01) to ensure the model did not return uninterpretable parameter values.

To test whether the parameters of the Gaussian curve were consistent across participants, we fit the asymmetrical Gaussian curve on all but one participant's data (e.g. P2-32) and then measured the sum of squared errors (observed vs. predicted pattern similarity values) when using this curve to predict the held-out participant's data (e.g. P1). We repeated this procedure for each choice of held-out participant to obtain an average error value. We compared this error to a null distribution, fitting the asymmetrical Gaussian curve to data in which the order of the

environments was shuffled. We defined a p value as the fraction of 10,000 shuffles which produced lower error than our original unshuffled fit.

We also performed statistical tests on the parameters of the correctly ordered Gaussians. We tested whether the amplitude was significantly above the different map baseline, and whether the asymptote was suppressed below the different map baseline, using one-sample t-tests across participants. We also tested whether the widths (σ) differed by brain region (visual cortex vs hippocampus), direction (forward vs backward), and their interaction using the following R-based formula (where “participant” indicates participant number):

$$lmer(\text{width} \sim \text{region} * \text{direction} + (1|\text{participant}), \text{data})$$

Cue Period - Searchlight

We conducted a whole-brain searchlight analysis with custom Python code to test whether brain regions beyond our ROIs represented temporal structure in the hypothesized asymmetrical Gaussian format. We looked for significant Gaussian representations in cubes with a side length of 7 voxels, moved throughout the whole brain volume with a step size of 2 voxels. We included only environment-reliable voxels within each cube, and only proceeded with the analysis of a cube if it contained at least 64 valid voxels. The parameters of fitted Gaussians within each searchlight, along with the goodness of fit, were assigned to each voxel in the searchlight. For voxels that were included in more than one searchlight, the final Gaussian parameters and goodness-of-fit were obtained by averaging the results across all the searchlights in which the voxel was included.

To determine which voxels exhibited significant Gaussian representations across participants, we first obtained a measure of goodness of fit by dividing the squared errors of the correctly ordered Gaussian by the average of the squared errors of the permuted Gaussians and

then subtracting the resulting value from 1 for each voxel included in the searchlight for each participant. Numbers above 0 indicate better fits to the correctly ordered vs permuted data. We then statistically tested whether the goodness of fit values were greater than 0 in each voxel, using the randomise function with threshold free cluster enhancement in FSL, which generates null distributions using 10,000 permutations and performs a one-sample t-test while enhancing clusters of significant voxels (Smith & Nichols, 2009). We then corrected for multiple comparisons using the family-wise error rate correction ($p<0.05$).

To determine whether widths (σ) were organized hierarchically within brain regions, we first averaged the forward and backward widths (σ) for each voxel for each participant. We opted to compute the average of the forward and backward widths (σ) because we did not find evidence for directional asymmetry in any brain regions, in both our ROI and searchlight analyses. Next, we determined whether the averaged widths (σ) became increasingly wider in more anterior, compared to posterior, voxels in the parahippocampal place area (PPA) and the retrosplenial cortex (RSC). We created PPA and RSC ROIs using pre-defined anatomical ventral visual stream ROIs (Julian et al., 2012), which we then resampled onto the same MNI grid as the functional data (MNI152NLin2009cAsym) and intersected with our map of environment-reliable voxels environment-reliable voxels ($r > 0.1$). We chose PPA and RSC because (1) the searchlight revealed significant Gaussian representations in the majority of voxels in these regions and (2) they have previously been implicated in both scene perception and memory (Baldassano et al., 2013; Epstein, 2008). In each region, for each participant, we obtained the Spearman rank-order correlation between the averaged forward and backward widths (σ) and the y coordinate (indicating a voxel's position on the posterior-anterior axis) across voxels. Finally, we determined whether the correlation was significant at the group level by comparing the

participant-specific r values to 0 using a one-sample t-test. A significantly positive r value would indicate that widths (σ) become increasingly wider in increasingly anterior aspects of regions.

Cue Period - Relationship to Behavior

We determined whether an individual's asymptote from their Gaussian model, indicating suppression of further environments from the cue, was related to response time costs for further environments. Response time costs were quantified with participant-specific regressions that predicted response time as a function of steps into the future. We then performed an individual differences analysis by obtaining the Spearman rank-order correlation between participants' response time costs and their asymptotes in (1) hippocampus and (2) visual cortex.

Cue Period - Context Dependent Analysis

To test whether multivariate patterns of activity were context-dependent, we obtained the correlation between a given participant's (e.g., P1) cue screen activity pattern for each trial type (a given environment cued on a given path) in the Anticipation Task and the remaining participants' (e.g., P2-P32) averaged patterns of activity for each of the environment templates (1) one and two steps away from the cue, (2) on the cued path and the uncued path, and (3) in the forward and backward direction (Figure 2.7a). This yielded eight pattern similarity values for each combination of step (one vs two), context (cued vs uncued), and direction (forward vs backward). For each step in each direction, we then obtained a measure of context-dependent representations by subtracting the pattern similarity values for environments on the uncued path from the pattern similarity values for environments on the cued path. For example, context-dependent representations for one step in the forward direction would be calculated as pattern similarity between the cued environment and the environment one step in the forward direction on the *cued* path minus the pattern similarity between the cued environment and the environment

one step in the forward direction on the *uncued* path. Positive values indicate prioritized representations for environments on the cued path, relative to the uncued path, whereas negative values indicate suppressed representations for environments on the cued path relative to the uncued path. We then statistically tested whether, at the group-level, there were prioritized or suppressed context-dependent representations as a function of direction (forward vs backward), step (1 step vs 2 step), and their interaction. We conducted these tests separately for each region (visual cortex, hippocampus, and insula) using the following R-based formula, where cueContextDiff indicates the cued minus uncued pattern similarity values:

*lmer(cueContextDiff ~ direction * step + (1 + direction + step | participant), data, subset = region)*

Blank Period - Context Dependent Analysis

We computed the same context-dependent analyses as described above, but using the blank screen activity pattern, rather than the cue screen, for each trial type (a given environment cued on a given path) of the Anticipation Task (Figure 2.7a). We hypothesized that we would be more likely to find one step, context-dependent representations in visual cortex during the blank screen period compared to the cue screen because (1) it is strongly modulated by visual input, potentially reducing our ability to detect anticipatory representations while images are on the screen and (2) past work has shown short timescale (e.g., one step) predictions in this region (Gavornik & Bear, 2014; Hindy et al., 2016). Based on these hypotheses, we statistically tested whether, at the group-level, visual cortex exhibited prioritized or suppressed context-dependent representations one step into the future in the forward direction using a one-sample t-test.

Chapter 3: Rapid Reconfiguration of Sequence Representations in the Hippocampus

Hannah Tarder-Soll, Christopher Baldassano*, & Mariam Aly*

*These authors contributed equally to this work

3.1 Introduction

The ability to flexibly use past experiences to guide future behavior is a cornerstone of adaptive memory. Memories for specific episodes can be reorganized and recombined to make novel inferences and decisions (Biderman et al., 2020; Morton et al., 2017). This ability to generalize across experiences enables the construction of internal models of our environment, which can be used to predict future events, even those that go beyond direct experience (Behrens et al., 2018; Epstein et al., 2017; Momennejad, 2020; O’Keefe & Nadel, 1979; Tolman, 1948). For example, we can learn the sequence of events along routes through a new city, allowing us to plan trajectories multiple steps ahead. Upon learning that two seemingly separate routes are linked (e.g., via a side street), we can plan trajectories along the integrated route, even if it has not been directly experienced (Epstein et al., 2017; Spiers & Gilbert, 2015; Tolman, 1948). From this perspective, learning about the side street allows us to reconfigure representations of the entire sequence to build an integrated and holistic understanding of the world. Here, we investigated the neural mechanisms by which new learning promotes the integration of temporally extended sequences to enable flexible anticipation of novel trajectories.

Despite the importance of flexibly reconfiguring memories for adaptive behavior, we have a limited understanding of how temporally extended experiences are reconfigured to support planning and anticipation in novel environments. Prior work from non-human primates suggests that linking a single node across two temporally extended sequences can promote their integration, enabling judgements across the sequence as whole (Treichler & Van Tilburg, 1996). However, it remains unknown how integrated sequence representations emerge in the brain and support flexible behavior. The hippocampus is a clear candidate region for representing integrated sequence structure, given its role in representing temporal structure broadly (Bellmund

et al., 2020; Brown et al., 2016; Davachi & DuBrow, 2015; Deuker et al., 2016; Schapiro et al., 2012; Thavabalasingam et al., 2019) and its ability to flexibly recombine overlapping experiences to promote inference and generalization (Eichenbaum, 2000; Eichenbaum et al., 1999; Schlichting & Preston, 2015). In humans, memories for narratives or sequences can become integrated in light of novel information or based on inferred structure, driving changes in hippocampal representations (Cohn-Sheehy et al., 2021; Fernandez et al., 2023; Milivojevic et al., 2015). However, these past studies have only examined how representational changes occur between relatively short, discrete episodes (i.e. A linked to B through X, Cohn-Sheehy et al., 2021; Milivojevic et al., 2015) or between short portions of longer integrated trajectories (Fernandez et al., 2023).

How does learning new connections change the neural representations of extended temporal sequences? Here, we test two competing hypotheses: one possibility is that the two sequences will maintain largely separate representations except for the linking event. This would be largely consistent with past work showing integration between discrete episodes (Schlichting & Preston, 2015). An alternative hypothesis is that learning a new connection will cause reconfiguration of the entire sequence, so that they come to share overlapping neural representations. In other words, reconfiguration could spread to distant events that were never themselves directly connected, suggesting the formation of an integrated internal model (Momennejad, 2020).

A secondary question was to determine how such reconfiguration occurs over time. Competing predictions suggest that such reconfiguration may either occur rapidly or require time to emerge. In support of the former, the hippocampus is well known to underlie single shot learning and rapid acquisition of simple associative memories through synaptic plasticity (Bliss & Collingridge, 1993), with rapid changes to hippocampal representations supporting inferences

across separate events (Milivojevic et al., 2015). Thus, it is possible that the hippocampus rapidly integrates representations of temporally extended sequences immediately after new learning to flexibly support planning of novel trajectories.

Muddying this prediction, however, other work suggests that the integration of internal models acquired separately in time may require time to emerge. During awake rest, the hippocampus replays trajectories that have never been directly experienced (Gupta et al., 2010; Y. Liu et al., 2019; Ólafsdóttir et al., 2015), suggesting that offline periods of rest play an important role in sampling and reconfiguring internal models used to guide adaptive behavior (Foster, 2017; Momennejad et al., 2018; Ólafsdóttir et al., 2018). Thus, another possibility is that integrated sequence representations in the hippocampus emerge, or are strengthened, with time. Importantly, these two possibilities are not mutually exclusive, in that there can be rapid initial updating followed by slower changes over time.

In the current study, we investigated how new information reconfigures representations of temporally extended sequences to support multistep anticipatory judgements. Participants learned four sequences of eight environments each (Figure 3.1a). They then learned novel transitions that linked two of the previously separate sequences into a single, integrated sequence (Figure 3.1a-b). The other two sequences remained separate and served as our baseline condition. While being scanned with fMRI, participants anticipated upcoming environments one to four steps into the future on both the non-integrated sequence and the integrated sequence; the latter required participants to mentally traverse novel transitions to make correct predictions (Integration Task; Figure 3.1c). Participants completed the Integration Task across four runs, allowing us to test how representations in the hippocampus are reconfigured to form integrated sequences over time. We used multivariate patterns of activity in the hippocampus to determine (1) whether the entire integrated sequence was represented more similarly than the non-

integrated sequence, (2) the temporal dynamics by which hippocampal sequence representations were reconfigured, and (3) the degree to which sequence reconfiguration supported prediction of novel trajectories.

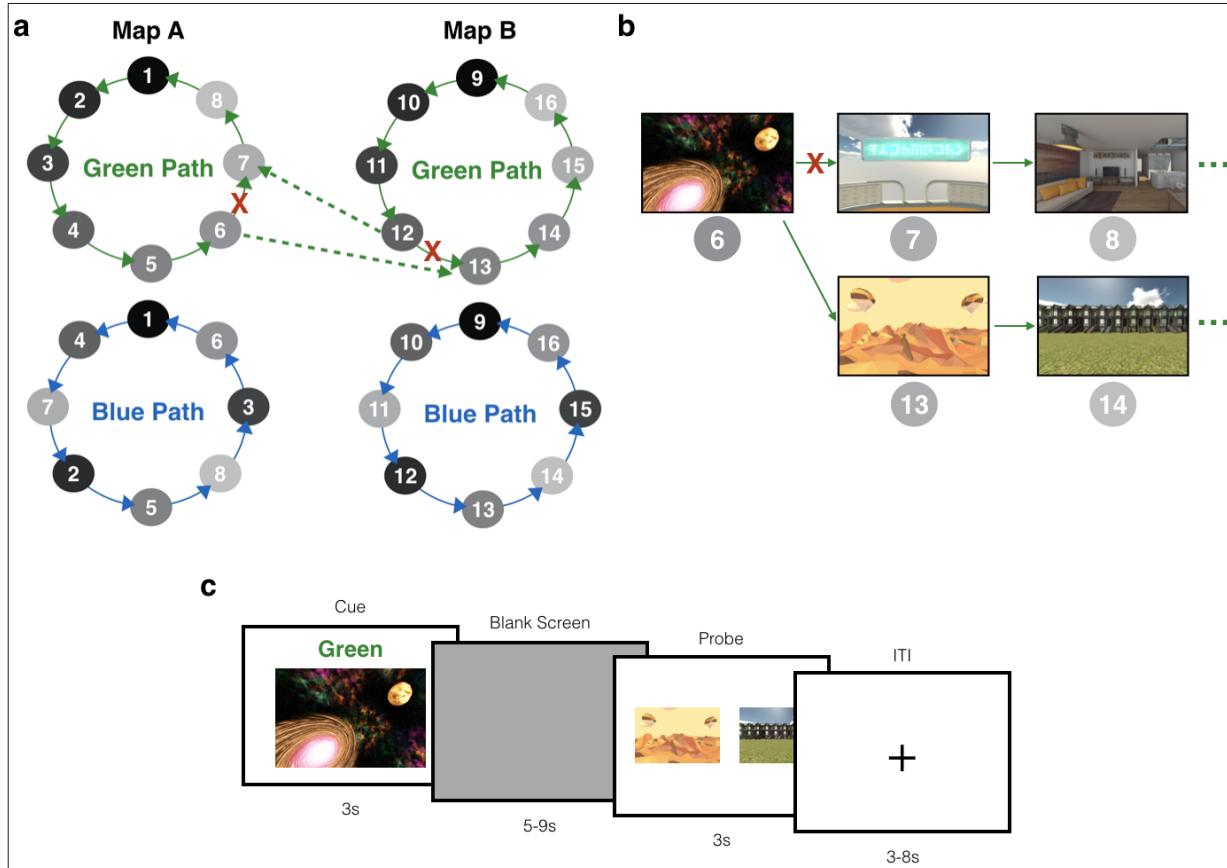


Figure 3.1: Sequence Structure and Integration Task. (a) Sequence structure. Participants learned four sequences of eight environments, indicated by the gray nodes. Participants learned four sequences in total: one green and blue path (green and blue solid lines) with a set of eight environments (Map A), and another green and blue path with a different set of eight environments (Map B). Following Sequence Learning and the Anticipation Task, participants learned that one of the environments in Map A was connected to one of the environments in Map B (dashed lines), and one of the environments in Map B was connected back to Map A, on either the green path or the blue path (counterbalanced across participants). Participants were told that the old connections between the bridge environments no longer worked and they could only use the new connections they had just learned, creating a single, integrated map. Participants were told that the non-integrated path (green or blue, counterbalanced across participants) remained the same as pre-integration. (b) Bridge Environment Example. Participants viewed one of the environments from Map A on either the green or blue path and were told that it now connected to a room in Map B. In this example, the participant would be shown environment 6 from Map A on the green path. They would be told that environment 6 now connects to environment 13 in Map B on the green, and that they could only follow this connection (thus going from

environment 6 to 13 to 14 and so on). They were told that the old connection from environment 6 to environment 7 on the green path no longer works (red X indicates broken connection) and that they could only follow the new connections, forcing participants to integrate Map A and Map B together on the green path. **(c) Integration Task.** Participants then completed an integration task in which they made anticipatory judgements inside the MRI scanner. Participants were cued with a 2D image of an environment from one of the sequences along with a path cue (Green or Blue) for 3 seconds. They then saw a blank screen for a variable duration of 5 to 9 seconds during which they were told to prepare their response. Participants were then probed with two images of upcoming environments and had 3 seconds to indicate which of the two environments was coming up sooner in the cued sequence, relative to the cue image. The correct answer could be 1 to 4 steps away from the cue image. On the integrated path (green or blue, counterbalanced across participants), participants could be cued with a room from one map (e.g. Map A) and have to anticipate upcoming rooms across the bridge environment into the other map (e.g. Map B).

3.2 Methods

Dataset

We used the dataset previously reported in Chapter 2 of this dissertation to investigate how representations of temporal structure are updated. In this dataset, thirty-two participants (21 female, 19-35 years old, mean = 24.17, sd = 4.11, 13-29 years of education, mean = 16.85, sd = 3.76) learned sequences of environments in immersive virtual reality (VR) and then, during fMRI scanning, anticipated upcoming environments in the different sequences. The order of the tasks was as follows: Sequence Learning, Anticipation Task, Integration Task, and Localizer Task. The analyses in the current chapter will mainly focus on the Integration Task (see below).

Task Overview

The Sequence Learning, Anticipation, and Localizer Tasks are briefly described here and described in full in Chapter 2. In the Sequence Learning Task, participants learned two sequences (“Green Path” and “Blue Path”) within each of two maps (Map A and Map B) in immersive VR (Figure 3.1a). Map A and Map B contained eight distinct environments each. Within each map, the Green Path and Blue Path contained the same environments in a different order. Participants learned the order of the four sequences of environments by generating stories

and then experiencing the environment sequences in immersive virtual reality using an Oculus Rift. Participants were then given a recall test to ensure they had learned all four sequences.

Participants returned 1 day later and completed the Anticipation Task in the MRI scanner (two runs, 32 trials per run). In the Anticipation Task, participants were shown an environment cue from a given map (A vs B) and a path cue (Green vs Blue). Participants then anticipated upcoming environments in the order of the cued map and path. The correct answer could be one to four steps into the future. See Chapter 2 for a full description of Sequence Learning and the Anticipation Task.

Integration Task

Following the Anticipation Task, participants completed the Integration Task, which is the task primarily analyzed in this chapter. In the Integration Task, participants were first told that one of the environments in Map A was now connected to one of the environments in Map B on either the Green or Blue Path. One of the environments in Map B also connected back to Map A, creating a single integrated path encompassing all the environments in both maps (see Figure 3.1a-b). The integrated path (Green or Blue) was counterbalanced across participants, with the other path serving as a control non-integrated path. For example, if a participant learned that Map A and Map B were integrated on the Green Path, the Blue Path would be the non-integrated path. The environment that connected map A to map B (a “bridge environment”) was randomly selected, as was the connecting environment in map B. The sequence of environments continued through all environments map B (in their previously-learned order), and then linked back to map A at the environment that had originally followed map A’s bridge environment, forming an integrated circle of all 16 environments (Figure 3.1a-b). To learn these new connections, participants viewed a video of the bridge environment in Map A transitioning to Map B on the

integrated path (Green or Blue). They then viewed a video of the bridge environment in Map B transitioning back to Map A. Participants were told to think about how the new connections could be integrated into their existing stories to link the environments in order, and then given a quiz in which they had to recall the novel connections linking Map A and Map B from memory.

Following the quiz, participants completed the Integration Task during fMRI (Figure 3.1c). The Integration Task was a modified version of the Anticipation Task in which participants anticipated upcoming environments on the integrated and non-integrated paths (four runs, 24 trials per run). Similar to the Anticipation Task, participants were cued with an environment from one of the maps (A or B) along with a path cue (Green or Blue) for 3 seconds. This cue indicated the starting point and sequence on that trial. Participants then viewed a blank (gray) screen for a variable duration (five to nine seconds). Then, participants were presented with two images of upcoming environments and were told to judge which of the two environments was coming up sooner in the cued sequence, relative to the cue image. Participants were given three seconds to make this judgment. This relatively short response deadline was implemented to encourage participants to use the blank screen period to generate predictions along the cued path in preparation for the forced choice decision. The correct answer could be one to four steps away from the cue image. The incorrect answer could be a maximum of five steps away from the cue image.

Critically, on the integrated path (Green or Blue, counterbalanced across participants), participants were told to use the new connections they just learned that connected Map A and B to each other. Thus, participants could be cued with a room from Map A and the correct answer could be in Map B (Figure 3.1c). On the non-integrated path, Map A and Map B remained separate, and the sequences were no different than in the Anticipation Task. There was a

uniformly sampled three to eight second jittered inter-trial interval (ITI), during which participants viewed a fixation cross. At the end of each run, there was a 60 second rest period during which participants viewed a blank screen.

There were four runs of the Integration Task. In each run, participants were cued with every environment from Map A and B on the Integrated Path (sixteen environments across both maps) and were cued with half of the environments from Map A and B on the non-integrated path (eight environments across both maps) for a total of 24 trials per run (96 trials total). In the probe phase, the correct answer was equally distributed across steps into the future (one to four). The incorrect answer was randomly sampled to be one to four steps away from the correct answer (two to five steps away from the cue). Within a run, participants completed trials from the integration and no integration conditions, blocked by condition. In each run, participants completed the Integration Task for all the environments in the integrated condition (16 trials), and half the environments in the no integration condition (8 trials). The order of the cues was randomized and intermixed across map (A or B) within each block. The order of the integrated vs non-integrated blocks was also randomized across runs and participants. A single run of the Integration Task was approximately 9 minutes, for a total of 36 minutes across all runs.

Following the Integration Task, participants completed four runs of the Localizer Task to obtain environment-specific template patterns of brain activity. In the Localizer Task, participants viewed all environments from the experiment in a randomized order and without a path cue (Green or Blue), allowing us to obtain a context-independent pattern of brain activity for each environment. The Localizer Task is described in full in Chapter 2.

Behavioral Analysis

Behavioral data were analyzed in the R programming language using t-tests and generalized linear and linear mixed effects models (GLMMs and LMMs, *glmer* and *lmer* function in the *lme4* package; Bates et al., 2014). For analyses that modeled multiple observations per participant, such as accuracy or response time on a given trial, models included random intercepts and slopes for all within-participant effects. All response time models examined responses on correct trials only.

We first ensured that participants performed effectively during the Integration Task overall, in both the integration and no integration conditions, by comparing response accuracy on the Probe Screen (see Figure 3.1c) to chance performance (50%) using a one-sample t-test. We then conducted two one-sample t-tests comparing performance on the Integration Task to chance within each integration condition (integration and no integration).

Next, we examined how performance in the integration and no integration conditions changed across runs on the Integration Task. We modeled accuracy as a function of condition (-0.5 = no integration, 0.5 = integration), run (-0.75 = 1, -0.25 = 2, 0.25 = 3, 0.75 = 4), and their interaction using a GLMM. We used the following R-based formulas (where “participant” indicates participant number):

```
glmer(correct ~ condition*run + (1+condition*run|participant), family = "binomial",  
data)
```

Within the integration condition, there were three different trial types: 1) trials in which both the probes were in the same map as the cue (same-map trials; these do not require crossing the bridge environment), 2) trials in which one of the probes was in a different map than the cue (one-different trials; these trials also do not require crossing the bridge environment; the

environment in the same map is always the correct answer), and 3) trials in which both of the probes were in a different map than the cue (different-map trials; these trials require crossing the bridge environment; Figure 3.2b). We tested whether accuracy differed as a function of trial type (0 = same-map, 1 = one-different, 2 = different-map), run (-0.75 = 1, -0.25 = 2, 0.25 = 3, 0.75 = 4), and their interaction. Trial type was treated as a factor, allowing us to investigate how performance changed on one-different and different-map trials, relative to same-map baseline trials.

```
glmer(correct ~ trialType*run + (1+trialType*run|participant), family = "binomial",
       data, subset = (condition == "Integration))
```

fMRI Analysis

We first preprocessed MRI data using fMRIPrep 1.5.2 (Esteban, Markiewicz, et al. 2018; Esteban, Blair, et al. 2018). MRI acquisition and preprocessing steps were identical to those reported in Chapter 2. After preprocessing, all fMRI analyses were performed in Python and R. Pattern similarity analyses were performed using custom code in Python 3. Statistical analysis comparing pattern similarity values across conditions, correlations between fMRI results and behavior, and visualizations were performed using custom code in R.

Conjunction ROI Definition

Our central region of interest (ROI) was the hippocampus. We defined our hippocampus ROI by looking for voxels that reliably responded to each environment in the experiment. First, we determined which voxels showed reliable, environment-specific patterns of activity during the Localizer Task, using an approach that identifies voxels that respond reliably to different conditions across runs of an experiment (Tarhan & Konkle, 2020). The analysis to identify environment-reliable voxels is described in full in Chapter 2. We then defined an anatomical

hippocampus ROI from the Harvard-Oxford probabilistic atlas in FSL (threshold: $p = 0.50$), resampled onto the same MNI grid as the functional data (MNI152NLin2009cAsym). Finally, we intersected the hippocampus ROI with our map of environment-reliable voxels (see Chapter 2) to create a conjunction ROI.

Integration Task Analysis

We conducted GLMs predicting whole-brain univariate BOLD activity from behavioral and nuisance regressors from the Integration Task using Python. For each participant, we modeled BOLD activity separately for each run of the Integration Task with regressors for the cue, blank screen, and probe periods, separately for each cued environment in Map A and B (1 to 16) and for each path (Green Path and Blue Path). This resulted in a total of 32 task regressors for each phase (cue, blank screen, probe) of the Integration Task per run (4 runs, 16 environments across Map A and Map B, with each environment modeled separately for the Green Path and the Blue Path). We also included nuisance regressors in the same model (translation and rotation along the X, Y, and Z axes and their derivatives, motion outliers as determined by fmriprep, CSF, white matter, framewise displacement, and discrete cosine-basis regressors for periods up to 125 seconds). The resulting beta weights for our task regressors for each run were examined within our hippocampal conjunction ROI.

Analyses were conducted on each trial of the Integration Task to examine how hippocampal sequence representations changed across maps in the integration vs no integration condition. For each participant, we obtained the correlation between (1) the blank screen pattern of activity on each trial and (2) the blank screen pattern of activity for all other trials in the same condition (integration or no integration) but from the different map (A or B) (Figure 3.3a). For example, if a given trial was from the integration condition in map A, we obtained the correlation

between that trial's blank screen activity patterns and the blank screen activity patterns for trials from the integration condition in map B. To explore how across-map pattern similarity changes across the Integration Task while accounting for temporal autocorrelation within a run, we conducted this analysis across neighboring runs separately for early and late runs. For example, if a trial was from run 1, we obtained the correlation with trials from the same condition and different map in run 2 and vice versa (Early Runs). If a trial was from run 3, we obtained the correlation with trials from the same condition and different map in run 4 and vice versa (Late Runs) (Figure 3.3a).

The resulting correlations indicate the similarity between patterns of activity for Map A and Map B across runs, with higher values indicating more similarity between maps. To test for significant differences in pattern similarity in the integrated vs non-integrated condition, we modeled across-map pattern similarity as a function of condition using a paired samples t-test. We then ran follow up paired sample t-tests to assess differences in pattern similarity for early and late runs in the integration condition, compared to the averaged pattern similarity across all runs in the no integration condition. We opted to average across runs in the no integration condition because (1) our preliminary analyses did not find differences in pattern similarity across runs and (2) we did not have enough trials to reliably estimate run-wise pattern similarity values in the no integration condition (8 trials per run; Dimsdale-Zucker & Ranganath, 2018).

Relationship Between Integration Task Pattern Similarity and Behavior

We determined whether map integration in the hippocampus was related to performance, and whether these representational changes were related to memory updating over the course of the Integration Task. To address these questions, we performed an individual differences analysis by obtaining the Spearman rank-order correlation between (1) participants' accuracy in the

integration condition, separately for early and late runs and (2) their hippocampal across map pattern similarity in the integration condition, separately for early and late runs. Finally, we conducted an exploratory analysis to examine integration behavior more specifically, by focusing on trials in which participants had to cross the bridge environment to get to the correct answer (different-map trials, see Behavioral Analysis). To that end, we obtained the Spearman rank-order correlation between participants' accuracy on different-map trials in early runs and their across-map pattern similarity in early runs.

3.3 Results

Integration Task Performance

As a first step, we determined whether participants performed well on the Integration Task. On average, participants chose the probe that was coming up sooner relative to the cued environment 85.97% of the time, which was significantly above chance performance (50%, $t(31) = 51.057$, $p < 0.000001$). Performance was high and significantly above chance in both the non-integration condition (average accuracy = 89.453%, $t(31) = 59.774$, $p < 0.000001$) and the integration condition (average accuracy = 84.228%, $t(31) = 41.814$, $p < 0.000001$).

To determine how sequence memories are updated with time and experience, we next determined how performance on the Integration Task varied by condition and run. Performance was significantly higher in the no integration compared to the integration condition ($\beta = -0.389$, 95% CI = [-0.741, -0.037], $p = 0.03$; Figure 3.2a). Overall, performance did not significantly differ across runs of the Integration Task ($\beta = 0.244$, 95% CI = [-0.024, 0.513], $p = 0.074$; Figure 3.2a) However, there was a significant interaction between condition and run ($\beta = 0.742$, 95% CI = [0.209, 1.276], $p = 0.006$; Figure 3.2a), such that performance increased across runs in the integration condition ($\beta = 0.62$, 95% CI = [0.329, 0.911], $p = 0.00003$), but

not in the no integration condition ($\beta = -0.122$, 95% CI = [-0.563, 0.320], $p = 0.588$). In fact, performance in the integration condition reached that of the no integration condition by the end of the task ($\beta = -0.150$, 95% CI = [-0.874, 0.573], $p = 0.684$). Strikingly, however, even in the first run of the Integration Task, performance was high and significantly above chance in the integration condition (average accuracy = 78.52%, $t(31) = 30.538$, $p < 0.000001$), suggesting that participants rapidly reconfigured their memories to reflect the updated temporal structure.

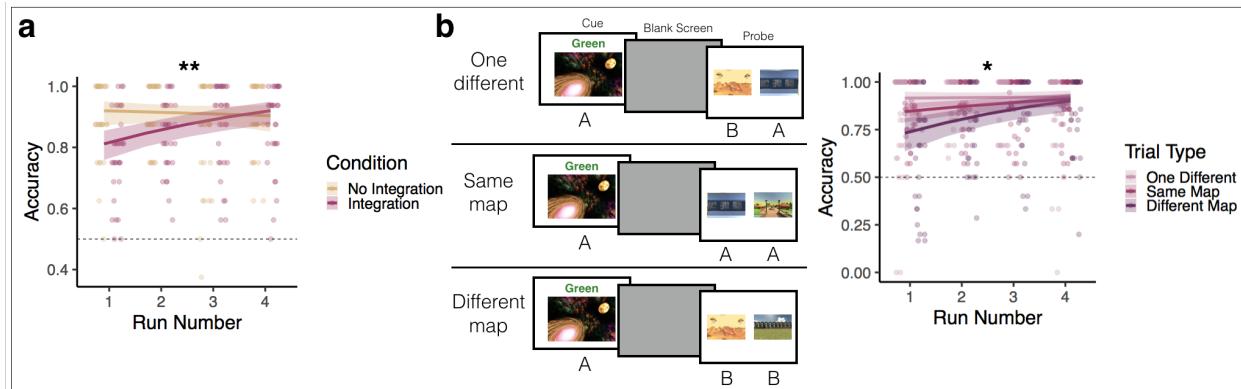


Figure 3.2: Integration Task Performance. **(a)** Overall Behavioral Performance. Participants accurately anticipated upcoming environments in the cued sequence in both the integration and the no integration conditions. Accuracy was higher in the no integration compared to the integration condition. However, accuracy improved over the course of the task, specifically in the integration condition, such that performance was similar across conditions by run 4. Notably, performance in the integration was still high even in run 1, suggesting rapid sequence updating. Solid pink lines and error ribbons indicate model predictions with 95% confidence intervals for the integration condition; pink points indicate individual participants' performance in the integration condition for each run of the Integration Task. The same conventions are used for the no integration condition, plotted in yellow. ** $p < .001$ **(b)** Performance by Trial Type in the Integration Condition. There were three trial types within the integration condition (left): one-different trials (top), same-map trials (middle), and different-map trials (bottom). On one-different trials, participants were cued with an environment from one map (A or B) and probed with one environment from the same map and one environment from the different map. On same-map trials, participants were cued and probed with environments from the same map. On different-map trials, participants were cued with an environment from one map (A or B) and probed with two environments from the other map. Performance was highest on one-different, followed by same-map, and then different-map trials. Performance improved the most across runs on different-map trials. Solid light pink lines and error ribbons indicate model predictions with 95% confidence intervals for one-different trials in the integration condition; light pink points indicate individual participants' performance on one-different trials for each run of the Integration Task. The same conventions are used for the same-map trials, plotted in dark pink, and different-map trials, plotted in purple. ** $p < .001$ * $p < 0.05$

To further probe the accuracy improvement in the integration condition, we next examined how performance changed as a function of run and trial type in the integration condition. Trials in the Integration Task differed in difficulty based on whether participants needed to mentally traverse between maps to correctly anticipate upcoming environments. There were three trial types within the integration condition: same-map trials, one-different trials, and different-map trials (see Methods; Figure 3.2b). For example, trials in which both the probes were in the same map as the cue (same-map trials) should be easier than trials in which the probes were in the different map as the cue (different-map trials), because different map trials required participants to cross the bridge environment to correctly respond. Trials in which only one of the probes is in the different map from the cue (one-different trials) should be comparatively easier than both same and different map trials, because a participant should be able to reject the incorrect probe on the basis of map identity alone, rather than temporal order. Specifically, because the correct answer was at most four steps away, it is never possible for the different-map environment to be closer than the same-map environment.

Indeed, accuracy on the Integrated Path was highest on one-different trials (mean = 88.9%), followed by same-map trials (mean = 87.8%), and lowest on different-map trials (mean = 78.5%). There was a trend toward lower accuracy on different-map compared to same-map trials ($\beta = -0.368$, 95% CI = [-0.744, 0.007], $p = 0.054$; Figure 3.2b), but there was no difference between one-different and same-map trials ($\beta = 0.284$, 95% CI = [-0.139, 0.708], $p = 0.188$). Critically, there was a trial type by run interaction ($\beta = 0.585$, 95% CI = [0.078, 1.091], $p = 0.024$; Figure 3.2b), such that performance on different-map trials increased compared to same-map trials across runs. There was no such interaction between same-map and one-different trials across runs ($\beta = -0.002$, 95% CI = [-0.650, 0.645], $p = 0.994$). In the first

run of the Integration Task, accuracy was significantly lower on different-map compared to same-map trials ($\beta = -0.8726$, 95% CI = [-1.361, -0.384], $p = 0.0005$), but performance on different-map trials reached that of same-map trials in the final run ($\beta = -0.086$, 95% CI = [-0.701, 0.529], $p = 0.784$). This shows that trials in which participants had to mentally traverse the bridge environments linking Map A and B improved the most over the course of the Integration Task.

Taken together, our behavioral results suggest that participants updated their memories of temporal structure after a single learning episode to successfully make anticipatory judgements. Importantly, sequence memories were rapidly reconfigured, such that they reflected the new sequence structure even in the first run of the Integration Task. They nevertheless continued to improve over the course of the experiment, implying the presence of both rapid and gradual updating of learned representations.

Reconfiguration of Hippocampal Sequence Representations

Having shown that individuals can rapidly integrate sequence memories in behavior, we next explored the reconfiguration of sequence representations in the hippocampus. We obtained the correlation between multivariate patterns of activity elicited during the blank screen of the Integration Task for trials in which cue environments came from Map A vs Map B, separately for the integration and no integration conditions (see Methods; Figure 3.3a). We found that across-map pattern similarity was higher in the integration condition compared to the no integration condition ($t(31) = 2.070$, $p = 0.041$; Figure 3.3b), indicating that hippocampal representations of the entire Map A and Map B sequences became more similar when they were connected by the bridge environments.

To probe how hippocampal sequence representations are updated over time, we separately compared across-map pattern similarity values in the integration condition for early and late runs to the averaged across-map pattern similarity in the no integration condition. We did not examine across-map pattern similarity separately for early vs late runs in the no integration condition because we did not have enough trials to reliably estimate the pattern similarity values in this condition (eight trials per run).

In early runs (runs 1 and 2), across-map pattern similarity was significantly greater in the integration condition compared to the average in the no integration condition ($t(31) = 2.665, p = 0.012$; Figure 3.3b). In late runs (runs 3 and 4), across-map pattern similarity was not different in the integration condition compared to the average in the no integration condition ($t(31) = 1.304, p = 0.203$; Figure 3.3b). This provides preliminary evidence that the difference between map similarity in the integration and no integration conditions was driven by the runs immediately after learning the bridge environment, suggesting that hippocampus rapidly reconfigures sequence representations in light of new information.

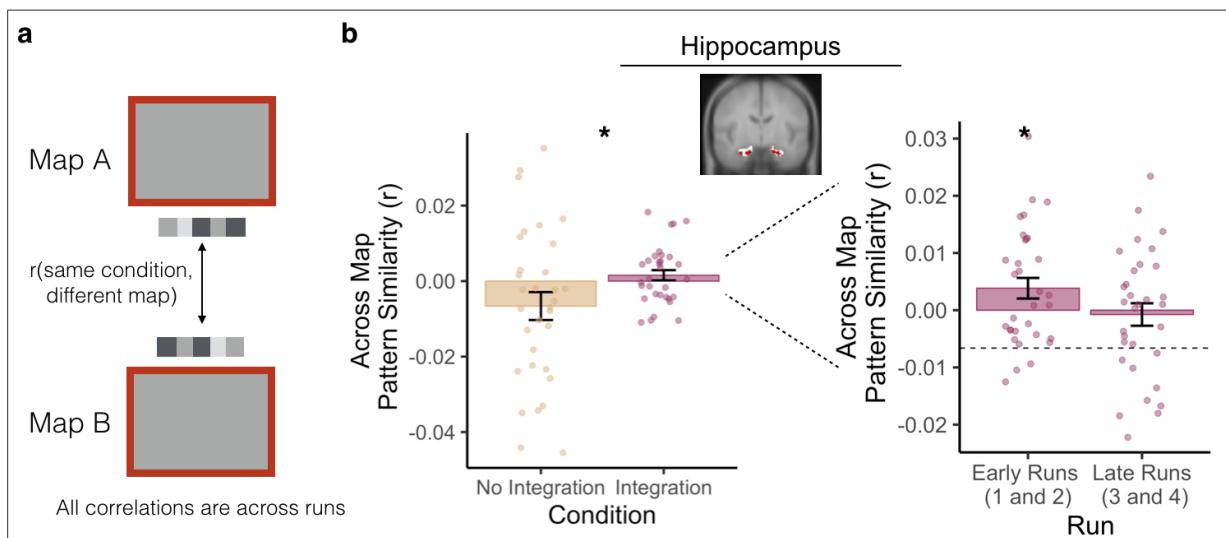


Figure 3.3: Sequence Reconfiguration in Hippocampus. (a) Pattern similarity analysis. To determine across-map pattern similarity, we obtained the correlation between (1) the blank screen pattern of activity on each trial and (2) the blank screen pattern of activity for all other trials in the

same condition (integration or no integration) but from the different map (A or B) for each participant. Correlations were run across neighboring runs to account. For example, we would correlate a trial from run 1 with all trials from the same condition and different map in run 2 and vice versa (Early Runs). The same analyses would be run between runs 3 and 4 (Late Runs). We refer to the resulting correlation values as across map pattern similarity, where higher values indicate more similarity between Map A and Map B representations. **(b)** Pattern similarity results. We investigated map pattern similarity within our hippocampus conjunction ROI, collapsed across early and late runs. Across map pattern similarity was higher in the integration condition compared to the no integration condition. We then compared across map pattern similarity in early and late runs of the integration condition, separately, to the averaged across-run similarity in the no integration condition. Across map pattern similarity was higher in early runs of the integration condition compared to the averaged no integration condition. There was no significant difference in late runs. Bars indicate average across map similarity across participants, error bars indicate standard error of the mean, and small, transparent points indicate each participant's across map pattern similarity. Dashed line indicates the averaged across-map pattern similarity across runs in the no-integration condition. * $p < .05$

Rapid Sequence Reconfiguration Supports Memory Updating

We explored whether rapid sequence updating in the hippocampus was related to Integration Task performance. To test this, we obtained the Spearman's rank-order correlation between early-run hippocampal across-map similarity in the integration condition and early-run performance on the Integration Task (see Methods). We then calculated the same Spearman's rank-order correlation for late runs. There was a trending relationship between early-run across-map similarity and early-run performance in the integration condition ($\rho = 0.345$, $p = 0.053$; Figure 3.4a). There was no relationship between late-run across-map similarity and late-run performance in the integration condition ($\rho = 0.213$, $p = 0.241$).

Finally, we investigated whether early-run across-map similarity in the hippocampus was related to early-run performance in each of the three integration condition trial types (one-different, same-map, different-map, see above). There was no relationship between early-run across-map similarity in the integration condition and early-run performance on one-different trials ($\rho = 0.126$, $p = 0.493$), or same-map trials ($\rho = 0.135$, $p = 0.460$). However, there was a positive correlation between early-run across-map similarity and early-run performance on

different-map trials ($\rho = 0.385$, $p = 0.030$; Figure 3.4b), suggesting that hippocampal reconfiguration of sequences is related to performance on trials that require integration (i.e. trials in which participants had to mentally traverse the bridge environment).

Taken together, our results suggest that the hippocampus rapidly updated sequence representations in light of new information, such that maps that were previously separate became integrated. The degree to which sequences were rapidly integrated in the hippocampus predicted anticipation performance along the integrated path, specifically on trials in which participants needed to traverse between maps.

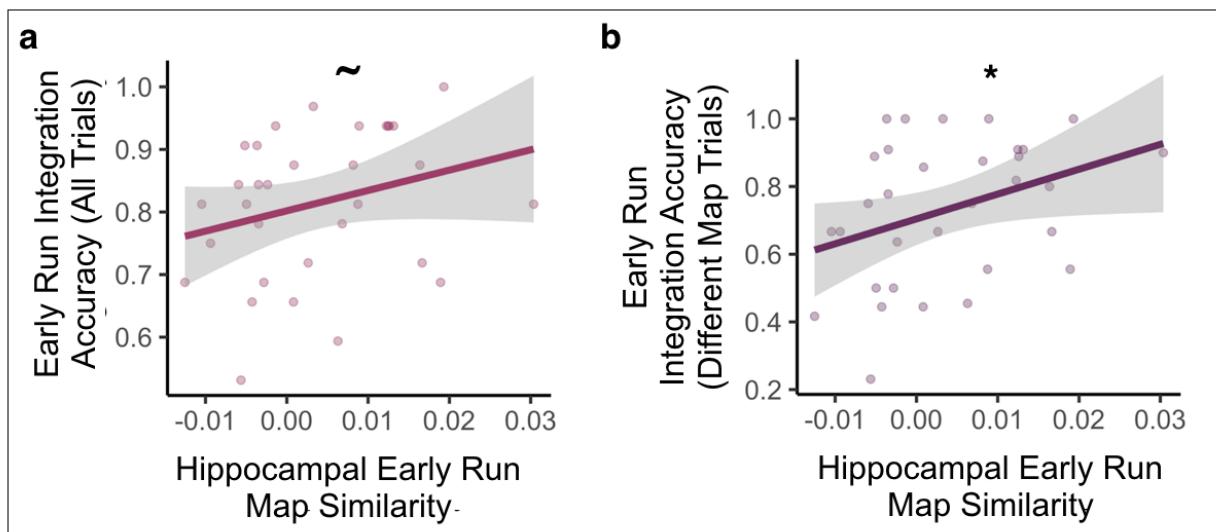


Figure 3.4: Rapid sequence configuration in the hippocampus is related to integration performance. (a) There was a trending relationship between across-map similarity in the hippocampus and performance on the integration condition of the Integration Task in early runs (runs 1 and 2). (b) Higher across map-pattern similarity in the hippocampus was related to performance on different-map trials in the integration condition in early runs (runs 1 and 2), suggesting that rapid sequence updating in the hippocampus is related to performance on trials that require integration across maps. Lines and gray error ribbons indicate the correlation with 95% confidence intervals; points indicate each participant's across-map pattern similarity and Integration Task performance. $\sim p < .10$, $*$ $p < .05$

3.4 Discussion

We investigated how temporally extended sequences become integrated in the brain to allow novel predictions. Participants rapidly learned novel transitions linking two previously

separate sequences, allowing them to successfully anticipate upcoming environments on the new integrated sequence even on the first post-integration run. They also continued to improve over the course of the experiment, suggesting both rapid and gradual updating of sequence representations. Multivariate fMRI analyses in the hippocampus revealed that the previously separate sequences became more similar when they were integrated, compared to a no-integration baseline condition. Importantly, this difference between conditions was strongest for early runs of the task, suggesting rapid reconfiguration of sequence structure in the hippocampus. Finally, hippocampal integration of sequences in early runs was related to performance on the integration task, particularly on trials in which participants had to mentally traverse the novel transition to make predictions about upcoming environments.

Our findings that the hippocampus rapidly reconfigured sequence representations are consistent with decades of research showing that the hippocampus does not just acquire direct associations, but rapidly generalizes across discrete episodes (Eichenbaum, 2000; Schlichting & Preston, 2015; Zeithamova & Bowman, 2020). Indeed, the flexibility of hippocampal representations allows such functions as inferential reasoning and insight among events that were not directly experienced together (Milivojevic et al., 2015; Schlichting et al., 2015; Shohamy & Wagner, 2008; Wimmer & Shohamy, 2012; Zeithamova et al., 2012). However, the majority of past work has focused on relatively discrete or short episodes, whereas our sequences were temporally extended (eight environments long pre-integration and 16 environments long post-integration). Thus, our findings extend prior work by showing that the hippocampus quickly reconfigured extended temporal structure, integrating representations of the environments that were far away from the novel transition. The increased pattern similarity across maps in the integration condition suggests that the whole internal model of the sequence was reconfigured

after learning a single novel transition, even though distant parts of the sequences were never directly linked. Future analyses can probe this effect further by looking at across-map pattern similarity as a function of distance to the linking environment. Thus, this work sheds light on how internal models of complex temporal structure are updated given new experiences.

We found that the increased hippocampal similarity for the integrated compared to non-integrated sequences was strongest in early runs. This is striking, considering that behavior on the Integration Task improved from early to late runs. One possible explanation for why across-map pattern similarity in the hippocampus did not continue to increase over runs is that integrated sequence representations were transferred to another brain region. For example, mPFC plays a role in memory integration (Audrain & McAndrews, 2022; Schlichting et al., 2015; Tompary & Davachi, 2017, 2022; Zeithamova et al., 2012), and representing schemas (Baldassano et al., 2018; Kesteren et al., 2012; Tse et al., 2011), suggesting that it too creates holistic representations of events across multiple discrete episodes. Further, recent work has shown that repeated retrieval may act as a form of memory consolidation (Antony et al., 2017), increasing both behavioral (Lifanov et al., 2021) and neural (Yu et al., 2022) markers of consolidation. Thus, the repetition of the Integration Task across runs may have led to the transfer of integrated sequence representations from hippocampus to mPFC, consistent with theories of memory consolidation (Robin & Moscovitch, 2017; Sekeres et al., 2018; Winocur et al., 2010). Future analyses can investigate this further by examining across-map pattern similarity in the mPFC in late runs.

Behaviorally, participants rapidly updated their sequence representations to successfully make predictions in the integrated sequences, but also continued to improve throughout the integration task. This suggests that rapid vs gradual updating of sequence representations need

not be mutually exclusive, in behavior or in the brain. Indeed, there are abundant demonstrations that there are multiple learning mechanisms across the brain that could support the rapid acquisition of new information alongside a gradual extraction of regularities (McClelland et al., 1995; Poldrack et al., 2001; Schapiro, Turk-Browne, et al., 2017), with even traditionally gradual systems showing learning within a single session (Hebscher et al., 2019). Both rapid and gradual mechanisms could underlie the integration of internal models of extended temporal structure to support the emergence of flexible behavior on multiple timescales.

These gradual learning systems may reside outside of the hippocampus (McClelland et al., 1995; Poldrack et al., 2001) or co-exist with rapid learning mechanisms within the hippocampus itself (Schapiro, Turk-Browne, et al., 2017). For example, prior work has shown that reactivation of sequences in the hippocampus during periods of awake rest go beyond trajectories that have been directly experienced (Gupta et al., 2010; Y. Liu et al., 2019; Ólafsdóttir et al., 2015), which may lead to the restructuring of representations over time (Foster, 2017). Further, coordinated replay of experience in the hippocampus and the cortex is proposed to underlie time-dependent changes to memory traces (Schlichting & Preston, 2016; Tambini et al., 2010; Tompary et al., 2015; van Kesteren et al., 2010), and has been specifically shown to promote the integration of overlapping memories (Audrain & McAndrews, 2022; Schlichting & Preston, 2016; Tompary & Davachi, 2022). Replay has been further theorized to specifically improve distally learned multistep predictions, especially after the transition structure of temporally extended experiences change (Momennejad et al., 2017, 2018). Thus, a likely mechanistic candidate for gradual sequence reconfiguration over time, in behavior and in the brain, is replay during awake rest. In our experimental design, we included brief rest periods following each run of the integration task (see Methods). In future analyses, we can look for

hippocampal reactivation of trajectories, as well as coordinated hippocampal-cortical reactivation, to determine how replay supports the emergence of integrated sequence structure over time.

In summary, we show that learning a single, new transition rapidly reconfigures sequence representations in the hippocampus. The extent to which sequences were rapidly updated was related to the behavioral ability to make multistep predictions about upcoming environments across sequences. Although participants quickly learned the new sequence structure, their ability to make predictions about upcoming environments across sequences also improved with time, implying both rapid and gradual updating of learned representations. This opens the door for future analyses investigating how internal models are both rapidly and gradually updated in the brain. Thus, our work sheds light on how internal models of the world that span multiple timescales are represented and updated, and how they are used to guide flexible and adaptive behavior.

Discussion

Overview

In this dissertation, I aimed to understand how temporal structure was represented in memory and the brain to enable multi-step predictions. Across three areas of research, I approached this question using novel experimental designs and analysis techniques to probe mnemonic and neural representations of structure across multiple timescales.

In Chapter 1, I explored how individuals build durable models of temporal structure to guide multi-step anticipation by testing predictions immediately and at a delay. Consolidation increased the efficiency of anticipation, particularly for events further in the future, but diminished access to perceptual features. To further probe mnemonic representations of temporal structure, I created a novel response time modeling approach and found that maintaining a link-based (vs. cue-based) model of the sequence after consolidation improved anticipation accuracy. These suggest that consolidation may promote efficient and durable models of temporal structure, thus facilitating anticipation of future events.

In Chapter 2, I used fMRI to examine how temporal structure was flexibly represented across the brain to support multi-step anticipation. Multivariate fMRI analyses revealed bidirectional representations of temporal structure within a context during prediction, with graded representations of environments in the forward and backward direction. Visual cortex primarily represented the current environment, whereas the hippocampus represented further environments in a graded fashion. Beyond visual cortex and hippocampus, a hierarchy of temporal structure was also apparent within individual visual regions: successively more anterior aspects of PPA and RSC represented further environments into the past and future. Together, this shows that neural representations of temporal structure are not necessarily skewed toward future

states, instead showing simultaneous past and future representations along a posterior to anterior hierarchy.

Chapter 3 used fMRI to investigate how temporally extended sequences become integrated in the brain in light of new information to allow novel predictions. Participants rapidly learned novel transitions linking two previously separate sequences, allowing them to successfully anticipate upcoming environments on the new integrated sequence even on the first post-integration run. They also continued to improve over the course of the experiment, suggesting both rapid and gradual updating of sequence representations. Multivariate fMRI analyses in the hippocampus revealed that, compared to a baseline of no integration, previously separate sequences became more similar when new connections were established between them. Importantly, this difference between conditions was strongest for early runs of the task, suggesting rapid reconfiguration of sequence structure in the hippocampus. These results shed light on how internal models of the world that span multiple timescales are represented and updated to guide adaptive behavior.

The work presented here advances our understanding of prediction and the adaptive nature of memory. We extended past work on temporal coding by investigating how multiple timescales of temporal structure are simultaneously represented across the brain, extending into both the past and the future. We further examined the flexibility of such representations by comparing sequences that had overlapping stimuli but afforded different predictions within a context. Overall, this work sheds light on how we represent durable, flexible, and temporally extended internal models of the world to guide adaptive behavior. Below, I suggest avenues for future work in light of the findings presented here.

Context-Specific Predictions

Although the idea that predictions should be relevant for the current context is intuitive, empirical research showing the context-specificity of predictions is sparse (Crivelli-Decker et al., 2023). Across all three chapters, we used the same sequence structure to assess the flexibility of context-dependent predictions. Pairs of sequences contained the same environments in a different order. Importantly, the sequences were constructed to be as distinct as possible: for a given environment the two preceding and two succeeding environments were different across the sequences. This allowed us to test the context-specificity of representations one and two steps into the future and the past. This is akin to navigating from your house to work vs navigating from your house to the grocery store: it is adaptive to prioritize events coming up on the route to work, but not the grocery store, on a weekday morning. Indeed, in Chapter 2, we found that visual cortex and insula represented context-dependent representations one step into the future but not into the past. Future work could investigate whether representations of the past are always context-independent, or if the degree to which context is flexibly prioritized changes with task demands. Another open area to explore is whether predictions further into the future lose their context-specificity: it may be the case that imminent predictions require context-dependent representations but further predictions are more general and context-independent.

In Chapter 2, we failed to find context-dependent predictions in the hippocampus. However, the hippocampus is a heterogeneous structure with well-defined regions that differ in their functional contributions to memory. Thus, another area for future work could investigate how context-dependent representations differ within diverse regions of the hippocampus. Some hippocampal subregions emphasize integration across contexts whereas others emphasize differentiation (Bein & Davachi, 2022; Brunec et al., 2020; Dimsdale-Zucker et al., 2018; Duncan & Schlichting, 2018; Molitor et al., 2021; Schlichting et al., 2015). Because of its sparse

coding scheme, dentate gyrus tends to differentiate related items within a context (Leutgeb et al., 2007; Treves & Rolls, 1992). Conversely, recurrent connections within CA3 facilitate memory retrieval based on partial cues, leading to pattern completion based on the current item rather than contextual information (O'Reilly & McClelland, 1994; Treves & Rolls, 1992). Therefore, differentiated dentate gyrus representations may favor specific, short timescale predictions, whereas CA3 may represent context-general predictions. CA1 receives input from both CA3 and the entorhinal cortex (Lavenex & Amaral, 2000). Therefore, CA1 may trade off between representing context-general predictions and representing the current moment (Bein et al., 2020). Speculatively, CA1 could potentially integrate perceptual information from the entorhinal cortex with context-general predictions from CA3 and select the most context-relevant predictions for the current moment. Such a trade-off may depend on the neurochemical state, and specifically the cholinergic state, of the hippocampus (Decker & Duncan, 2020; Duncan et al., 2012; Hasselmo et al., 1995; Hasselmo & Schnell, 1994; Tarder-Stoll*, Jayakumar*, et al., 2020). Future work can investigate how context-dependent predictions are represented within hippocampal subfields, and how they are modulated by the distance of predictions and neurochemical state of the hippocampus.

In the present body of work, we operationalized context by creating distinct sequences, in which individuals can make differing predictions about the same stimuli based on the overarching sequence context. However, context also changes over time during continuously unfolding experience. Such temporal context shapes neural representations and memory for the temporal order of events (Clewett & Davachi, 2017; Ezzyat & Davachi, 2014). Future work can investigate how predictions generated within vs across temporal contexts are differentially represented. For example, recent work has shown similar patterns of activity in CA3 for items

within an event after repetitions of a sequence (Bein & Davachi, 2022). Thus, CA3 might represent predictions into the future up to the next event boundary, or events into the past up to the preceding event boundary. This would be in line with our findings in Chapter 2 that the hippocampus represented bidirectional representations of temporal structure in the forward and backward direction. mPFC, on the other hand, exhibits common patterns of activity for different events that follow the same script, such as ordering food at a restaurant or going to an airport (Baldassano et al., 2018). mPFC may therefore represent schematic predictions that extend further into the future across event boundaries. Indeed, this proposed distinction between hippocampal and mPFC predictions would be consistent with our behavioral findings from Chapter 1. Specifically, our findings may reflect a consolidation-mediated hand-off between perceptually detailed predictions in hippocampus and further-reaching but schematic predictions in mPFC.

Multiple Representational Hierarchies in the Brain

The present studies are inspired by recent theoretical and empirical work suggesting that multiple timescales of past information are represented hierarchically, with shorter timescales in posterior regions and successively longer timescales in successively more anterior brain regions (Hasson et al., 2015). Recent work has extended these theories to show that the brain also predicts future information along a posterior to anterior hierarchy (Brunec & Momennejad, 2022; Lee et al., 2021). In line with this, we found in Chapter 2 that the brain represented *both* forward and backward temporal structure along a posterior to anterior hierarchy, such that longer timescales in both forward and backward directions were represented in more anterior regions. Note that previous work on representational hierarchies of the past investigated recently

experienced information held in working memory (Hasson et al., 2015), whereas our experiments showed past representations from long term memory.

Relatedly, other influential theories suggest that a similar posterior to anterior gradient organizes the representational granularity of memories. For example, posterior hippocampus may be specialized to represent detailed aspects of memories, whereas anterior hippocampus represents memories at a coarser level of granularity, in a more gist-like manner (Brunec et al., 2018; Poppenk et al., 2013; Strange et al., 2014; but see Thorp et al., 2022 for a differing view). mPFC, which lies more anterior to hippocampus, has been implicated in representing memories with even coarser granularity, at the level of schemas and scripts (Robin & Moscovitch, 2017; Sekeres et al., 2018).

Future work could test the extent to which representations of detail and timescale converge along these proposed hierarchies. For example, events that are closer in time may be remembered in a more detailed manner and those further in the future may be remembered in a more schematic manner (Durkin et al., 2020; Trope & Liberman, 2010). This would suggest a convergence between short timescales and details, and long timescales and abstract representations. Contrary to this prediction, in Chapter 1 we found that the extent to which perceptual features were incorporated into predictions, which may indicate the amount of detail represented in memory, did not change depending on prediction distance. However, the perceptual features in our experiment were applied to the entire image and were consistent across a given sequence, raising the possibility that they were represented in a relatively schematic way. Future studies could incorporate smaller, image-specific details to test how details, timescales, or both, shape the representations of memories along a hippocampal and cortical posterior to anterior hierarchy.

Replay as a Mechanism for Building Durable and Efficient Models of the World

In Chapter 1, we found that consolidation enhanced multi-step predictions, such that anticipatory judgements became more efficient at a delay, especially those further into the future. We additionally found that maintaining a link-based model after consolidation predicted improved performance on the anticipation task, perhaps suggesting that consolidation enhances the efficiency or stability of the links between successive items in a sequence. What mechanisms might contribute to the strengthening of internal models over a delay?

One possibility is replay: experiences are reactivated in the brain during offline periods of wakeful rest or sleep (Foster & Wilson, 2006; Schapiro et al., 2018; Skaggs & McNaughton, 1996; Tambini & Davachi, 2013), with some work suggesting that replay serves to selectively strengthen otherwise weak memories (Schapiro et al., 2018; Yu et al., 2022). Offline replay recapitulates the order of sequential experience (Foster & Wilson, 2006; Skaggs & McNaughton, 1996), which may allow it to update internal models of the world (Foster, 2017; Momennejad, 2020). Future research can test whether offline replay of the entire sequence is related to the efficiency improvement across steps into the future observed in Chapter 1. For example, it might be the case that offline sequential reactivation drives patterns of activity closer together for linked items in a sequence. This in turn might make it easier to access events in the far future if their representations become similar to the cue item. This could be one way that the brain builds efficient models of temporal structure to guide prediction.

In addition to the post-consolidation improvement observed in Chapter 1, we also saw an online improvement in multi-step predictions after environments changed in Chapter 3. Future work can similarly address how replay of the linking environment, or of the entire novel

trajectory, is related to neural and behavioral markers of rapid memory updating. Thus, investigating how sequential replay influences representations of temporal structure to guide predictions could be a fruitful avenue of research to advance our understanding of how the brain builds durable and efficient models of complex environments.

What are we Predicting Towards?

Much of the prior work investigating predictions multiple steps into the future has examined how goals shape our predictions. For example, during the explicit planning of trajectories, the hippocampus represents the goal state as well as sub-goal information (Brown et al., 2016; Crivelli-Decker et al., 2023). However, in the current body of work, there was never an explicit goal that individuals had to reach. Our sequences were circular, so there was no particular endpoint that participants were predicting towards. Furthermore, participants were not told how far into the future the correct answer would be on a given trial, so there was no clear stopping point for predicting upcoming environments. In Chapter 2, we showed bidirectional representations of temporal structure, even though participants were instructed to predict future events without a specific goal in mind. This finding may differ from prior work that emphasized anticipatory coding (Brown et al., 2016; Crivelli-Decker et al., 2023) because our circular sequences without end goals may have encouraged maintenance of both past and future information: an environment from the past could also be in the future if one predicted far enough. Future work could compare how extended temporal sequences are represented in the presence vs the absence of explicit end points to determine the extent to which goals skew representations of temporal regularities toward the forward direction.

The presence or absence of an explicit end point could also influence how sequences are represented within computational frameworks. In all three chapters, individuals predicted from a

different start location to a different end location across trials. One possibility is that a model-based representation of temporal structure, in which an individual maintains an internal representation of the entire sequence in memory and rolls out the sequence link by link at decision time (Daw & Dayan, 2014; Doll et al., 2012), may be more likely to guide behavior when start and end locations change, requiring subsampling from the larger internal model. Indeed, in Chapter 1, we found that the link model (akin to a model-based strategy) was the optimal internal model in our task, in which participants did not have an explicit start and end location. In contrast, successor representations, in which individuals cache information about predicted future states (Dayan, 1993), may be more likely to form when there is a clear start and end point, allowing representations towards the end point to become embedded in the start point. In line with this prediction, past studies have shown a preference for successor representation strategies when short sequences have a clear beginning and end (Momennejad et al., 2017). Future work can create temporally extended sequences with and without consistent start and end states to determine whether a preference for a cue or link strategy differentially explains behavior.

Conclusion

In this dissertation, I investigated how temporal sequence structure is represented in memory and in the brain to guide predictions over multiple timescales. Across three chapters, I demonstrated that (1) memory for temporal structure is enhanced over the course of consolidation to efficiently guide predictions multiple steps into the future, (2) temporal structure is represented bidirectionally in the brain across a posterior to anterior hierarchy, and (3) internal models are rapidly updated in the hippocampus after new learning, which can guide predictions

along novel trajectories. These findings advance our understanding of how internal models spanning multiple timescales are represented and used to guide efficient, flexible, and ultimately adaptive behavior.

References

- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8.
- <https://www.frontiersin.org/articles/10.3389/fninf.2014.00014>
- Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7), 1363–1377.
- <https://doi.org/10.1016/j.neuropsychologia.2006.10.016>
- Aitken, F., & Kok, P. (2022). Hippocampal representations switch from errors to predictions during acquisition of predictive associations. *Nature Communications*, 13(1), 3294.
- <https://doi.org/10.1038/s41467-022-31040-w>
- Alsbury-Nealy, K., Wang, H., Howarth, C., Gordienko, A., Schlichting, M. L., & Duncan, K. D. (2022). OpenMaze: An open-source toolbox for creating virtual navigation experiments. *Behavior Research Methods*, 54(3), 1374–1387. <https://doi.org/10.3758/s13428-021-01664-9>
- Aly, M., Chen, J., Turk-Browne, N. B., & Hasson, U. (2018). Learning Naturalistic Temporal Structure in the Posterior Medial Network. *Journal of Cognitive Neuroscience*, 30(9), 1345–1365. https://doi.org/10.1162/jocn_a_01308
- Aly, M., & Turk-Browne, N. B. (2016a). 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. (2016b). Attention Stabilizes Representations in the Human Hippocampus. *Cerebral Cortex*, 26(2), 783–796. <https://doi.org/10.1093/cercor/bhv041>
- Aminoff, E., Gronau, N., & Bar, M. (2007). The Parahippocampal Cortex Mediates Spatial and Nonspatial Associations. *Cerebral Cortex*, 17(7), 1493–1503.
<https://doi.org/10.1093/cercor/bhl078>
- Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, 17(8), 379–390.
<https://doi.org/10.1016/j.tics.2013.06.009>
- Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a Fast Route to Memory Consolidation. *Trends in Cognitive Sciences*, 21(8), 573–576.
<https://doi.org/10.1016/j.tics.2017.05.001>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Arciuli, J., & Simpson, I. C. (2012). Statistical learning is lasting and consistent over time. *Neuroscience Letters*, 517(2), 133–135. <https://doi.org/10.1016/j.neulet.2012.04.045>
- Audrain, S., & McAndrews, M. P. (2022). Schemas provide a scaffold for neocortical integration of new memories over time. *Nature Communications*, 13(1), 1.
<https://doi.org/10.1038/s41467-022-33517-0>
- Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (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>

- Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. *NeuroImage*, 75, 228–237.
<https://doi.org/10.1016/j.neuroimage.2013.02.073>
- Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, 95(3), 709-721.e5. <https://doi.org/10.1016/j.neuron.2017.06.041>
- Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of Real-World Event Schemas during Narrative Perception. *Journal of Neuroscience*, 38(45), 9689–9699.
<https://doi.org/10.1523/JNEUROSCI.0251-18.2018>
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280–289. <https://doi.org/10.1016/j.tics.2007.05.005>
- Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes Unseen: The Parahippocampal Cortex Intrinsically Subserves Contextual Associations, Not Scenes or Places Per Se. *Journal of Neuroscience*, 28(34), 8539–8544. <https://doi.org/10.1523/JNEUROSCI.0987-08.2008>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4* (arXiv:1406.5823). arXiv. <https://doi.org/10.48550/arXiv.1406.5823>
- Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, 100(2), 490–509.
<https://doi.org/10.1016/j.neuron.2018.10.002>
- Behzadi, Y., Restom, K., Liau, 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., & Davachi, L. (2022). *Event integration and temporal pattern separation: How hierarchical knowledge emerges in hippocampal subfields through learning* (p. 2022.07.18.500527). bioRxiv. <https://doi.org/10.1101/2022.07.18.500527>
- Bein, O., Duncan, K., & Davachi, L. (2020). Mnemonic prediction errors bias hippocampal states. *Nature Communications*, 11(1), 3451. <https://doi.org/10.1038/s41467-020-17287-1>
- Bellmund, J. L. S., Polti, I., & Doeller, C. F. (2020). Sequence Memory in the Hippocampal–Entorhinal Region. *Journal of Cognitive Neuroscience*, 32(11), 2056–2070. https://doi.org/10.1162/jocn_a_01592
- 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>
- Bliss, T. V. P., & Collingridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature*, 361(6407), 31–39. <https://doi.org/10.1038/361031a0>
- Bonasia, K., Blommesteyn, J., & Moscovitch, M. (2016). Memory and navigation: Compression of space varies with route length and turns. *Hippocampus*, 26(1), 9–12. <https://doi.org/10.1002/hipo.22539>
- 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., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z.-X., Grady, C., Rosenbaum, R. S., Winocur, G., Barense, M. D., & Moscovitch, M. (2018). Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans. *Current Biology*, 28(13), 2129-2135.e6. <https://doi.org/10.1016/j.cub.2018.05.016>
- Brunec, I. K., & Momennejad, I. (2022). Predictive Representations in Hippocampal and Prefrontal Hierarchies. *Journal of Neuroscience*, 42(2), 299–312.
<https://doi.org/10.1523/JNEUROSCI.1327-21.2021>
- Brunec, I. K., Robin, J., Olsen, R. K., Moscovitch, M., & Barense, M. D. (2020). Integration and differentiation of hippocampal memory traces. *Neuroscience & Biobehavioral Reviews*, 118, 196–208. <https://doi.org/10.1016/j.neubiorev.2020.07.024>
- 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>
- Caucheteux, C., Gramfort, A., & King, J.-R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*, 7(3), 3.
<https://doi.org/10.1038/s41562-022-01516-2>
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & Rosario, H. D. (2020). *pwr: Basic Functions for Power Analysis* (Version 1.3-0) [Computer software]. <https://CRAN.R-project.org/package=pwr>
- Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology*, 27(15), 2307-2317.e5. <https://doi.org/10.1016/j.cub.2017.06.057>

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>

Clarke, A., Crivelli-Decker, J., & Ranganath, C. (2022). Contextual Expectations Shape Cortical Reinstatement of Sensory Representations. *Journal of Neuroscience*, 42(30), 5956–5965.

<https://doi.org/10.1523/JNEUROSCI.2045-21.2022>

Clewett, D., & Davachi, L. (2017). The ebb and flow of experience determines the temporal structure of memory. *Current Opinion in Behavioral Sciences*, 17, 186–193.

<https://doi.org/10.1016/j.cobeha.2017.08.013>

Cohn-Sheehy, B. I., Delarazan, A. I., Reagh, Z. M., Crivelli-Decker, J. E., Kim, K., Barnett, A. J., Zacks, J. M., & Ranganath, C. (2021). The hippocampus constructs narrative memories across distant events. *Current Biology*, 31(22), 4935-4945.e7.

<https://doi.org/10.1016/j.cub.2021.09.013>

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292), 1464–1468.

<https://doi.org/10.1126/science.aaf0941>

Crivelli-Decker, J., Clarke, A., Park, S. A., Huffman, D. J., Boorman, E. D., & Ranganath, C. (2023). Goal-oriented representations in the human hippocampus during planning and navigation. *Nature Communications*, 14(1), 2946. <https://doi.org/10.1038/s41467-023-35967-6>

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, 9(2), 179–194.

<https://doi.org/10.1006/nim.1998.0395>

- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, 16(6), 693–700. <https://doi.org/10.1016/j.conb.2006.10.012>
- Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: The role of prediction and context. *Trends in Cognitive Sciences*, 19(2), 92–99. <https://doi.org/10.1016/j.tics.2014.12.004>
- Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 20130478. <https://doi.org/10.1098/rstb.2013.0478>
- Dayan, P. (1993). Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation*, 5(4), 613–624. <https://doi.org/10.1162/neco.1993.5.4.613>
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, 22(9), 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>
- Decker, A. L., & Duncan, K. (2020). Acetylcholine and the complex interdependence of memory and attention. *Current Opinion in Behavioral Sciences*, 32, 21–28. <https://doi.org/10.1016/j.cobeha.2020.01.013>
- Deuker, L., Bellmund, J. L., Navarro Schröder, T., & Doeller, C. F. (2016). An event map of memory space in the hippocampus. *eLife*, 5, e16534. <https://doi.org/10.7554/eLife.16534>
- Diba, K., & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, 10(10), 1241–1242. <https://doi.org/10.1038/nn1961>
- 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>

Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018).

CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nature Communications*, 9(1), 294.

<https://doi.org/10.1038/s41467-017-02752-1>

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, 22(6), 1075–1081.

<https://doi.org/10.1016/j.conb.2012.08.003>

Drosopoulos, S., Windau, E., Wagner, U., & Born, J. (2007). Sleep Enforces the Temporal Order in Memory. *PLOS ONE*, 2(4), e376. <https://doi.org/10.1371/journal.pone.0000376>

DuBrow, S., & Davachi, L. (2014). Temporal Memory Is Shaped by Encoding Stability and Intervening Item Reactivation. *Journal of Neuroscience*, 34(42), 13998–14005.

<https://doi.org/10.1523/JNEUROSCI.2535-14.2014>

Dudai, Y., Karni, A., & Born, J. (2015). The Consolidation and Transformation of Memory. *Neuron*, 88(1), 20–32. <https://doi.org/10.1016/j.neuron.2015.09.004>

Duncan, K. D., & Schlichting, M. L. (2018). Hippocampal representations as a function of time, subregion, and brain state. *Neurobiology of Learning and Memory*, 153(Pt A), 40–56.

<https://doi.org/10.1016/j.nlm.2018.03.006>

Duncan, K., Sadanand, A., & Davachi, L. (2012). Memory's Penumbra: Episodic Memory Decisions Induce Lingering Mnemonic Biases. *Science*, 337(6093), 485–487.

<https://doi.org/10.1126/science.1221936>

Durkin, C., Hartnett, E., Shohamy, D., & Kandel, E. R. (2020). An objective evaluation of the beholder's response to abstract and figurative art based on construal level theory.

Proceedings of the National Academy of Sciences, 117(33), 19809–19815.

<https://doi.org/10.1073/pnas.2001772117>

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2013). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cerebral Cortex (New York, N.Y.: 1991)*, 23(10), 2467–2478. <https://doi.org/10.1093/cercor/bhs244>

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, 49(5), 1322–1331.

<https://doi.org/10.1016/j.neuropsychologia.2011.02.015>

Eichenbaum, H. (2000). A cortical–hippocampal system for declarative memory. *Nature Reviews Neuroscience*, 1(1), 1. <https://doi.org/10.1038/35036213>

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, 23(2), 209–226. [https://doi.org/10.1016/s0896-6273\(00\)80773-4](https://doi.org/10.1016/s0896-6273(00)80773-4)

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The Medial Temporal Lobe and Recognition Memory. *Annual Review of Neuroscience*, 30(1), 123–152.

<https://doi.org/10.1146/annurev.neuro.30.051606.094328>

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, 8(1), 15276.

<https://doi.org/10.1038/ncomms15276>

Ekman, M., Kusch, S., & de Lange, F. P. (2023). Successor-like representation guides the prediction of future events in human visual cortex and hippocampus. *ELife*, 12, e78904.
<https://doi.org/10.7554/eLife.78904>

Elliott Wimmer, G., & Büchel, C. (2019). Learning of distant state predictions by the orbitofrontal cortex in humans. *Nature Communications*, 10(1), 1.
<https://doi.org/10.1038/s41467-019-10597-z>

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, 12(10), 388–396.
<https://doi.org/10.1016/j.tics.2008.07.004>

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, 20(11), 1504–1513.
<https://doi.org/10.1038/nn.4656>

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*. Zenodo. <https://doi.org/10.5281/zenodo.6834519>

- Ezzyat, Y., & Davachi, L. (2014). Similarity Breeds Proximity: Pattern Similarity within and across Contexts Is Related to Later Mnemonic Judgments of Temporal Proximity. *Neuron*, 81(5), 1179–1189. <https://doi.org/10.1016/j.neuron.2014.01.042>
- Fernandez, C., Jiang, J., Wang, S.-F., Choi, H. L., & Wagner, A. D. (2023). Representational integration and differentiation in the human hippocampus following goal-directed navigation. *eLife*, 12, e80281. <https://doi.org/10.7554/eLife.80281>
- Fonov, V., Evans, A., McKinstry, R., Almlí, 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)
- Foster, D. J. (2017). Replay Comes of Age. *Annual Review of Neuroscience*, 40(1), 581–602. <https://doi.org/10.1146/annurev-neuro-072116-031538>
- Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084), 680–683. <https://doi.org/10.1038/nature04587>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Galea, J. M., Albert, N. B., Ditye, T., & Miall, R. C. (2010). Disruption of the Dorsolateral Prefrontal Cortex Facilitates the Consolidation of Procedural Skills. *Journal of Cognitive Neuroscience*, 22(6), 1158–1164. <https://doi.org/10.1162/jocn.2009.21259>
- Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience*, 17(5), 732–737. <https://doi.org/10.1038/nn.3683>

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, 9(6), 641–651.

<https://doi.org/10.1177/1745691614551642>

Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A., & Sederberg, P. B. (2012). The Successor Representation and Temporal Context. *Neural Computation*, 24(6), 1553–1568. https://doi.org/10.1162/NECO_a_00282

Gilboa, A., & Moscovitch, M. (2021). No consolidation without representation: Correspondence between neural and psychological representations in recent and remote memory. *Neuron*, 109(14), 2239–2255. <https://doi.org/10.1016/j.neuron.2021.04.025>

Glasser, M. F., Sotiroopoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80, 105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127>

Gorgolewski, K., Burns, C., Madison, C., Clark, D., Halchenko, Y., Waskom, M., & Ghosh, S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, 5.

<https://www.frontiersin.org/articles/10.3389/fninf.2011.00013>

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>

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal Replay Is Not a Simple Function of Experience. *Neuron*, 65(5), 695–705.

<https://doi.org/10.1016/j.neuron.2010.01.034>

- Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, 104(5), 1726–1731. <https://doi.org/10.1073/pnas.0610561104>
- Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, 11(7), 299–306. <https://doi.org/10.1016/j.tics.2007.05.001>
- Hasselmo, M. E., & Schnell, E. (1994). Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: Computational modeling and brain slice physiology. *Journal of Neuroscience*, 14(6), 3898–3914.
<https://doi.org/10.1523/JNEUROSCI.14-06-03898.1994>
- Hasselmo, M. E., Schnell, E., & Barkai, E. (1995). Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience*, 15(7), 5249–5262. <https://doi.org/10.1523/JNEUROSCI.15-07-05249.1995>
- 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>
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, 303(5664), 1634–1640.
<https://doi.org/10.1126/science.1089506>
- Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A Hierarchy of Temporal Receptive Windows in Human Cortex. *Journal of Neuroscience*, 28(10), 2539–2550.
<https://doi.org/10.1523/JNEUROSCI.5487-07.2008>

Hebscher, M., Wing, E., Ryan, J., & Gilboa, A. (2019). Rapid Cortical Plasticity Supports Long-Term Memory Formation. *Trends in Cognitive Sciences*, 23(12), 989–1002.

<https://doi.org/10.1016/j.tics.2019.09.009>

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, 19(5), 5.

<https://doi.org/10.1038/nn.4284>

Hobson, J. A., & Friston, K. J. (2012). Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress in Neurobiology*, 98(1), 82–98.

<https://doi.org/10.1016/j.pneurobio.2012.05.003>

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., Doyle, W. K., Rubin, N., Heeger, D. J., & Hasson, U. (2012). Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. *Neuron*, 76(2), 423–434.

<https://doi.org/10.1016/j.neuron.2012.08.011>

Janacsek, K., & Nemeth, D. (2012). Predicting the future: From implicit learning to consolidation. *International Journal of Psychophysiology*, 83(2), 213–221.

<https://doi.org/10.1016/j.ijpsycho.2011.11.012>

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>

Johnson, A., & Redish, A. D. (2007). Neural Ensembles in CA3 Transiently Encode Paths Forward of the Animal at a Decision Point. *Journal of Neuroscience*, 27(45), 12176–12189. <https://doi.org/10.1523/JNEUROSCI.3761-07.2007>

- Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, 60(4), 2357–2364. <https://doi.org/10.1016/j.neuroimage.2012.02.055>
- Kesteren, M. T. R. van, 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>
- Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, 461(2), 145–149. <https://doi.org/10.1016/j.neulet.2009.06.030>
- Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, 13(2), e1005350. <https://doi.org/10.1371/journal.pcbi.1005350>
- Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002). Memory and Temporal Experience: The Effects of Episodic Memory Loss on an Amnesic Patient's Ability to Remember the Past and Imagine the Future. *Social Cognition*, 20(5), 353–379. <https://doi.org/10.1521/soco.20.5.353.21125>
- Kóbor, A., Janacsek, K., Takács, Á., & Nemeth, D. (2017). Statistical learning leads to persistent memory: Evidence for one-year consolidation. *Scientific Reports*, 7(1), 1. <https://doi.org/10.1038/s41598-017-00807-3>
- Kok, P., Brouwer, G. J., Gerven, M. A. J. van, & Lange, F. P. de. (2013). Prior Expectations Bias Sensory Representations in Visual Cortex. *Journal of Neuroscience*, 33(41), 16275–16284. <https://doi.org/10.1523/JNEUROSCI.0742-13.2013>

- Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *Journal of Cognitive Neuroscience*, 26(7), 1546–1554.
https://doi.org/10.1162/jocn_a_00562
- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, 75(2), 265–270.
<https://doi.org/10.1016/j.neuron.2012.04.034>
- Kok, P., & Turk-Browne, N. B. (2018). Associative Prediction of Visual Shape in the Hippocampus. *Journal of Neuroscience*, 38(31), 6888–6899.
<https://doi.org/10.1523/JNEUROSCI.0163-18.2018>
- 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>
- Lavenex, P., & Amaral, D. G. (2000). Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus*, 10(4), 420–430. [https://doi.org/10.1002/1098-1063\(2000\)10:4<420::AID-HIPO8>3.0.CO;2-5](https://doi.org/10.1002/1098-1063(2000)10:4<420::AID-HIPO8>3.0.CO;2-5)
- 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>
- Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *Journal of Neuroscience*, 31(8), 2906–2915. <https://doi.org/10.1523/JNEUROSCI.3684-10.2011>
- Leutgeb, J. K., Leutgeb, S., Moser, M.-B., & Moser, E. I. (2007). Pattern Separation in the Dentate Gyrus and CA3 of the Hippocampus. *Science*, 315(5814), 961–966.
<https://doi.org/10.1126/science.1135801>

Lifanov, J., Linde-Domingo, J., & Wimber, M. (2021). Feature-specific reaction times reveal a

semanticisation of memories over time and with repeated remembering. *Nature Communications*, 12(1), 1. <https://doi.org/10.1038/s41467-021-23288-5>

Lisman, J., & Redish, A. d. (2009). Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1193–1201.

<https://doi.org/10.1098/rstb.2008.0316>

Liu, H., Forest, T. A., Duncan, K., & Finn, A. S. (2023). What sticks after statistical learning:

The persistence of implicit versus explicit memory traces. *Cognition*, 236, 105439.

<https://doi.org/10.1016/j.cognition.2023.105439>

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. J. (2019). Human Replay

Spontaneously Reorganizes Experience. *Cell*, 178(3), 640-652.e14.

<https://doi.org/10.1016/j.cell.2019.06.012>

Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D., & Dolan, R. J. (2021). Experience replay is

associated with efficient nonlocal learning. *Science*, 372(6544), eabf1357.

<https://doi.org/10.1126/science.abf1357>

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal Mismatch Signals Are Modulated by

the Strength of Neural Predictions and Their Similarity to Outcomes. *Journal of Neuroscience*, 36(50), 12677–12687. <https://doi.org/10.1523/JNEUROSCI.1850-16.2016>

Lutz, N. D., Wolf, I., Hübner, S., Born, J., & Rauss, K. (2018). Sleep Strengthens Predictive

Sequence Coding. *Journal of Neuroscience*, 38(42), 8989–9000.

<https://doi.org/10.1523/JNEUROSCI.1352-18.2018>

Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory

patterns in temporal lobe reveal context reinstatement during memory search.

Proceedings of the National Academy of Sciences, 108(31), 12893–12897.

<https://doi.org/10.1073/pnas.1015174108>

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside Looking In: Landmark

Generalization in the Human Navigational System. *Journal of Neuroscience*, 35(44),

14896–14908. <https://doi.org/10.1523/JNEUROSCI.2270-15.2015>

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,

419–457. <https://doi.org/10.1037/0033-295X.102.3.419>

Michelmann, S., Hasson, U., & Norman, K. A. (2023). Evidence That Event Boundaries Are

Access Points for Memory Retrieval. *Psychological Science*, 34(3), 326–344.

<https://doi.org/10.1177/09567976221128206>

Milivojevic, B., Vicente-Grabovetsky, A., & Doeller, C. F. (2015). Insight Reconfigures

Hippocampal-Prefrontal Memories. *Current Biology*, 25(7), 821–830.

<https://doi.org/10.1016/j.cub.2015.01.033>

Molitor, R. J., Sherrill, K. R., Morton, N. W., Miller, A. A., & Preston, A. R. (2021). Memory

Reactivation during Learning Simultaneously Promotes Dentate Gyrus/CA2,3 Pattern

Differentiation and CA1 Memory Integration. *Journal of Neuroscience*, 41(4), 726–738.

<https://doi.org/10.1523/JNEUROSCI.0394-20.2020>

Momennejad, I. (2020). Learning Structures: Predictive Representations, Replay, and

Generalization. *Current Opinion in Behavioral Sciences*, 32, 155–166.

<https://doi.org/10.1016/j.cobeha.2020.02.017>

- Momennejad, I., & Howard, M. W. (2018). *Predicting the Future with Multi-scale Successor Representations* (p. 449470). bioRxiv. <https://doi.org/10.1101/449470>
- Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *eLife*, 7, e32548.
<https://doi.org/10.7554/eLife.32548>
- Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, 1(9), 9. <https://doi.org/10.1038/s41562-017-0180-8>
- Morton, N. W., Sherrill, K. R., & Preston, A. R. (2017). Memory integration constructs maps of space, time, and concepts. *Current Opinion in Behavioral Sciences*, 17, 161–168.
<https://doi.org/10.1016/j.cobeha.2017.08.007>
- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–175.
[https://doi.org/10.1016/0006-8993\(71\)90358-1](https://doi.org/10.1016/0006-8993(71)90358-1)
- O'Keefe, J., & Nadel, L. (1979). Précis of O'Keefe & Nadel's The hippocampus as a cognitive map. *Behavioral and Brain Sciences*, 2(4), 487–494.
<https://doi.org/10.1017/S0140525X00063949>
- Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D., & Spiers, H. J. (2015). Hippocampal place cells construct reward related sequences through unexplored space. *eLife*, 4, e06063. <https://doi.org/10.7554/eLife.06063>
- Ólafsdóttir, H. F., Bush, D., & Barry, C. (2018). The Role of Hippocampal Replay in Memory and Planning. *Current Biology*, 28(1), R37–R50.
<https://doi.org/10.1016/j.cub.2017.10.073>

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, 4(6), 661–682.

<https://doi.org/10.1002/hipo.450040605>

Ouden, H. E. M. den, Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal Prediction Error Modulates Cortical Coupling. *Journal of Neuroscience*, 30(9), 3210–3219. <https://doi.org/10.1523/JNEUROSCI.4458-09.2010>

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Crespo Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, 414(6863), 546–550. <https://doi.org/10.1038/35107080>

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, 17(5), 230–240.

<https://doi.org/10.1016/j.tics.2013.03.005>

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>

Rauss, K., & Born, J. (2017). A Role of Sleep in Forming Predictive Codes. In N. Axmacher & B. Rasch (Eds.), *Cognitive Neuroscience of Memory Consolidation* (pp. 117–132). Springer International Publishing. https://doi.org/10.1007/978-3-319-45066-7_8

- 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>
- Romano, J. C., Howard, J. H., & Howard, D. V. (2010). One-year retention of general and sequence-specific skills in a probabilistic, serial reaction time task. *Memory (Hove, England)*, 18(4), 427–441. <https://doi.org/10.1080/09658211003742680>
- Sanchez, D. J., Gobel, E. W., & Reber, P. J. (2010). Performing the unexplainable: Implicit task performance reveals individually reliable sequence learning without explicit knowledge. *Psychonomic Bulletin & Review*, 17(6), 790–796. <https://doi.org/10.3758/PBR.17.6.790>
- 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>
- Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 773–786.
<https://doi.org/10.1098/rstb.2007.2087>
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, 8(9), 657–661.
<https://doi.org/10.1038/nrn2213>

- Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron*, 76(4), 677–694. <https://doi.org/10.1016/j.neuron.2012.11.001>
- 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., McDevitt, E. A., Chen, L., Norman, K. A., Mednick, S. C., & Rogers, T. T. (2017). Sleep Benefits Memory for Semantic Category Structure While Preserving Exemplar-Specific Information. *Scientific Reports*, 7(1), 1. <https://doi.org/10.1038/s41598-017-12884-5>
- Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C., & Norman, K. A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications*, 9(1), 3920. <https://doi.org/10.1038/s41467-018-06213-1>
- Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, 16(4), 486–492. <https://doi.org/10.1038/nn.3331>
- 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>

- Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, 26(1), 3–8.
<https://doi.org/10.1002/hipo.22523>
- Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, 6(1), 8151.
<https://doi.org/10.1038/ncomms9151>
- Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, 1, 1–8.
<https://doi.org/10.1016/j.cobeha.2014.07.005>
- Schlichting, M. L., & Preston, A. R. (2016). Hippocampal–medial prefrontal circuit supports memory updating during learning and post-encoding rest. *Neurobiology of Learning and Memory*, 134, 91–106. <https://doi.org/10.1016/j.nlm.2015.11.005>
- Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, 91(6), 1402–1412.
<https://doi.org/10.1016/j.neuron.2016.08.019>
- Scoville, W. B., & Milner, B. (1957). LOSS OF RECENT MEMORY AFTER BILATERAL HIPPOCAMPAL LESIONS. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20(1), 11–21.
- Sekeres, M. J., Bonasia, K., St-Laurent, M., Pishdadian, S., Winocur, G., Grady, C., & Moscovitch, M. (2016). Recovering and preventing loss of detailed memory: Differential rates of forgetting for detail types in episodic memory. *Learning & Memory*, 23(2), 72–82. <https://doi.org/10.1101/lm.039057.115>

- Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters*, 680, 39–53.
<https://doi.org/10.1016/j.neulet.2018.05.006>
- Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences*, 14(10), 464–472. <https://doi.org/10.1016/j.tics.2010.08.002>
- Shohamy, D., & Wagner, A. D. (2008). Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron*, 60(2), 378–389.
<https://doi.org/10.1016/j.neuron.2008.09.023>
- Skaggs, W. E., & McNaughton, B. L. (1996). Replay of Neuronal Firing Sequences in Rat Hippocampus During Sleep Following Spatial Experience. *Science*, 271(5257), 1870–1873. <https://doi.org/10.1126/science.271.5257.1870>
- Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1), 83–98. <https://doi.org/10.1016/j.neuroimage.2008.03.061>
- Spiers, H. J., & Gilbert, S. J. (2015). Solving the detour problem in navigation: A model of prefrontal and hippocampal interactions. *Frontiers in Human Neuroscience*, 9.
<https://www.frontiersin.org/articles/10.3389/fnhum.2015.00125>
- Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, 20(11), 11. <https://doi.org/10.1038/nn.4650>
- Steel, A., Billings, M. M., Silson, E. H., & Robertson, C. E. (2021). A network linking scene perception and spatial memory systems in posterior cerebral cortex. *Nature Communications*, 12(1), 2632. <https://doi.org/10.1038/s41467-021-22848-z>

- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437(7063), 7063.
<https://doi.org/10.1038/nature04286>
- Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience*, 15(10), 655–669.
<https://doi.org/10.1038/nrn3785>
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Tambini, A., & Davachi, L. (2013). Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. *Proceedings of the National Academy of Sciences*, 110(48), 19591–19596. <https://doi.org/10.1073/pnas.1308499110>
- Tambini, A., Ketz, N., & Davachi, L. (2010). Enhanced Brain Correlations during Rest Are Related to Memory for Recent Experiences. *Neuron*, 65(2), 280–290.
<https://doi.org/10.1016/j.neuron.2010.01.001>
- Tarder-Stoll, H., Baldassano, C., & Aly, M. (2023). *Consolidation enhances multistep anticipatory judgements but diminishes access to perceptual features*. PsyArXiv.
<https://doi.org/10.31234/osf.io/x2f7s>
- Tarder-Stoll, H., Jayakumar, M., Dimsdale-Zucker, H. R., Günseli, E., & Aly, M. (2020). Dynamic internal states shape memory retrieval. *Neuropsychologia*, 138, 107328.
<https://doi.org/10.1016/j.neuropsychologia.2019.107328>
- Tarhan, L., & Konkle, T. (2020). Reliability-based voxel selection. *NeuroImage*, 207, 116350.
<https://doi.org/10.1016/j.neuroimage.2019.116350>
- Thavabalasingam, S., O’Neil, E. B., Tay, J., Nestor, A., & Lee, A. C. H. (2019). Evidence for the incorporation of temporal duration information in human hippocampal long-term

memory sequence representations. *Proceedings of the National Academy of Sciences*, 116(13), 6407–6414. <https://doi.org/10.1073/pnas.1819993116>

Thorp, J. N., Gasser, C., Blessing, E., & Davachi, L. (2022). Data-Driven Clustering of Functional Signals Reveals Gradients in Processing Both within the Anterior Hippocampus and across Its Long Axis. *Journal of Neuroscience*, 42(39), 7431–7441. <https://doi.org/10.1523/JNEUROSCI.0269-22.2022>

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55, 189–208. <https://doi.org/10.1037/h0061626>

Tolman, E. C., & Gleitman, H. (1949). Studies in learning and motivation: I. Equal reinforcements in both end-boxes, followed by shock in one end-box. *Journal of Experimental Psychology*, 39(6), 810–819. <https://doi.org/10.1037/h0062845>

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. I. Orientation and the short-cut. *Journal of Experimental Psychology*, 36(1), 13–24. <https://doi.org/10.1037/h0053944>

Tompry, A., & Davachi, L. (2017). Consolidation Promotes the Emergence of Representational Overlap in the Hippocampus and Medial Prefrontal Cortex. *Neuron*, 96(1), 228-241.e5. <https://doi.org/10.1016/j.neuron.2017.09.005>

Tompry, A., & Davachi, L. (2022). *Consolidation-dependent behavioral integration of sequences related to mPFC neural overlap and hippocampal-cortical connectivity*.

Tompry, A., Duncan, K., & Davachi, L. (2015). Consolidation of Associative and Item Memory Is Related to Post-Encoding Functional Connectivity between the Ventral Tegmental Area and Different Medial Temporal Lobe Subregions during an Unrelated Task. *Journal of Neuroscience*, 35(19), 7326–7331. <https://doi.org/10.1523/JNEUROSCI.4816-14.2015>

- Townsend, J. T., & Ashby, F. G. (1978). Methods of Modeling Capacity in Simple Processing Systems. In *Cognitive Theory*. Psychology Press.
- Treichler, F. R., & Van Tilburg, D. (1996). Concurrent conditional discrimination tests of transitive inference by macaque monkeys: List linking. *Journal of Experimental Psychology: Animal Behavior Processes*, 22(1), 105–117. <https://doi.org/10.1037/0097-7403.22.1.105>
- Treves, A., & Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, 2(2), 189–199. <https://doi.org/10.1002/hipo.450020209>
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440–463. <https://doi.org/10.1037/a0018963>
- Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., & Morris, R. G. M. (2011). Schema-Dependent Gene Activation and Memory Encoding in Neocortex. *Science*, 333(6044), 891–895. <https://doi.org/10.1126/science.1205274>
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology / Psychologie Canadienne*, 26(1), 1–12. <https://doi.org/10.1037/h0080017>
- Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, 53(1), 1–25. <https://doi.org/10.1146/annurev.psych.53.100901.135114>
- Tulving, E. (2005). Episodic Memory and Autonoesis: Uniquely Human? In *The missing link in cognition: Origins of self-reflective consciousness* (pp. 3–56). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195161564.003.0001>

- Turk-Browne, N. B., Simon, M. G., & Sederberg, P. B. (2012). Scene Representations in Parahippocampal Cortex Depend on Temporal Context. *Journal of Neuroscience*, 32(21), 7202–7207. <https://doi.org/10.1523/JNEUROSCI.0942-12.2012>
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., 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>
- van Kesteren, M. T. R., Fernández, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences*, 107(16), 7550–7555. <https://doi.org/10.1073/pnas.0914892107>
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with Sleep Makes Perfect: Sleep-Dependent Motor Skill Learning. *Neuron*, 35(1), 205–211. [https://doi.org/10.1016/S0896-6273\(02\)00746-8](https://doi.org/10.1016/S0896-6273(02)00746-8)
- Walker, M. P., Brakefield, T., Seidman, J., Morgan, A., Hobson, J. A., & Stickgold, R. (2003). Sleep and the Time Course of Motor Skill Learning. *Learning & Memory*, 10(4), 275–284. <https://doi.org/10.1101/lm.58503>
- Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex*, 25(10), 3911–3931. <https://doi.org/10.1093/cercor/bhu277>
- Wikenheiser, A. M., & Redish, A. D. (2015). Hippocampal theta sequences reflect current goals. *Nature Neuroscience*, 18(2), 289–294. <https://doi.org/10.1038/nn.3909>

- Wimmer, G. E., Liu, Y., McNamee, D. C., & Dolan, R. J. (2023). Distinct replay signatures for prospective decision-making and memory preservation. *Proceedings of the National Academy of Sciences*, 120(6), e2205211120. <https://doi.org/10.1073/pnas.2205211120>
- Wimmer, G. E., Liu, Y., Vehar, N., Behrens, T. E. J., & Dolan, R. J. (2020). Episodic memory retrieval success is associated with rapid replay of episode content. *Nature Neuroscience*, 23(8), 1025–1033. <https://doi.org/10.1038/s41593-020-0649-z>
- Wimmer, G. E., & Shohamy, D. (2012). Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. *Science*, 338(6104), 270–273. <https://doi.org/10.1126/science.1223252>
- Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia*, 48(8), 2339–2356. <https://doi.org/10.1016/j.neuropsychologia.2010.04.016>
- Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492. <https://doi.org/10.1109/CVPR.2010.5539970>
- Yu, W., Zadbood, A., Chanales, A. J. H., & Davachi, L. (2022). Repetition accelerates neural markers of memory consolidation (p. 2022.12.14.520481). bioRxiv. <https://doi.org/10.1101/2022.12.14.520481>
- Zeithamova, D., & Bowman, C. R. (2020). Generalization and the hippocampus: More than one story? *Neurobiology of Learning and Memory*, 175, 107317. <https://doi.org/10.1016/j.nlm.2020.107317>

- Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron*, 75(1), 168–179. <https://doi.org/10.1016/j.neuron.2012.05.010>
- 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>

Appendix A: Supplementary Materials for Chapter 1

Story examples

“I went to a **restaurant** that served delicious food. The restaurant was located in a **big city**.

Since I was already in a restaurant in a big city I decided to go back to the **kitchen** to ask the chef a question. The chef in the kitchen told me that he had gone to school in a big **auditorium** to learn his craft. The chef's school auditorium was right next to a **beach** they would play at.

The beach's water ran through to a **forested mountain**. At the forested mountain there was a **ski resort** and I went skiing. The ski resort in the mountain was so high up that you needed to take a **plane** to get there. At the top of the mountain I had to take a plane to get to the top of was an old **castle** that I wanted to stay the night in. Since I'm staying the night in the castle I needed to sleep in a **bed**. The bed in the castle was right next door to a **restaurant** that was lovely to eat in.”

“In the middle of the **forest** was a beautiful kingdom with a king who lived in a **castle**. The castle was atop a hill in the middle of a **city**. The king's city was bordered by a beautiful **beach**. The beach that was next to the king's city was so beautiful that people came on **planes** to see it. Part of the reasons tourists came on planes was to eat at the wonderful **restaurants**. The restaurants were staffed by the kingdom's most promising students. The students of the restaurant **school** are also all **skiing experts**. The restaurant school is staffed by skiing experts, it is also a boarding school with very comfortable **beds**. The boarding school has a **kitchen** where the restaurant chefs can learn their craft. The kitchen is one of the things that makes the **forest** kingdom great.”

“I was walking through my **kitchen** when I looked outside the window and saw a **forest**. While looking at the forest I got a bit peckish and decided that I would go to a **restaurant**. While thinking about what restaurant I should go to I thought about the food on **planes** and just planes in general. While thinking about planes and the horrid food that they have I remembered my school and the food that they have, especially the food in the **auditorium**. The auditorium was right next to a **beach**. And the beach was right next to the **castle**, The castle was across a **snowy mountain** with people skiing. After watching the people ski on the mountain I decided to look at the **city skyline**. After looking at the city skyline, I grew tired and decided to go back home to my **bedroom**. In order to get to my bedroom I had to pass through my **kitchen**.”

Appendix B: Supplementary Materials for Chapter 2

Story Examples

“I was relaxing in the meadows one day when I came upon an interesting hallway; I decide to walk through. When I go through the last door of the hallway, I land on this futuristic platform that tells me to pick a letter on the screen. I pick the third letter, and get transported into a huge, dark, scary cave. I frantically run to the end of the cave, and it leads me to a beautiful sunny park. At the park, I am asked to pick between the yellow gazebo and the red gazebo. I pick the yellow one, and I find myself in a tunnel. Once I go through the archway of this tunnel, I realize that it’s a portal. I zoom through the portal, and end up in my house on Christmas day of 2005. I realize that I’ve time travelled. After enjoying a meal with my family, I enter the fireplace and am transported back to the meadow where I began.”

“I am being held captive by an intergalactic military convoy. I am being taken to a prison on a remote forest world that is covered in darkness every day. I am dragged along a dark path by the brutish guards, and after a night of travel they leave me in this cell. What it presents as upscale boutique aesthetics festers into cold minimalism and rots the brain of the inhabitant. But I escape. A cleverly timed throwing-a-chair-through-the-window leaves me stranded in the barren hills that make up this planet. I find myself staggering across each hill, thinking each vista might hold sight of my salvation. One of them does, in the form of an old pub with a blinking neon sign and a game of dominoes going on outside. I order a drink. A man, dressed in a monocle and festive top hat turns to me, asks me my story. I tell it. He seems amused. He invites me to his summer palace on a brighter part of the planet, that is only reachable through hot air balloon.

It is warm. I can feel the suns tapping warmly on my cheek. The eccentric man puts his hands on my shoulders. You will love it there, he says. When we arrive, I can see why he might have thought so. But why come all this way to take a mere stranger back with you. It has been days. It is beautiful on his estate. He never takes off his festive tophat, not even over eggs in the morning or with a nightcap over the fire. After a week he invites me into his study, in the wing he had banned me from entering. Upon seeing the walls, it washes over me in an instant. This is the leader of the Christmas empire, he knew of my escape and took me captive. He laughs. Guards enter the room and take me aboard the ship that will take me to prison. This is the last time I let my guard down for a festive top hat."