From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Psychology

2018

A Case of Memory Reconstruction: Sharing Memories via Communication

Zadbood, Asieh

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01rv042w95x

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

A Case of Memory Reconstruction: Sharing Memories via Communication

Asieh Zadbood

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPRTMENT OF PSYCHOLOGY

Advisor: Uri Hasson

Secondary Advisor: Kenneth A. Norman

June 2019

© Copyright by Asieh Zadbood, 2019. All rights reserved.

Abstract

The brain actively interprets and re-interprets our personal narrative. Incoming information is reshaped based on prior experience during the encoding, and this update process proceeds through consolidation, retrieval, and reconsolidation. But how does the brain implement this constructive process? In this dissertation I studied reconstructive memory from two perspectives. First, I asked whether memory and construction have shared neural mechanisms. To address this question, I studied how memories of past events (in a movie) are transmitted to other individuals, who mentally construct the events through listening to an audio recording of a participant's recall in fMRI scanner. I observed reliable similarities between the scene specific patterns of brain response during encoding, recall, and construction of the same real-life episodes in the DMN. These mutual patterns of representation, suggest that a mutual mechanism to perform different construction-related functions might be at work. In the second study, I used fMRI to track the neural representations of the memory update during naturalistic recall. Participants watched a movie and at the end of the movie were exposed to a dramatic "twist" that changed the interpretation of the previously encoded movie. Participants were then instructed to recall several scenes in the movie in their own words (in the scanner). Most participants updated their recall to incorporate the twist. Two control groups performed similar tasks: one group watched a "spoiled" version of the movie (knowing the twist from the beginning), and a second group never saw the twist. Both control groups performed the spoken recall task in the scanner as well. This enabled us to compare how the twist transformed neural patterns during encoding and during recall. Our results demonstrated that providing participants with information about the twist beforehand (spoiled condition) altered the neural response patterns during the encoding phase (movie-viewing). Moreover, providing participants with information about the twist after the end of the encoding phase (post-viewing) transformed the neural patterns of the encoded information during recall. The similarity between

activation patterns during encoding and recall was attenuated in participants who better updated their recall to accommodate the new information. This suggests that event representations are updated to retroactively accommodate new narrative information. Overall, the work presented in this dissertation provides a window into the neural mechanisms underlying social and constructive memory processes.

Acknowledgments

My time at Princeton was enriched due to the help and support of many people. I will not be able to name them all here, and even if I do, I cannot express my gratitude properly. First, I would like to thank my advisor, Uri Hasson, who was my main supporter. Without his support and guidance, none of the things I did during my PhD would be possible. I would also thank my secondary advisor Ken Norman, that made my path at Princeton more fruitful by his mentoring and support. I am grateful to Janice Chen, who taught me a lot like a mentor and helped me a lot as a friend. I thank my oral committee, Adele Goldberg and Christopher Honey, for their great feedback. I appreciate the help and support that I received from my friends and colleagues in the lab, Liat Hasenfratz, Mor Regev, Mai Nguyen, Sam Nastase, Yuan Chang Leong, and other members of Hasson and Norman labs. At the end, I would like to especially thank my friends and family. I will not be able to show my feelings and gratitude to them by words.

To my parents

My life’s light and source of inspiration

Table of Content

Chapter 1 Introduction ……..…...…..1

1.1 Construction as the basis of Memory……..……..2

1.2 To study memory as an adaptive process in an ever-changing world……...……....7

Chapter 2 How we transmit memories to other brains: constructing shared neural representations via communication

2.1 Introduction…...…….……..11

2.2 Materials and Methods ……..……15

Stimuli…….……..…….……15

Participants…….……..…….…15

Procedure…….…….…….……..……...16

Post-scan behavioral memory test ...…….…….……20

Data analysis…….…….…….……20

2.3 Results…….……..……..27

Pattern similarity between spoken-recall and movie-viewing……..……..27

Pattern similarity between spoken-recall and listening……..……..….……..30

Pattern similarity between listening and movie-viewing…….…..……..31

Relationship between pattern similarity and behavioral performance …...…….…..……..35

Shared neural response across three conditions (triple shared pattern analysis)…...……37

2.4 Discussion…….…….…….39

Chapter 3 Memory update during naturalistic recall

3.1 Introduction..…….……..…..…….……..45

3.2 Material and Methods…...……..…….…….…….49

Stimuli ……...…….…….…….50

Participants…..…..……...…….……..51

Procedure…..…..…….…….……..52

Data Analysis…..…..……...…….……..54

3.3 Results…..…..……..…….……..61

Recall behavior…..…..……...…….……...61

Effect of the twist on movie viewing…….…….……...64

Effect of the twist on recall representations……...…….……...68

Interaction between movie encoding and movie recall……..…….……...70

Effect of the twist on memory reinstatement ……....…….……...73

3.4 Discussion……..…….……...78

4. Concluding remarks……...…….……...84

5. References …….…….……...86

Chapter 1: Introduction

Memories are often seen as intact entities; as information stored in the past that will be retrieved at the time of *remembering* or will not be recalled if *forgetting* occurs. However, this dichotomy is challenged in the case of 'memory distortions'- when the information retrieved is incorrect. Distorted memories share common features with real memories, and people who recall these distorted memories believe, sometimes with high confidence, that their false memories are veridical accounts (Gallo, 2010; Kim & Cabeza, 2007; Loftus, 2005). Childhood memories such as being lost in a mall or being hospitalized overnight can be entirely and unintentionally fabricated or even planted in one's mind (Loftus, 1997, 2005; Loftus & Pickrell, 1995). The formation of such illusions can have serious consequences and has been of particular concern regarding the evaluation of eyewitness testimonies in the courts (Lacy & Stark, 2013), or the psychotherapy in cases of childhood sexual abuse (Lindsay & Read, 1995; D. L. Schacter, Koutstaal, & Norman, 1996). Furthermore, confabulations can have debilitating effects on the life of brain damaged, Alzheimer's patients, and a number of other neurological disorders (Burgess & Shallice, 1996; Kern, Van Gorp, Cummings, Brown, & Osato, 1992; Daniel L. Schacter, 2012; Schnider, 2017).

While memory distortions may appear as instances of failure in an unreliable system, or "sins of memory" (D.L. Schacter, 1999), they can be seen from a broader adaptive perspective. Frederic Bartlett was a pioneer in bringing the concept of *memory reconstruction* into the light (F. C. Bartlett, 1932). He emphasized the dynamic, constructive aspect of memory which defines memories *not* as 'carbon copies' of events in the past, but rather as dynamic recollections of past experience that incorporate new information to achieve adaptive goals (Bartlett 1932). According to him, a literal reproduction is useless in a world that constantly changes. He viewed humans as

agents who *actively* react to stimuli and are thus able to adapt their responses according to their current needs (Bartlett 1932, 1933; S. F. C. Bartlett & Bartlett, 1995; Wagoner, 2017). Under this framework, memories are dynamically changing to meet real life needs. This very same property, however, also makes them vulnerable to unwanted distortions.

The constructive notion of memory is widely acknowledged today and memory distortions are accepted as "by-products of an otherwise adaptive system" (Conway & Pleydell-Pearce, 2000; Hassabis & Maguire, 2007; D. L. Schacter, Norman, & Koutstaal, 1998; Daniel L. Schacter, 2012). With this perspective, studying memory distortions does not necessarily require direct bifurcation of true and false memories under different task conditions, as currently common in many paradigms of studying memory distortion (Dennis, Bowman, & Turney, 2015). Therefore, studying the constructive aspect of memory in normative real-life situations may shed light on the mechanisms of memory distortions as well. In addition, this approach could help us to better understand how memory system interacts with other cognitive functions to meet adaptive goals.

In this dissertation I adopted two related approaches to studying reconstructive memory. First, I studied the relationship between retrieval of real-life memories and the construction of new imagined events. Second, I studied the process by which we update and modify our memories as we gather more information about the structure of events over time. These two directions will be further elaborated at the end of this introduction (section 1.2).

1.1 Construction as the basis of Memory

Dynamicity of memory can be discussed at all three levels of encoding, storage or consolidation, and retrieval. A critical concept in implementing Bartlett's notion of reconstructive memory is

schema. Bartlett specifically emphasized the adaptivity of schemas and envisioned them as active representations rather than static storage-like organizations (S. F. C. Bartlett & Bartlett, 1995; Wagoner, 2017). Evolution of schemas through life experiences (learning) was later elaborated by Piaget who defined *assimilation* as integration of new information in current schemas, and *accommodation* in which schemas are modified to incorporate new knowledge (Piaget, 1952). In modern psychology and neuroscience, schemas are broadly defined as abstract knowledge structures formed through various life experiences that affect perception and interpretation of current events and their future retrieval (Gilboa & Marlatte, 2017).

At the encoding level, where memory representations are created under the influence of relevant prior knowledge or schemas, the passive recording of sensory input gives way to *construction* (Alba & Hasher, 1983; Bonasia et al., 2018; McKenzie et al., 2014; Poppenk, McIntosh, Craik, & Moscovitch, 2010; Tse et al., 2011; van Kesteren, Ruiter, Fernández, & Henson, 2012). Representations that are congruent with prior knowledge, often improve comprehension and future recollection (Alba & Hasher, 1983; Sommer, 2017; van Kesteren et al., 2012). However, congruence with schemas can also distort memories by forming prototypical representations that match well with prior experiences but their details have been removed and thus can lead to biases in perception and memory (Friedman, 1979; Ghosh & Gilboa, 2014; Hemmer & Steyvers, 2009; Konkle & Oliva, 2007).

The medial prefrontal cortex has been established as a key region to process incoming information in relation to an existing schema (Bonasia et al., 2018; Kesteren, Rijpkema, Ruiter, & Fernández, 2010; Tse et al., 2011; van Kesteren et al., 2012). It has been shown that during assimilation of new information into the established prior knowledge framework, immediate early

gene expression in rats' mPFC is upregulated (Tse et al., 2011). Patients with mPFC damage are unable to make connections between words used in everyday life and their associated schemas (Ghosh, Moscovitch, Colella, & Gilboa, 2014). However, if they are asked to perform the Deese–Roediger–McDermott (DRM) (Roediger & McDermott, 1995) paradigm, which is widely used to study false memories, their performance is better than that of healthy controls (Melo, Winocur, & Moscovitch, 1999; Warren, Jones, Duff, & Tranel, 2014). In this paradigm a list of semantically related words is memorized. During the test, in which subjects must determine if a word is "new" (unseen) or "old" (on the prev. mentioned list), semantically related lures that were not present in the actual set, are often falsely reported as "old" with high confidence. As the associative connection of the lures to the study items increases, so too does the chance of a false recall (Roediger, Watson, McDermott, & Gallo, 2001). In this paradigm, semantic associations between items on the list are created to improve memory performance. At the same time, these associations form gist-like representations, that could lead to incorrect detection of lures as old items due to reduced specificity. The absence of this mechanism, caused by mPFC damage (as stated above) or temporal deactivation using TMS (Berkers et al., 2017), protects participants from making memory errors. However, this memory benefit comes at the cost of the ability to evaluate and integrate incoming information in relation to prior knowledge, which is an important adaptive function.

Formation of long-term stable memories occurs through the process of systems consolidation (Dudai, Karni, & Born, 2015; McGaugh, 2000), and is also affected by prior knowledge frameworks or schemas. In a pioneering study in 2007, Tse and colleagues showed that memories formed within an existing framework of knowledge were immune to hippocampal damage occurring 24 hours later. This was surprising given that under normal circumstances it takes much longer than 24 hours to consolidate memories rendering them hippocampal

independent. Therefore, this finding suggests that new memory consolidation is faster if memories are congruent with the previously learned association of a complex environment or schema (Tse et al., 2007). It is not clear whether the reorganization of memory during systems consolidation modifies the memory content (Hardt & Nadel, 2018).

The act of remembering may lead to memory modification as well. At the molecular level, it has been shown that retrieval makes memories susceptible to change. The glutamate receptor subtype NMDA (N-methyl-D-aspartate) is important to forming of long-term memories. When memories are consolidated, they become presumably NMDA independent and therefore an injection of an NMDA-receptor antagonist should not affect them. However, while this antagonist was administered to rats right after retrieval of consolidated memories, their subsequent retrieval was impaired (Przybyslawski & Sara, 1997). According to Przybyslawski and Sara, the memories were *reconsolidated*, which was confirmed by later findings (K. Nader, Schafe, & Le Doux, 2000). Reconsolidation fits well with Bartlett's notion of reconstructive memory. Memories, regardless of their age, may be vulnerable to change when reactivated during retrieval and become exposed to new information (Besnard, Caboche, & Laroche, 2012; Hupbach, Gomez, & Nadel, 2015; Karim Nader & Einarsson, 2010). The factors that lead to attenuation or updating memories (integration of new information into the encoded memories) is not yet entirely clear. One hypothesis suggests that compatibility of the new information with encoded memories could facilitate integration, while incompatibility may cause attenuation (Hupbach, 2011).

Another line of evidence supporting the constructive nature of memory is through other cognitive functions, which appear to be intertwined with memory. For instance, prospection or future thinking (the ability to simulate future events in one's mind) is of great adaptive value.

(Atance & O'Neill, 2001; Gilbert & Wilson, 2007; Schacter, Addis, & Buckner, 2008; Schacter, Benoit, & Szpunar, 2017; Szpunar, Spreng, & Schacter, 2014; Szpunar, Watson, & McDermott, 2007). Predicting the consequences of hypothetical actions and events leads to improved planning and decision making, which benefits the action outcomes. Imagined future events rely on varying degrees of episodic and semantic memory information (Szpunar et al., 2014). This reliance depends on the nature of the imagined event, whether it is heavily episodic like a future family gathering, or semantic like the future of the world under climate change. It has been suggested that retrieving the past experience and reframing it into the representation of a potential future event better enables future thinking (Schacter & Addis, 2007). Accordingly, the network involved in simulation of future episodes overlaps with episodic memory encoding, the retrieval network (Maguire, 2001; Svoboda, McKinnon, & Levine, 2006), and the semantic network (Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009).

The network of brain regions involved in episodic encoding and retrieval, imagining future events, scene construction (imagining hypothetical scenes), and theory of mind, greatly overlaps with the DMN (default mode network) (Hassabis, Kumaran, & Maguire, 2007; Daniel L Schacter & Addis, 2007, 2007; Spreng, Mar, & Kim, 2009; Szpunar et al., 2007) . The DMN was initially defined as internalizing network or a set of regions that is sensitive to inner thoughts rather than external stimuli. However, it is now well established that this network is involved in multiple higher-level cognitive function (R. Buckner, Andrews-Hanna, & Schacter, 2008; Raichle et al., 2001). The same set of regions have been indicated in accumulation of information over long time windows (TRW) during processing of naturalistic stimuli such as movies or audio narratives (Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Lerner, Honey, Silbert, & Hasson, 2011). Recently it was shown that during processing of different schemas, regions on this network contain

representations that, regardless of the story, are shared within schemas rather than between schemas (Baldassano, Hasson, & Norman, 2018). Construction appears to be a key mutual property of many cognitive processes related to DMN (Hassabis & Maguire, 2009), which highlights the importance of a more holistic approach to the study of the brain, and to an understanding of the neural underpinnings of memory reconstruction.

1.2 To study memory as an adaptive process in an ever-changing world

"If this view of remembering as a constructive activity is correct, the whole experimental setting of the problem of recall is changed."(Bartlett 1932)

In studying memory, Bartlett believed in a functionalist approach. That is to say, how memory is utilized in real life and how it adapts to the task goals. He criticized tasks that defined certain dependent and independent variables but ignored the active role of the participant in influencing the results (Bartlett 1932). Therefore, he sought to design more ecologically valid tasks to study memory retrieval. He valued qualitative results, such as the particular ways in which participants perform the recall, over quantitative results, such as the number of correct items recalled. In his famous series of experiments called "The War of the Ghosts" (Bartlett 1932), participants read a short story that was quite confusing and lacked a rational plot line but was filled with characters, dialogues, and vivid descriptions. He later asked participants to write down everything that they remembered from the story and recorded how participants were dealing with this type of narrative, given that it was incongruent with their normal logical comprehension template. Bartlett observed that based on their impression from the story, participants omitted or altered parts of the narrative

to better fit it into a more logical understanding of the story (Bartlett 1932, S. F. C. Bartlett & Bartlett, 1995; Wagoner, 2017)

According to Bartlett, another important dimension of memory adaptivity is its function in social life (Bartlett 1932, S. F. C. Bartlett & Bartlett, 1995). In a follow-up experiment, he asked participants naïve to the story to read a written recall and later write their own recall of the story. Then the next participant would read this recall and write their own recall of it and so on (similar to a game of Telephone). His goal in this experiment was to observe how flexibility in memory could play in the social and cultural aspects of life. He reported that changes in recalls were even more dramatic than the changes when a subject repeatedly recalled the story herself.

Sharing past experiences by talking is very common in real life. We could be recounting a remote childhood memory in a family gathering, or a recent memory of events that occurred during the day. People may not have identical memories of the same event. Even the same person may recall an event differently in two occasions for a different audience. From a reconstructive memory perspective, sharing memories fits well into the definition of a constructive adaptive process. Memories get compressed, abstracted, omitted or perhaps altered during the recall to meet behavioral goals. Therefore, sharing memories could be used as an excellent ecologically valid experimental set up to study memory reconstruction.

While the neural processes underlying free spoken recall might appear difficult to study in experimental settings, in 2017 Chen and colleagues successfully used this paradigm to study memory reinstatement. Participants watched a 50-min long movie and recounted it in the fMRI scanner in their own words. Their recalls ranged from 10 to 43 minutes and contained a wide range

of information and details of each movie event. Memory reinstatement, defined as similar patterns of brain response during the encoding (watching the movie) and subsequent recall of the same events, was observed in DMN regions with fMRI. Interestingly, despite the differences in recall length and content, a systematic transformation of perception into memory was reported, as representations of the recall of the same event between participants were more similar that recall to encoding representations (Chen et al., 2017)

In this dissertation, I have studied the reconstructive nature of memories using naturalistic stimuli (movies and audio narratives) and spoken recall. In **Chapter 2**, I first studied how memories of past events (in a movie) are transmitted to other individuals, who mentally construct the events through listening to an audio recording of a participant's recall. I observed reliable similarities between the scene specific patterns of brain response during encoding, recall, and construction of the same real-life episodes in the DMN. As discussed in section 1.1, prior research suggests that mutual networks of brain regions are implicated during memory and other construction-related cognitive functions (Hassabis & Maguire, 2009). Here I explored the neural processes underlying memory further by evaluating the patterns of brain representations rather than the amplitude of brain response. Indeed, my results support mutual patterns of representation across encoding, retrieval, and construction. This suggests that a mutual mechanism to perform different functions is at work, rather than the operation of different subsets of representations in the same regions. Findings were replicated across two independent sets of participants and stimuli.

In **Chapter 3**, I evaluated the phenomenon of memory update in a natural and adaptive setting. Participants watched a movie and were exposed to a dramatic "twist" that changed the interpretation of the previously encoded movie. Participants were instructed to recall each scene

in the movie in their own words spoken and recorded in an fMRI scanner. One of the main questions in this research study was whether participants could update their recall to convey the most accurate information (in this case, the new interpretation). In this study, the main group watched the movie and learned about the twist only at the end of the movie. I included two control groups, one group that watched a "spoiled" version of the movie (knowing the twist from the beginning), and a second group that never saw the twist. This enabled a comparison of how the twist transformed neural patterns during encoding and during recall. Results showed that providing participants with information about the twist before the encoding phase (spoiled condition), transformed the neural patterns of the encoded information as they watched the movie. Similar results were seen in a previous study when subjects received different contextual information prior to listen to an ambiguously narrated story (Yeshurun et al 2017). Furthermore, the results showed that providing participants with information about the twist after the end of the encoding phase (post viewing), transformed the neural patterns of the encoded information during recall. The similarity between the patterns of brain response during encoding and recall was attenuated in participants who better updated their recall to accommodate the new information. This finding suggests that update during retrieval has changed the event representations and made them dissimilar to the original representation of the previously encoded events.

Chapter 2: How we transmit memories to other brains: constructing shared neural representations via communication

2.1 Introduction

Sharing memories of past experiences with each other is foundational for the construction of our social world. What steps comprise the encoding and sharing of a daily life experience, such as the plot of a movie we just watched, with others? To verbally communicate an episodic memory, the speaker has to recall and transmit via speech her memories of the events from the movie. At the same time, the listener must comprehend and construct the movie's events in her mind, even though she did not watch the movie herself. To understand the neural processes that enable this seemingly effortless transaction, we need to study three stages: 1) the speaker's encoding and retrieval (Bird, Keidel, Ing, Horner, & Burgess, 2015; Chen et al., 2017); 2) the linguistic communication from speaker to listener (Silbert, Honey, Simony, Poeppel, & Hasson, 2014; Stephens, Silbert, & Hasson, 2010); and 3) the listener's mental construction of the events (Chow et al., 2014; Kosslyn, Ganis, & Thompson, 2001; Mar, 2004). To date, there has been no work addressing the direct links between the processes of memory, verbal communication, and construction (in the listener's mind) of a single real-life experience. Therefore, it remains a mystery how information from a past experience stored in one person's memory is propagated to another person's brain, and to what degree the listener's neural construction of the experience from the speaker's words resembles the original encoded experience.

To characterize this cycle of memory transmission, neural patterns during encoding, spoken recall, and mental construction of each scene in a movie were compared (Figure 2.1). To closely mimic a real-life scenario, the study consisted of *movie-viewers* who watched a continuous movie narrative, a person (*speaker*) watching and then freely verbally recalling the same movie, and

finally naïve *listeners*, who had never seen the movie, listened to the audio recording of the spoken description. Scene-specific neural patterns common across the three conditions were evaluated. To ensure the robustness of the results, the full study was replicated using a second movie. This design allowed us to map the neural processes by which information is transmitted across brains in a real-life context, and to examine relationships between neural patterns underlying encoding, communication, and construction.

Why should we expect scene-specific neural patterns in high-order areas to be similar during the encoding, spoken recall, and mental construction of a given event? Resemblance between neural patterns elicited during encoding and retrieval have been shown in numerous studies using different types of stimuli (Polyn, Natu, Cohen, & Norman, 2005; J. D. Johnson, McDuff, Rugg, & Norman, 2009; Buchsbaum, Lemire-Rodger, Fang, & Abdi, 2012; St-Laurent, Abdi, Bondad, & Buchsbaum, 2014; Bird et al., 2015) over the past decade. More recently, it was demonstrated that *scene-specific* neural patterns elicited during encoding of complex natural stimuli (an audio-visual movie) are reinstated during free spoken recall (Chen et al., 2017). Free spoken recall of a movie differs from the original experience (movie) in many aspects. Not only are they different modalities, but the content of each scene is altered during recall: recall is usually shorter and less detailed, with some elements emphasized and others minimized, based on the recaller's preference or retrieval success. What are the neural processes underlying listening to such a recalled narrative? While no study has compared scene-specific patterns of brain responses during mental construction of a story with the patterns elicited during initial encoding or subsequent recall of the same event, recent studies suggest that the same areas that encode and retrieve episodic memories are also involved in the construction of imaginary and future events (Addis, Wong, & Schacter, 2007; Cabeza & St Jacques, 2007; Hassabis & Maguire, 2007, 2009; Rugg & Vilberg, 2013; Daniel L. Schacter, Addis, & Buckner, 2007; Spreng et al., 2009; Svoboda et al., 2006; Szpunar et al.,

2007). These areas include retrosplenial and posterior parietal cortices, ventromedial prefrontal cortex, bilateral hippocampus, and parahippocampal gyrus, known as default mode network (R. Buckner et al., 2008; Raichle et al., 2001). Why are the same brain areas active during episodic encoding, retrieval, and mental construction? One possibility is that the same brain areas are involved in encoding, retrieval, and construction, but these areas assume different activity states during each process; in this case, one would expect that neural representations present during encoding and retrieval of specific scenes would not match those present during mental construction of those scenes. Another possibility is that the same neural activity patterns underlie the encoding, retrieval, and construction of a given scene. This hypothesis has never been tested.

Our communication protocol (Figure 2.1) provides a testbed for this latter hypothesis. In my experiment, during the spoken recall phase, the speaker must retrieve and reinstate her episodic memory of the movie events. At the same time, the listeners, who never experienced the movie events, must construct (imagine) the same events in their minds. Thus, if the same neural processes underlie both retrieval and construction, then similar activity patterns might emerge in the speaker's brain and the listener's brains while recalling/constructing each event. Furthermore, if the speaker successfully communicates her experiences of the original events to the listeners, then we should predict similarity between the neural patterns during the encoding phase (movie viewing) and construction phase (listening to the verbal description without viewing). These questions were addressed by implementing whole brain analysis as well as additional ROI examination. ROI analysis was mainly focused on PCC, a key node in DMN (R. Buckner et al., 2008), as previous research suggest this region is highly involved in natural communication (Stephens et al., 2010), retrieval of autobiographical memories (Maddock, Garrett, & Buonocore, 2001), neural reinstatement (Chen et al., 2017), and memory consolidation (Bird et al., 2015).

In the current study it was witnessed, for the first time, how an event-specific pattern of activity can be traced throughout the communication cycle: from encoding, to spoken recall, to comprehending and constructing (Figure 2.1). Our work reveals the intertwined nature of memory, mental construction, and communication in real-life settings, and explores the neural mechanisms underlying how we transmit information about real-life events to other brains.

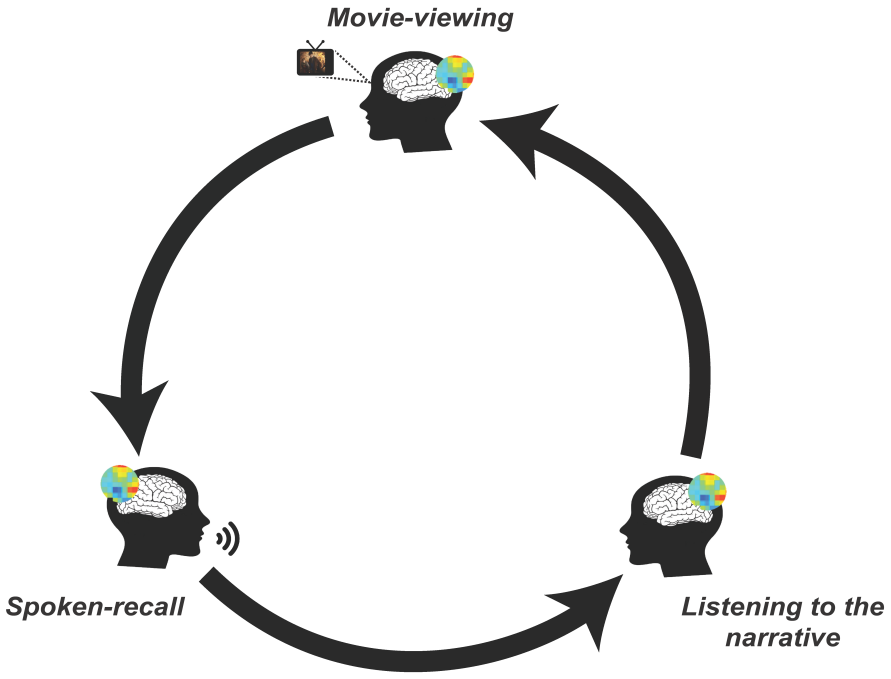


Fig 2.1: Circle of communication Depiction of the entire procedure during sharing of an experience. Participants encode the movie and then reinstate it during recall. By listening to the audio recall, listeners construct the movie events in their mind. Mental representations related to the movie are shared throughout this cycle and transmitted across the brains via communication.

2.2 Materials and Methods

Stimuli

Two audio-visual movies, the first episodes of Sherlock BBC (24-min length) and Merlin (25-min length) BBC were used. These movies were chosen to have similar levels of action, dialogue, and production quality. Audio recordings were obtained from a participant who watched and recounted the two movies in the scanner (free-recall). The outcome was an 18-min audio recording of the Sherlock story, and a 15-min audio recording of the Merlin story. Thus the stimuli consisted of a total of two movies (Sherlock and Merlin) and two corresponding audio recordings. This allowed us to internally replicate the results across the two datasets.

Participants

A total of 52 participants (age 18 – 45) who were all right-handed native English speakers with normal or corrected to normal vision were scanned. Before contacting participants, their previous exposure to both movie stimuli was screened and only people without any self-reported history of watching either of the two movie stimuli were recruited. From the total group, 4 were dropped due to head motions larger than 3 mm (voxel size), 1 was dropped due to anomalous anatomy, 4 fell asleep, 5 were dropped due to failure in post scan memory test (recall levels < 1.5 SD below the mean), and 2 were dropped who had watched the movie but did not report it before the scan session. Subjects who were dropped due to poor recall had scores close to zero (Merlin scores: max = 25, min = 0.4, mean = 11.9, std = 7.1 Sherlock scores: max = 21.4, min = 0, mean = 11.18, std = 5.6). Informed consent was acquired from all participants, which was approved by Princeton University Institutional Review Board.

Procedure

Experimental design. Eighteen participants watched a 25-minute audiovisual movie (from the first episode of BBC's *Merlin*) while undergoing fMRI scanning (*movie-viewing*, Figure 2.2-A). One participant separately watched the movie and then recalled it aloud inside the scanner (unguided, without any experimenter cues) and her spoken description of the movie was recorded (*spoken-recall*). Another group of participants (N = 18) who were naïve to the content of the movie listened to the recorded narrative of ~ 15 minutes (*listening*). The entire procedure was repeated with a second movie (from the first episode of BBC's *Sherlock*), with the same participant serving as the speaker. This design allowed us to internally replicate each of our findings and demonstrate the robustness of the results.

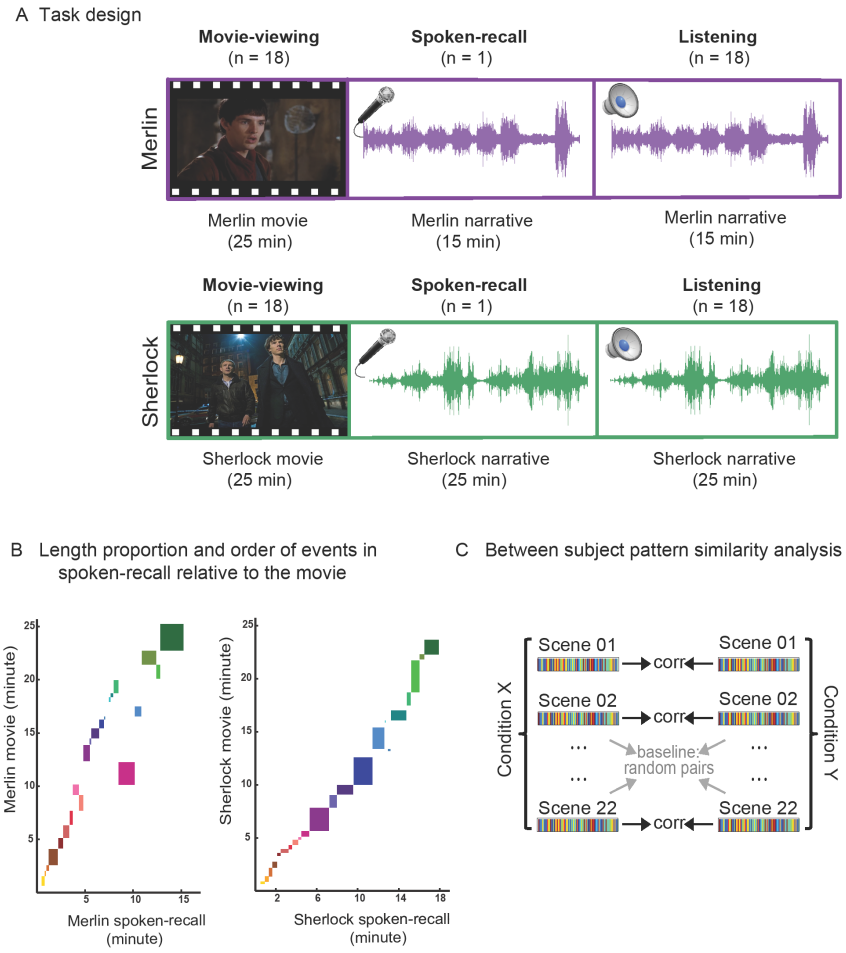


Fig 2.2: Experiment design and analysis A. 18 participants watched a 25-minute audiovisual movie (Merlin) while undergoing fMRI scanning (movie-viewing). One participant separately watched the movie and then recalled it inside the fMRI scanner and her spoken description of the movie was recorded (spoken-recall). Then a group of 18 participants who were naïve to the content of the movie listened to the recorded narrative. The entire procedure was repeated with a second movie (Sherlock) by recruiting new groups of participants. **B.** Depiction of the length of each event in the movie (y-axis) relative to its corresponding event (if remembered) in spoken-recall (x-axis) for each movie. Each box denotes a different event. Boxes that are out of the continuous diagonal string of events depict the events that were recalled in an order different from their original place in the movie. **C.** Schematic for the main analysis. Brain data were averaged within each scene in the data set of each condition (e.g. condition x = movie-viewing and condition Y = spoken-recall). Averaging resulted in a single pattern of brain response across the brain for each scene

for each condition. Then these two patterns were compared and correlated using a searchlight method. Significant values were computed by shuffling the scene labels and comparing the non-matching scenes. Similar analyses were performed for all other comparisons (spoken-recall to listening, listening to movie-viewing)

After the main experiment, participants listened to a short audio stimulus (15 minutes) in the scanner. Data from this run was collected for a separate experiment and was not used in this paper. Participants were randomly assigned to watch Sherlock (n = 18) or Merlin (n =18). The speaker was chosen from the participants of a previous project in our lab (Chen et al 2017). In that study the task for all participants was to watch the Sherlock movie and perform the spoken-recall in the same scan session, describing as many scenes as they could remember. None of the participants in that study had seen the Sherlock movie before. One of them, selected based on recall performance, was called back to watch and recall Merlin movie. Her spoken recollections were recorded and used as the audio narratives in the current study. The speaker's speech was recorded during the fMRI scan using a customized MR-compatible recording system (FOMRI II; Optoacoustics Ltd). The MR recording system uses two orthogonally oriented optical microphones. The reference microphone captures the background noise, whereas the source microphone captures both background noise and the speaker's speech utterances (signal). A dual-adaptive filter subtracts the reference input from the source channel (using a least mean square approach). To achieve an optimal subtraction, the reference signal is adaptively filtered where the filter gains are learned continuously from the residual signal and the reference input. To prevent divergence of the filter when speech is present, a voice activity detector is integrated into the algorithm. A speech enhancement spectral filtering algorithm further preprocesses the speech output to achieve a real-time speech enhancement. Finally, after the recording, the remaining noise was further cleaned

using noise removal software (Adobe Audition). The resulted recording was good enough to be comprehend by all 18 naïve listeners as indicated by our post listening comprehension test. During listening, sound level was adjusted separately for each participant to assure a complete and comfortable understanding of the stimuli. An anatomical scan was performed at the end of the scan session. Before the experiment, participants were instructed to watch and/or listen to the stimuli carefully and were told that there would be memory tests for each part separately.

There was no memory task (or any task) inside the scanner and there was no specific instruction about fixating to the center. Participants were asked to watch the stimuli through the mirror which was reflecting the rear screen. The movie was projected to this screen located at the back of the magnet bore via a LCD projector. In-ear headphones were used for the audio stimuli. Eye-tracking was performed during all the runs (recording during the movie, observing the eye during the audio) using iView X MRI-LR system (SMI Sensomotoric Instruments). Eye-tracking was implemented to ensure that participants were paying full attention and not falling asleep. They were asked to keep their eyes open even during the audio runs (no visual stimuli). The movie and audio stimuli were presented using Psychophysics Toolbox [http://psychtoolbox.org] in MATLAB, which enabled us to coordinate the onset of the stimuli (movie and audio) and data acquisition.

MRI acquisition: MRI data was collected on a 3T full-body scanner (Siemens Skyra) with a 16-channel head coil. Functional images were acquired using a T2*-weighted echo planar imaging (EPI) pulse sequence (TR 1500 ms, TE 28 ms, flip angle 64, whole-brain coverage 27 slices of 4 mm thickness, in-plane resolution 3 x 3 mm2, FOV 192 x 192 mm2). Anatomical images were acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) pulse sequence (0.89 mm3 resolution). Anatomical images were acquired in an 8-minute scan after the functional scan with no stimulus on the screen.

Post-scan behavioral memory test

Memory performance was evaluated using a free recall test in which participants were asked to write down the events they remembered from the movie and audio recording with as much detail as possible. There was no time limitation and they were asked to ensure they wrote everything that they remembered. Three independent raters were asked to read the transcripts of participants' free recalls and to assign memory scores to each participant. The raters were given general instructions to assess the quality of the comprehension and accuracy of each response, and a few examples. Suggested items to consider were the number of scenes remembered, importance of the remembered scenes, details provided, and overall comprehension/memory level. However, the raters could flexibly choose how they score each of these items. This flexibility was provided to obtain a better coverage of various dimensions in evaluating such complex stimuli (which could be subjective without sharply defined right or wrong answers), as well as having a common ground for rating. They reported a score for each participant and these numbers were rescaled to the same scale across (minimum and maximum) the three raters. Ratings generated by of the three raters were highly correlated (Cronbach's Alpha = 0.85 and 0.87 for Merlin and Sherlock respectively) and averaged to be used in further analysis.

Data analysis

Preprocessing was performed in FSL [http://fsl.fmrib.ox.ac.uk/fsl], including slice time correction, motion correction, linear detrending, and high-pass filtering (140 s cutoff). These were followed by coregistration and transformation of the functional volumes to a template brain (MNI). The rest of the analysis was coded and performed using Matlab software (MathWorks). All the time courses were despiked before further analysis. We briefly review the analytical methods and objectives. Before running the searchlight analysis, brain time-courses were averaged within each scene for all the participants and conditions.

Pattern similarity searchlight

Movie-viewing and *spoken-recall* data are not aligned across time-points; it took the speaker 15 minutes to describe the 25-minute Merlin movie, and 18 minutes to describe the 24-minute Sherlock movie (Figure 2.2-B). To compare the brain responses across different conditions, data obtained during the watching of each movie (*movie-viewing)* were divided into 22 scenes (Figure 2.2-C), following major shifts in the narrative (e.g., location, topic, and/or time, as defined by an independent rater). The same 22 scenes were identified in the audio recordings of the recall session based on the speaker's verbal narration. Calculating the arithmetic mean of time points within each scene provided a single pattern of brain response for each scene during movie viewing, recall and listening. Before averaging, time-courses of brain response were shifted 3 TRs to account for HRF delay. Pattern similarity analysis was conducted by calculating the Pearson correlation between the patterns elicited during movie-viewing and the patterns observed during the recall in a searchlight analysis (15 x 15 x 15 mm cubes centered on every voxel in the brain, (Kriegeskorte, Goebel, & Bandettini, 2006; Kriegeskorte, Mur, & Bandettini, 2008). The same analysis was performed to compute pattern similarity between brain responses during recall and listening, and also listening and movie-viewing. This analysis reveals regions containing scene-specific patterns, as statistical significance is only reached if matching scenes (same scene in movie and recall) can be differentiated from non-matching scenes. Significance thresholds were calculated using a permutation method (Kriegeskorte et al., 2008) by shuffling the scene labels and correlating non-matched scenes to create a null distribution of r-values; the p-value was extracted by calculating the proportion of the null distribution that was above the observed level of pattern similarity for matching scenes. This procedure was implemented for all the searchlight cubes for which 50% or more of their volume was inside the brain. Thus individual p values were generated for each voxel (center of searchlight cube) and were corrected for multiple comparisons using False Discovery

Rate (Benjamini & Hochberg, 1995), $q < 0.05$. This analysis aims to confirm the event-specificity of our findings by demonstrating that correlation between matching scenes is significantly higher that non-matching scenes.

Encoding to recall pattern similarity was calculated by executing the searchlight analysis to compare the *spoken-recall* data with each subjects' *movie-viewing* (encoding data) and then averaging across subjects. Pattern similarity analysis was performed across subjects. Therefore, the speaker's movie viewing data was not included in the movie-viewing set. After performing the shuffling and permutation test, the average map was plotted with specific p-values for each voxel, with the threshold corrected using FDR (Figure 2.5.A-B). To compare *speaking-to-listening*, the pattern similarity searchlight was used to compare the *speaker*'s recall data with each of the *listeners'* listening data and then averaged and statistically thresholded (Figure 2.6.A-B). In *listening-to-viewing* condition, each *viewer*'s data was correlated with the average of all the *listeners* listening data. The procedure was done for all the participants in the group and then statistical analysis and averaging was performed to compute the p-value maps (Figure 2.7.A-B). After averaging, maps were thresholded based on significance (FDR correction, $q<0.05$).

Speaker-listener time-course lag

Previous work suggests that during communication, the neural responses observed in the listener follows the speaker's neural response timecourses with a delay of a few seconds (Dikker, Silbert, Hasson, & Zevin, 2014; Silbert et al., 2014; Stephens et al., 2010). To see whether this response lag was also present in the listeners' brains, I calculated the correlation in PCC between the scene-specific neural patterns during *spoken-recall* and *listening* in the spatial domain, with TR-by-TR shifting of listeners' neural timecourses. Figure 2.3-A depicts the r values in the PCC ROI as the TR shift in the listeners was varied from -20 to 20 TRs (-30 to 30 seconds). In agreement with prior findings, we observed a lag between *spoken-recall* and *listening.* In the Merlin movie

correlation peaked (r = 0.17) at a lag of ~5 TRs (7.5 seconds). A similar speaker-listener peak lag correlation at ~5 TRs was replicated in the *listeners* of the Sherlock movie (Fig. 2.3-B). To account for the listeners' lag response, we used this 5 TR lag across the entire brain in all analyses.

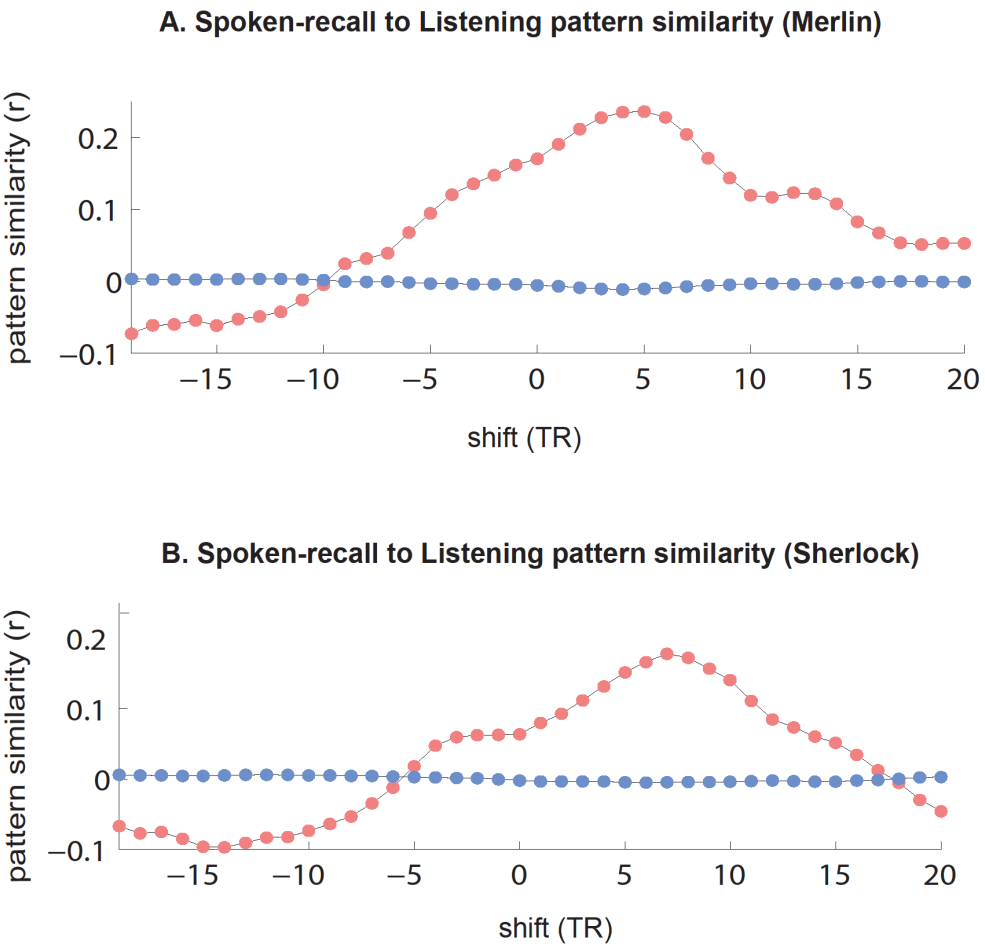


Fig 2.3: Pattern similarity in PCC with shifting of the listening data This figure depicts the r values for pattern similarity between the spoken-recall data and average of all listeners' listening data on y axis. Listening data has been shifted -20 to +20 TRs (x-axis) before calculating pattern similarity. R values for each shift are depicted as a separate dot. R values peak at TR = 5.

ROI-based pattern similarity

In addition, pattern similarity was separately calculated at subject level in posterior cingulate cortex ROI. This analysis was performed by calculating Pearson correlation between patterns of brain response across the entire ROI in each *viewer* to the *speaker* (Figure 2.5.C-D), each *listener* to the *speaker* (Figure 2.6.C-D), and each *viewer* to the average of all *listeners* (Figure 2.7.C-D). The same analysis was performed in auditory cortex, mPFC, angular gyrus, hippocampus and parahippocampal gyrus ROIs (Figure 2.4 A-F). The posterior cingulate and mPFC ROIs were taken from a resting state connectivity atlas (Shirer et al. 2011 May 26). Auditory cortex, hippocampus, and parahipocampal gyrus ROIs were taken from the Harvard-Oxford anatomical atlas (distributed with the FSL software package; http://fsl.fmrib.ox.ac.uck/fsl/). Angular gyrus was defined using LONI Probabilistic Brain Atlas (Shattuck et al. 2008). For each ROI in each condition, significance of the average pattern similarity across subjects was calculated using a permutation test. Scene labels were shuffled and the average of correlation values for matching scenes was calculated 100,000 times. The true averages larger than 99% of the values in their corresponding null distribution were considered significant (asterisks in Figure 2.4).

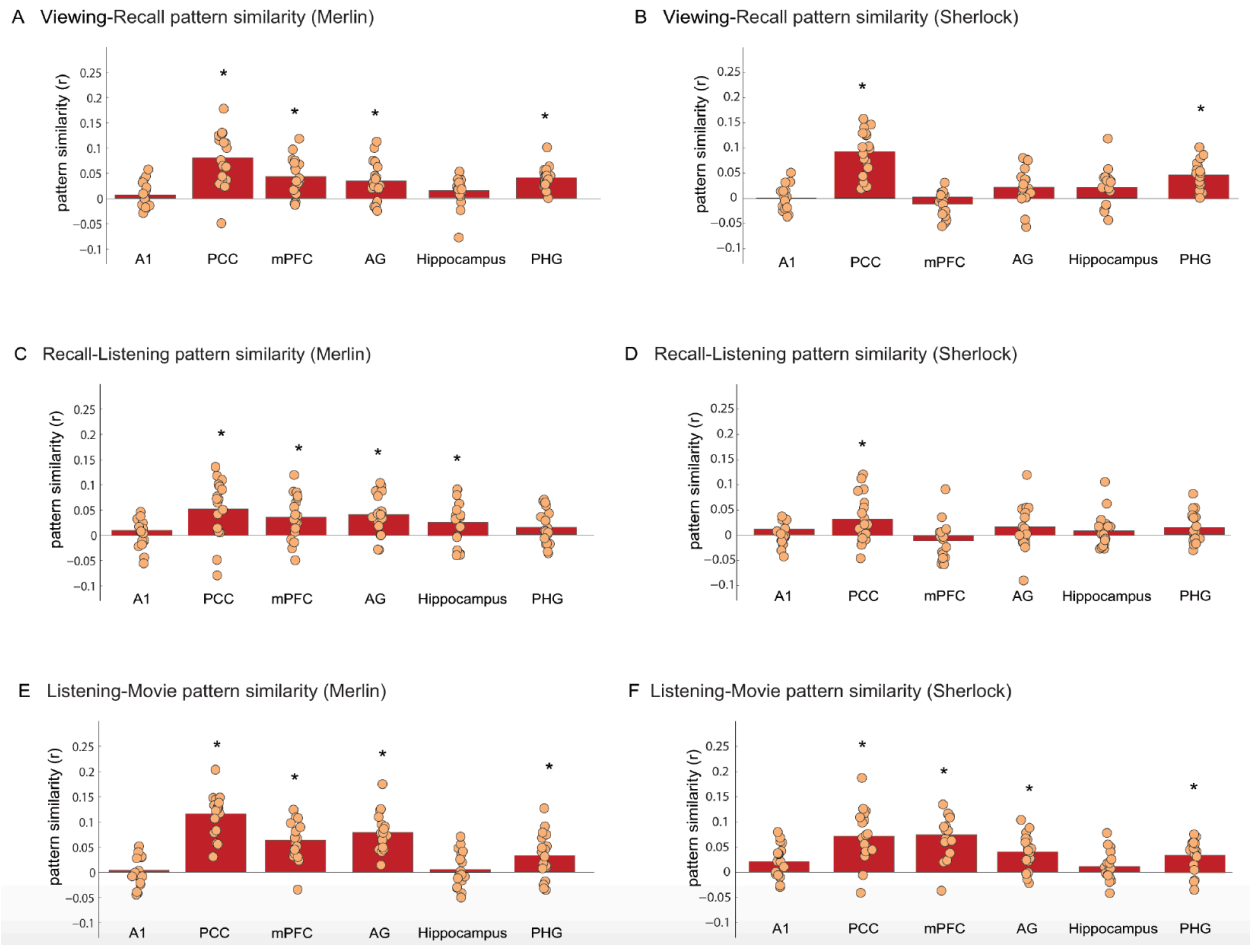


Fig 2.4: ROI based pattern similarity

Red bars show average correlation of matching scenes. Circles depict similarity values for individual subjects. Asterisks indicate $p<0.01$ (averages standing above 99% of null distribution). Panel A,C,E depicts data for the Merlin movie and panel B,D,F depicts data for the Sherlock movie.

Behavioral correlation

To compute the behavioral correlation, patterns similarity in each viewer (including the speaker's viewing) to each listener was calculated in PCC. These patterns similarity values for each viewer (18 values because of the correlation of each viewer with each of the 18 listeners) were then correlated with the listeners' behavioral scores. Figure 2.9-A and 2.9-C show the

correlation for the speaker's viewing and the listeners. Figure 2.9-B and 2.9-D depict the sorted outcome r values of correlation for each movie viewer and show the speaker's viewing in red as one of the highest values.

Classification analysis

To investigate the discriminability of neural patterns for individual scenes, we first averaged the time-course of brain response within each scene during movie viewing and listening. The patterns were then averaged across participants in each group to make an averaged pattern for viewing and an averaged pattern for listening. Pairwise correlation between the two groups for all 22 scenes was computed. Classification was considered successful if the pairwise correlation of any given scene between the movie and listening (matching scene) was higher than their correlation with any other scene (out of 22 possibilities, chance 4.5%). Overall accuracy was then computed by the number of scenes with the highest correlation (rank = 1 for matching scene) divided by the number of scenes. Chance was calculated by shuffling the scene labels and computing the pairwise correlation 10000 times. Null distribution was created using the classification accuracy of the shuffled scenes. Real classification accuracies of both movies stand beyond 99.9 percentile on this distribution.

Triple shared pattern searchlight

The triple shared pattern analysis was performed to directly compare the neural patterns across the three conditions (*movie-viewing, spoken-recall, listening*). We sought to find voxels within each searchlight cube that were correlated across the three conditions. For each scene, the brain response was z-scored across voxels (spatial patterns) within each cube. If the same patterns are present across conditions, then the z-scored activation value for a given voxel should have the

same sign across conditions. To measure this property, we implemented the following computation: For a given voxel in each cube, if it showed all positive or all negative values across the three conditions, we calculated the product of the absolute values of brain response in that voxel. Otherwise (if a voxel did not exhibit all positive or negative signs across the three conditions), the product value was set to zero. The final value for each voxel was then created by averaging these product values across scenes. To perform significance testing, the order of scenes in each condition was randomly shuffled (separately for each condition) and then the same procedure was applied (calculating the product value and averaging). By repeating the shuffling 10000 times and creating the null distribution, p values were calculated for each voxel. The resulting p-values were then corrected for multiple comparisons using FDR ($q < 0.05$).

2.3 Results

Pattern similarity between spoken-recall and movie-viewing

We first asked whether brain patterns elicited during *spoken-recall* (memory retrieval) were similar to those elicited during *movie-viewing* (encoding). To this end, we needed to compare corresponding content across the two datasets, i.e., compare brain activity as the *movie-viewing* participants encoded each movie event to the brain activity as the speaker recalled the same event during *spoken-recall*. Previous work from our lab (Chen et al., 2017) has shown that neural patterns elicited by watching a movie are highly similar across participants at the individual scene level. Therefore, to increase the reliability of the movie-viewing-related patterns, we used the data from 18 viewers (not including the speaker's viewing data) and compared them with the recall data in the speaker.

Pattern similarity analysis (details in method section) revealed a large set of brain regions that exhibited significant scene-specific resemblance between the patterns of brain response during *movie-viewing* and *spoken-recall*. Figure 2.5-A shows the scene-specific *movie-viewing* to *spoken-recall* pattern similarity for the Merlin movie; Figure 2.5-B replicates the results for the Sherlock movie. These areas included posterior medial cortex, medial prefrontal cortex, parahippocampal cortex, and posterior parietal cortex; collectively, these areas strongly overlap with default mode network (DMN). In the posterior cingulate cortex (PCC), a major region of interest (ROI) in the DMN, we observed a positive reinstatement effect in 17 of the 18 subjects in the *Merlin* condition (Fig. 3-C), and 18 out of the 18 subjects in the *Sherlock* condition (Fig. 2.5-D). The DMN has been previously shown to be active in episodic retrieval tasks (Maguire, 2001; Rugg & Vilberg, 2013; Svoboda et al., 2006). Our finding of similar brain activity patterns between encoding and recall of a continuous movie narrative supports previous studies showing reinstatement of neural patterns during recall using simpler stimuli such as words, images, and short videos (Bird et al., 2015; Polyn et al., 2005; St-Laurent et al., 2014; Wing, Ritchey, & Cabeza, 2014). In addition, the result replicates a previous study from our lab that used a different dataset where both movie-viewing and recall were scanned for each participant (Chen et al., 2017).

The above result shows that scene-specific brain patterns presented during the encoding of the movie were reinstated during the spoken free recall of the movie. Next we asked whether listening to a recording of the recalled (verbally described) movie would elicit these same event-specific patterns in an independent group of listeners who had never watched it (*listeners*).

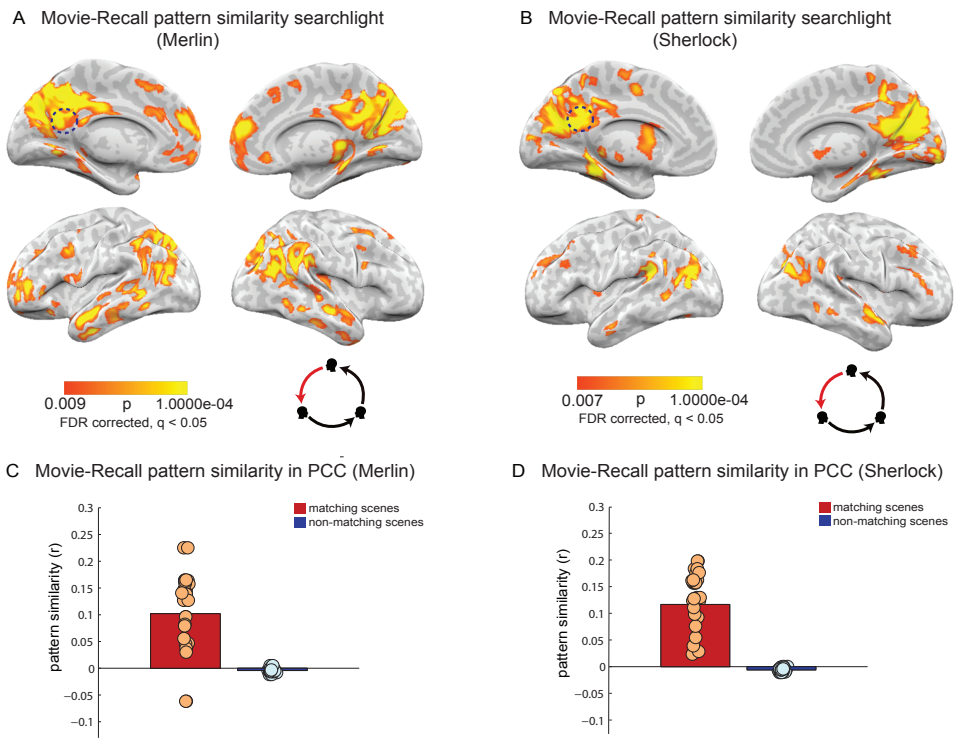


Fig 2.5: Movie-viewing to spoken-recall pattern similarity analysis A-B. Pattern similarity searchlight map, showing regions with significant between-participant, scene-specific correlations (p-values) between spoken-recall and movie-viewing (searchlight was a 5x5x5 voxel cube). Panel A depicts data for the Merlin movie and panel B depicts data for the Sherlock movie. Dotted circle shows the approximate location of the PCC ROI that was used in the analysis in panel C-D **C-D.** Pattern similarity (r-values) of each participant's encoding (movie-viewing) data to the brain response during spoken-recall (in the speaker) in posterior cingulate cortex. Red bar shows average correlation of matching scenes and blue bar depict average correlation of non-matching scenes, averaged across subjects. Circles depict values for individual subjects. Panel C depicts data for the Merlin movie and panel D depicts data for the Sherlock movie.

Pattern similarity between spoken-recall and listening

Previous studies have provided initial evidence for neural alignment (correlated responses in the temporal domain using inter-subject correlation) between the responses observed in the speaker's brain during the production of a story and the responses observed in the listener's brain during the comprehension of the story (Silbert et al., 2014; Stephens et al., 2010). Moreover, it has been shown that higher speaker-listener neural coupling predicts successful communication and narrative understanding (Stephens et al., 2010). However, it is not known whether similar scene-specific *spatial* patterns will be observed across communicating brains, and where in the brain such similarity exists. To test this question, we implemented pattern similarity analysis (see methods) as in the previous section; however, for this analysis we correlated the average scene-specific neural patterns observed in the speaker's brain during spoken recall with the average scene-specific neural patterns observed in the listeners' brains as they listened to a recording of the spoken recall.

We observed significant scene-specific correlation between the speaker's neural patterns during the spoken recall and the listeners' neural patterns during speech comprehension. Scene-specific neural patterns were compared between the *spoken-recall* and *listening* conditions using a searchlight and were corrected for multiple comparisons using FDR ($q<0.05$). Figure 2.6-A shows the scene-specific *spoken-recall* to *listening* pattern similarity for the Merlin movie; Figure 2.6-B replicates the results for the Sherlock movie. Similarity was observed in many of the areas that exhibited the memory reinstatement effect (movie-spoken recall correlation, Figure 2.6), including angular gyrus, precuneus, retrosplenial cortex, posterior cingulate cortex and mPFC.

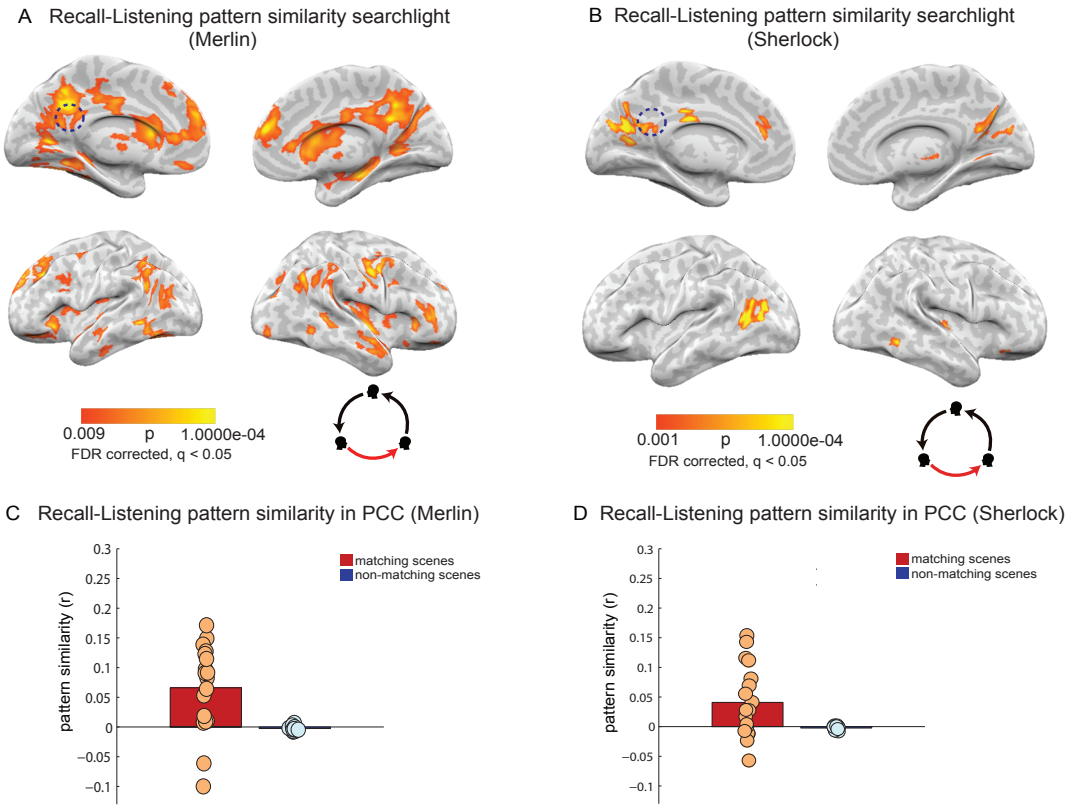


Fig 2.6: Spoken-recall to listening pattern similarity analysis A-B. Pattern similarity searchlight map, showing regions with significant between-participant, scene-specific correlations (p-values) between spoken-recall and listening (searchlight was a 5x5x5 voxel cube). Panel A depicts data for the Merlin movie and panel B depicts data for the Sherlock movie. Dotted circle shows the approximate location of the PCC ROI that was used in the analysis in panel C-D **C-D.** Pattern similarity (r-values) of each participant's listening data to the brain response during spoken-recall (in the speaker) in posterior cingulate cortex. Red bar show average correlation of matching scenes and blue bar depict average correlation of non-matching scenes, averaged across subjects. Circles depict values for individual subjects. Panel C depicts data for the Merlin movie and panel D depicts data for the Sherlock movie.

Pattern similarity between listening and movie-viewing

So far we have demonstrated that event-specific neural patterns observed during encoding in high-order brain areas were reactivated in the speaker's brain during spoken recall; and that some

aspects of the neural patterns observed in the speaker were induced in the listeners' brains while they listened to the spoken description of the movie. If speaker-listener neural alignment is a mechanism for transferring event-specific neural patterns encoded in the memory of the observer to the brains of naive listeners, then I predict that the neural patterns in the listeners' brains during the construction of each event will resemble the movie-viewers' neural patterns during each scene. To test this, I compared the patterns of brain responses when people listened to a verbal description of that event (*listening)* with those when people encoded the actual event while watching the movie (*movie-viewing)*.

It was found that the event-specific neural patterns observed as participants watched the movie were significantly correlated with neural patterns of naïve listeners who listened to the spoken description of the movie. Figure 2.7-A shows the scene-specific *listening* to *movie-viewing* pattern similarity for the Merlin movie; Figure 2.7-B replicates the results for the Sherlock movie. Similarity was observed in many of the same areas that exhibited memory reinstatement effects (*movie-viewing* to *spoken-recall* correlation Figures 2.6) and speaker-listener alignment (Figures 2.6), including angular gyrus, precuneus, retrosplenial cortex, posterior cingulate cortex and mPFC. Computing the scene-specific *listening* to *movie-viewing* pattern similarity within the same PCC ROI shows that effect was positive for each of the individual subjects in each of the movies (Figure 2.7 C-D).

To confirm that the relationship between the viewing and listening patterns was scene-specific, I assessed whether I could classify *which scene* participants were hearing about (in the listening condition) by matching scene-specific patterns from the listening condition to scene-specific patterns from the viewing condition. Average patterns in posterior cingulate cortex for each scene were created separately for viewing and listening groups. On average, the neural pattern observed during movie-viewing of a particular scene was most similar to the pattern observed

when listening to a verbal description of the scene (average classification accuracy for Merlin = 27%, p = 0.0002 1-tailed, Sherlock = 22%, p = 0.001 1-tailed, chance level = 4.5%, Fig 2.8), even though participants listening to the verbal description had not previously seen the movie.

I conducted ROI-based pattern similarity analyses (same as for PCC in Figure 2.5 – 2.7 C-D) for angular gyrus, mPFC, hippocampus, parahippocampal gyrus, and auditory cortex (Figure 2.4 A-F) for all three comparisons: movie-viewing to spoken-recall, spoken-recall to listening, and listening to movie-viewing. As expected, the control ROI (auditory cortex) did not exhibit reliable pattern similarity in any comparison. In contrast, areas within DMN such as PCC, AG, and mPFC showed significant pattern similarity in most comparisons. PHG showed significant pattern reinstatement during recall and pattern similarity between movie encoding and story encoding. However, hippocampus did not exhibit reliable pattern similarity in most of the comparisons, with the exception of recall-listening in the Merlin movie.

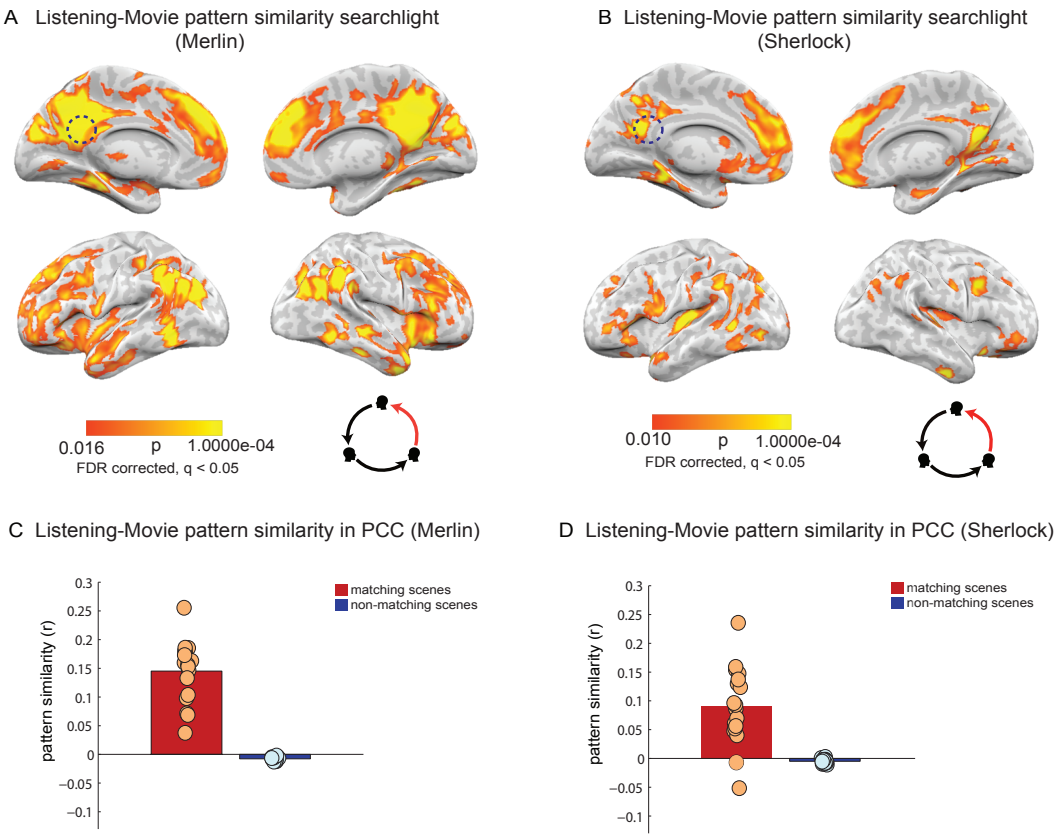


Fig 2.7: Listening to movie-viewing pattern similarity analysis A-B. Pattern similarity searchlight map, showing regions with significant between-participant, scene-specific correlations (p-values) between movie-viewing and listening (searchlight was a 5x5x5 voxel cube). Panel A depicts data for the Merlin movie and panel B depicts data for the Sherlock movie. Dotted circle shows the approximate location of the PCC ROI that was used in the analysis in panel C-D **C-D.** Pattern similarity (r-values) of each participant's movie-viewing data to the average of all other listeners in posterior cingulate cortex. Red bar show average correlation of matching scenes and blue bar depict average correlation of non-matching scenes, averaged across subjects. Circles depict values for individual subjects. Panel C depicts data for the Merlin movie and panel D depicts data for the Sherlock movie.

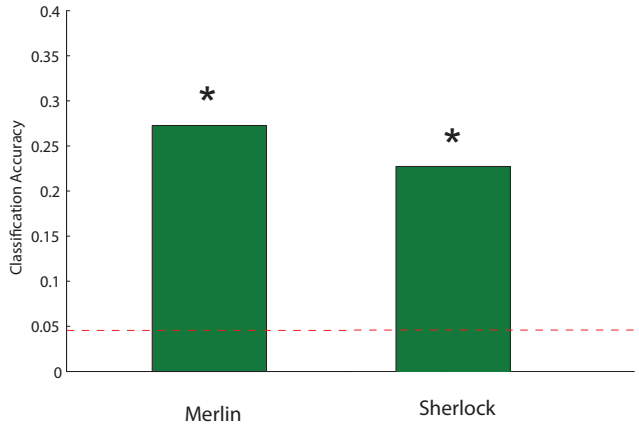


Fig 2.8: Viewing to Listening Scene Classification Bar graphs depict overall classification accuracy between viewing and listening. This shows that, on average, a given scene could be correctly distinguished from other scenes even across modalities. The red line indicates the probability of correctly identifying a scene among 22 scenes based on guessing alone (4.5%). Based on a permutation test, the average classification accuracy for both movies was well above chance ($p = 0.0002$ and $p = 0.001$ for Merlin and Sherlock respectively).

Relationship between pattern similarity and behavioral performance

Given that the speaker's success in transmitting her memories may vary across listeners, I next asked whether the level of correlation between the neural responses of each listener and the speaker's neural responses while encoding the movie can predict the listeners' comprehension level. To test this question, I looked at the posterior cingulate cortex (PCC). The PCC was chosen as the region of interest since previous research has shown that the strength of similarity between spatial patterns of brain response during encoding and rehearsal in this area could predict the subsequent memory performance (Bird et al. 2015). Indeed, within the posterior cingulate cortex, speaker-listener neural alignment (correlation) predicted the level of comprehension in the listeners, as measured with an independent post-scan test of memory and comprehension (Figure

2.9-A and 2.9-C, R = 0.71 and P = 0.001 for the Merlin movie, R = 0.54 and P = 0.022 for the Sherlock movie).

Different people could vary in the way they encode and memorize the same events in the movie. These idiosyncrasies would then be transmitted to listeners when a particular speaker recounts her memory. A successful transmission of a particular episodic memory, therefore, may entail a stronger correspondence between the neural responses of the listeners with those of the *speaker watching the movie*, as opposed to with other viewers watching the same movie. To test this hypothesis, I compared the listeners' comprehension levels with the correlation between neural patterns in posterior cingulate cortex of each movie viewer (including the speaker, N = 19) and each listener during listening. I observed that the listeners' comprehension levels were predicted the best when I compared the listeners' neural patterns with those of the *actual speaker viewing the movie*, relative to all other 18 viewers (Figure 2.9-B); and among the top three in the replication study (Figure 2.9-D). These results indicate that during successful communication the neural responses in the listeners' brains were aligned with neural responses observed in the speaker's brain during encoding (viewing) the movie, even before recall had begun. While I was able to replicate these results across both studies, I should interpret them with caution given the relatively small number of subjects (n=18) in each study. Future studies are needed to better understand how the idiosyncratic way in which a speaker encodes an episode may shape the way it is transmitted to the listeners.

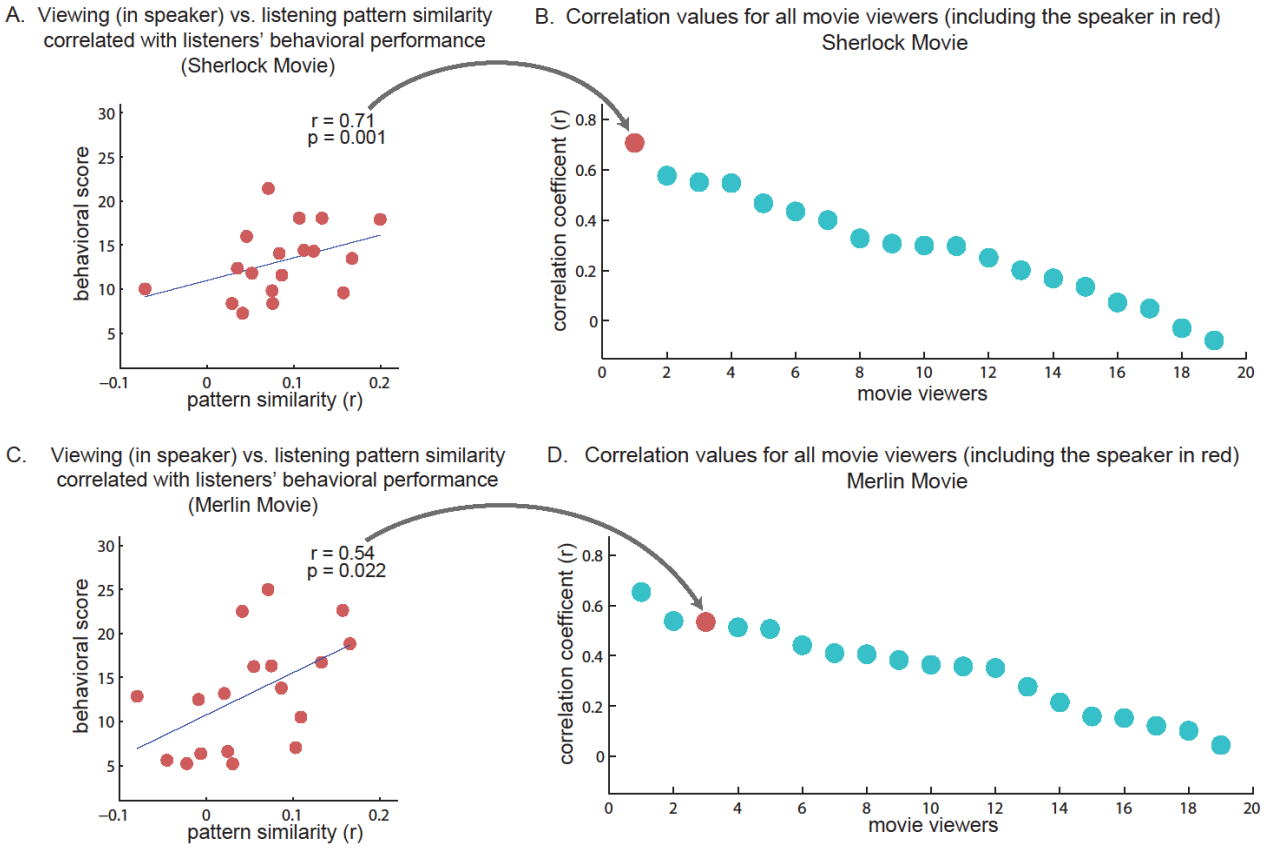


Fig 2.9: Pattern similarity of movie viewers to listeners – relationship to behavior. A,C Correlation between the comprehension score of listeners and degree of similarity between the speaker neural responses during the encoding phase (i.e. while watching the movie) and all listeners in PCC, for each movie. **B,D** Rank order correlation values of the same analysis as in A,C for each of the viewers (including the speaker, red circle). Note that correlating the listeners' brain responses with the actual speaker's brain responses during encoding phase better predicted comprehension levels than the correlation with other viewers.

Shared neural response across three conditions (triple shared pattern analysis)

In Figures 2.5, 2.6 and 2.7 I show the pairwise correlations between encoding, speaking, and constructing. The areas revealed in these maps are confined to high order areas, which overlap with the default mode network, and include the TPJ, angular gyrus, retrosplenial, precuneus, posterior cingulate cortex and mPFC. Such overlap suggests that there are similarities in the neural

patterns, which are shared at least partially, across conditions. Correlation, however, is not transitive (beside the special case when the correlation values are close to 1). That is, if x is correlated with y, y is correlated with z, and z is correlated with x, one can't conclude that a shared neural pattern is common across all three conditions. To directly quantify the degree to which neural patterns are shared across the three conditions, I developed a new, stringent three-way similarity analysis to identify shared event-specific neural patterns across all three conditions (movie encoding, spoken recall, naïve listening). The analysis looks for shared neural patterns across all conditions, by searching for voxels that fluctuate together (either going up together or down together) in all three conditions (see methods for details). Figure 2.10-A shows all areas in which the scene-specific neural patterns are shared across all three conditions in the Merlin movie; Figure 2.10-B replicates the results in the Sherlock movie. These areas substantially overlap with the pairwise maps (Figures 2.5, 2.6 and 2.7), thereby indicating that similarities captured by the pairwise correlations include patterns that are shared across all three conditions. Note that the existence of shared neural patterns across conditions does not preclude the existence of additional response patterns that are shared across only two of the three conditions (e.g. shared responses across the speaker-listener which are not apparent during movie encoding), and revealed in the pair-wise comparisons (Figures 2.5, 2.6 and 2.7).

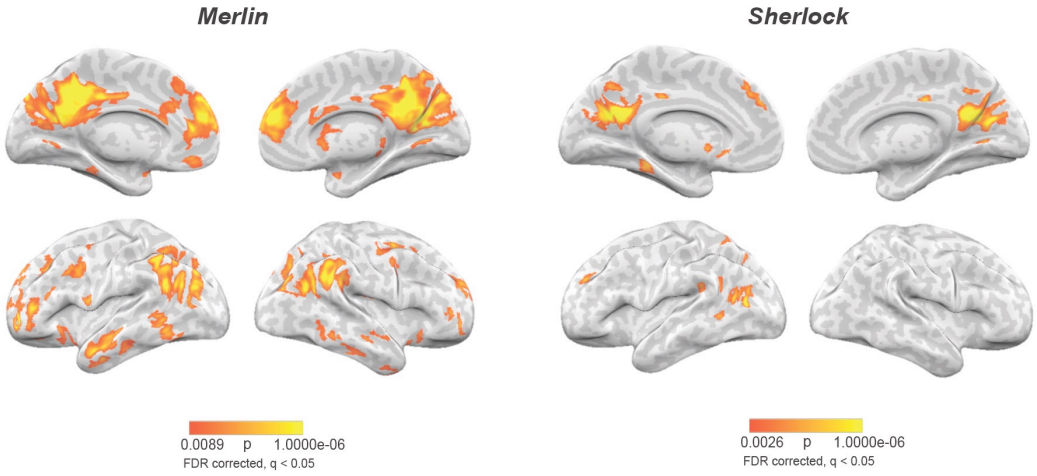


Fig 2.10: Shared neural patterns across all conditions Regions showing scene-specific pattern correlations across movie-viewing, spoken-recall, and listening for **A.** the Merlin movie **B.** the Sherlock movie.

2.4 Discussion

This study reports, for the first time, that shared event-specific neural patterns are observed in the default mode network (DMN) during the encoding, reinstatement (spoken recall), and new construction of the same real-life episode. Furthermore, across participants, higher levels of similarity between the speaker's neural patterns during movie viewing and the listeners' neural patterns during mental construction were associated with higher comprehension of the described events in the listeners (i.e., successful "memory transmission"). Prior studies have shown that neural patterns observed during the encoding of a memory are later reinstated during recall (Bird et al., 2015; Buchsbaum et al., 2012; Chen et al., 2017; J. D. Johnson et al., 2009; Polyn et al., 2005; St-Laurent et al., 2014). Furthermore, it has been reported that the same areas that are active during recall are also active during prospective thinking and mental construction of imaginary

events (Addis et al., 2007; R. L. Buckner & Carroll, 2007; Hassabis, Kumaran, & Maguire, 2007; Daniel L. Schacter et al., 2007; Spreng et al., 2009; Szpunar et al., 2007). This study is the first to directly compare scene-specific neural patterns observed during *mental construction (imagination) of a verbally-described but never experienced event* directly to patterns elicited during *audio-visual perception of the original event*. This comparison, which was necessarily performed across-participants, revealed brain areas throughout the DMN, including posterior medial cortex, mPFC, and angular gyrus, where spatial patterns were shared across both spoken recall and mental construction of the same event.

Why do we see such a strong link between memory encoding, spoken recall and construction? By identifying these shared event-specific neural patterns, I hope to illustrate an important purpose of communication: to transmit and share one's thoughts and experiences with other brains. In order to transmit memories to another person, a speaker needs to convert between modalities, using speech to convey what she saw, heard, felt, smelled, or tasted. In the experimental setup, during the spoken recall, the speaker focused primarily on the episodic narrative (e.g., the plot, locations and settings, character actions and goals), rather than on fine sensory (visual and auditory) details. Accordingly, *movie-viewing* to *spoken-recall* pattern correlations were not found in low level sensory areas, but instead were located in high level DMN areas, which have been previously found to encode amodal abstract information (Binder et al., 2009; Honey, Thompson, Lerner, & Hasson, 2012; Regev, Honey, Simony, & Hasson, 2013). Future studies could explore whether the same speech-driven recall mechanisms can be used to reinstate and transmit detailed sensory memories in early auditory and visual cortices.

Spoken words not only enabled the reinstatement of scene-specific patterns during recall, but also enabled the construction of the same events and neural patterns as the listeners imagined those

scenes. For example, when the speaker says "Sherlock looks out the window, sees a police car, and says, well now it's four murders", she uses just a few words to evoke a fairly complex situation model. Remarkably, a few brief sentences such as this are sufficient to elicit neural patterns, specific to this particular scene, in the listener's DMN that significantly resemble those observed in the speaker's brain during the scene encoding. Thus, the use of spoken recall in the study exposes the strong correspondence between memories (event reconstruction) and event construction (imagination). This intimate connection between memory and imagination (Hassabis, Kumaran, & Maguire, 2007; Hassabis, Kumaran, Vann, & Maguire, 2007; Hassabis & Maguire, 2009; Romero & Moscovitch, 2012) allows us not only to share our memories with others, but also to invent and share imaginary events with others. Areas within the DMN have been proposed to be involved in creating and applying "situation models" (Ranganath & Ritchey, 2012; Zwaan & Radvansky, 1998), and changes in the neural patterns in these regions seem to mark transitions between events or situations (Baldassano et al., 2016; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). An interesting possibility is that the (re)constructed "situation model" is the "unit" of information transferred from the speaker to the listener, a transfer made compact and efficient by taking advantage of their shared knowledge.

I showed that, despite the differences between the verbal recall and the movie, listening to a recalled narrative of the movie triggered mental construction of the events in the listeners' brain and enabled them to partly experience a movie they have never watched. Similarity between patterns of brain response during perception and imagination has been reported before (Cichy, Heinzle, & Haynes, 2012; M. R. Johnson & Johnson, 2014; Stokes, Thompson, Cusack, & Duncan, 2009; Vetter, Smith, & Muckli, 2014). These studies have mostly focused on visual or auditory imagery of static objects, scenes and sounds. In this study, I directly compared, for the first time,

scene-specific neural patterns during mental construction of rich episodic content, which describes the actions and intentions of characters as embedded in real-life dynamical movie narrative.

In agreement with the hypothesis that the speaker's verbal recall transmitted her own idiosyncratic memory of the movie, I found the listeners correlated better with the speaker's neural patterns during the encoding of the movie, relative to neural responses in other viewers that watched the movie. Furthermore, the ability of the speaker to successfully transmit her memories can vary as a function of how successful the listeners are in constructing the information in their minds. And indeed I observed that the strength of speaker-listener neural alignment correlated with the listeners' comprehension as measured by a post-scan memory and comprehension tests. Taken together, these results suggest that the alignment of brain patterns between the speaker and listeners can capture the quality of transfer of episodic memories across brains. This finding extends previous research that showed a positive correlation between communication success and speaker-listener neural coupling in the temporal domain (Dikker et al., 2014; Silbert et al., 2014; Stephens et al., 2010) in posterior medial cortex, and is also consistent with research showing that higher levels of encoding-to-recall pattern similarity in posterior cingulate cortex positively correlate with behavioral memory measures (Bird et al., 2015). The result highlights the importance of the subjectivity and uniqueness of the original experience (how each person perceives the world, which later on affects how they retrieve that information) in transmission of information across different brains. Additional research is needed to assess the variability in such capacity among speakers as they share their memories with other subjects.

What causes some listeners to have weaker or stronger correlation with the speaker's neural activity? Listeners may differ in terms of their ability to construct and understand second hand information that is transmitted by the speaker. The speaker's recall is biased toward those parts of

the movie which are more congruent with her own prior knowledge, and the listener's comprehension and memory of the speaker's description is also influenced by his/her own prior knowledge (Alba & Hasher, 1983; F. C. Bartlett & Burt, 1933; John D. Bransford, 1972; Romero & Moscovitch, 2012). Thus, the coupling between speaker and listener is only possible if the interlocutors have developed a shared understanding about the meaning and proper use of each spoken (or written) sign (Clark, 2006; Clark & Krych, 2004; Pickering & Garrod, 2004). For example, if instead of using the word "police officers" the speaker uses the British synonym "bobbies", she is likely to be misaligned with many of the listeners. Thus, the construction of the episode in the listeners' imagination can be aligned with speaker's neural patterns (associated with the reconstruction of the episode) only if both speaker and listener share the rudimentary conceptual elements that are used to compose the scene.

Finally, it is important to note that information may change in a meaningful or useful way as it passes through the communication cycle; the three neural patterns associated with encoding, spoken recall, and construction are similar but not identical. For example, in a prior study I documented systematic transformations of neural representations between movie encoding and movie recall (Chen et al., 2017). In the current study, I observed that the verbal description of each scene seemed to be compressed and abstracted relative to the rich audio-visual presentation of these events in the movie. Indeed, at the behavioral level, I found that most of the scene recalls were shorter than the original movie scene (e.g., in this study it took the speaker ~15-18 minutes to describe a ~25-minute movie). Nevertheless, the spoken descriptions were sufficiently detailed to elicit replay of the sequence of scene-specific neural patterns in the listeners' DMNs. Because the DMN integrates information from multiple pathways (Binder & Desai, 2011; Margulies et al., 2016), I propose that, as stimulus information travels up the cortical hierarchy of timescales during

encoding, from low-level sensory areas up to high-level areas, a form of compression takes place (Hasson, Chen, & Honey, 2015). These compressed representations in the DMN are later reactivated (and perhaps further compressed) using spoken words during recall. It is interesting to note that the listeners may benefit from the speaker's concise speech, as it allows them to bypass the step of actually watching the movie themselves. This may be an efficient way to spread knowledge through a social group (with the obvious risk of missing on important details), as only one person needs to expend the time and run the risks in order to learn something about the world, and can then pass that information on to others.

Overall, this study tracks, for the first time, how real-life episodes are encoded and transmitted to other brains through the cycle of communication. Sharing information across brains is a challenge that the human race has mastered and exploited. This study uncovers the intimate correspondences between memory encoding and narrative construction, and highlights the essential role that our shared language plays in that process. By demonstrating how I transmit mental representations of previous episodes to others through communication, this study lays the groundwork for future research on the interaction between memory, communication, and imagination in a natural setting.

Chapter 3: Memory update during naturalistic recall

3.1 Introduction

As humans, we are constantly barraged with new information and compelled to update our knowledge and understanding of the world so that we can make better decisions. It has been proposed that higher-order abstract representations of knowledge, called *schemas,* are created throughout life and influence the processing of incoming information as well as their retrieval (Bartlett 1932, Piaget 1952). Schemas are dynamic and can be updated to *assimilate* of new congruent information into the existing knowledge, or reshaped to *accommodate* new incongruent information (Piaget 1952). Schemas are now defined broadly as abstract frameworks of knowledge that could influence the later processing, which include different forms of knowledge such as gist of narratives, concepts, categories, scripts, perspectives, and even arbitrary relationship between items or locations (Gilboa & Marlatte, 2017).

Memories are vulnerable to modifications and alter in a way that reflects changes in the world (Bartlett 1932). The effects of prior knowledge on encoding has been widely studied using different paradigms (Alba & Hasher, 1983; Bonasia et al., 2018; McKenzie et al., 2014; Poppenk et al., 2010; Tse et al., 2011; van Kesteren et al., 2012). By using schemas based on naturalistic stimuli and differentiating perspectives, a recent fMRI study showed that prior knowledge affected the neural processing of two identical audio narratives. Yeshurun and colleauges successfully decoded these distinct perspectives through neural data gathered from the DMN, which revealed the effect of prior knowledge on transforming the brain responses of the same stimulus during encoding (Yeshurun et al., 2017). It is also known that schematic information effects the

reorganization of the neural responses even after the initial encoding (Hupbach et al., 2015). It has been shown that memories compatible with a previously learned framework of knowledge become hippocampally independent much faster than incongruent memories (Tse et al., 2007).

Schemas themselves may also be dynamic representations (Ghosh & Gilboa, 2014). During the consolidation through the time of recall, the schemas or the structures of knowledge in the world may change as well. For instance, one might encode an episode according to a certain schema and later on update their perspective and realize that the old schema is incorrect. Memory as an adaptive reconstructive process should be able to update the old representations in light of the new information and this is often reflected by changing the recall accordingly (Bartlett 1932; D. L. Schacter, Norman, & Koutstaal, 1998).

Whether participants can update their recall adaptively, was tested by Anderson and Picheret in a behavioral perspective manipulation experiment. Participants in this experiment were instructed to adopt the mind-set of a prospective home-buyer or of a prospective burglar prior to reading a story about the life of a family in their house. The description of the home included items that could be considered very important for someone in the market for a home and other items that mattered more for a thief. All participants recalled the story after the encoding. Then they were asked to take a new perspective (switch roles) and redo the recall without re-reading the story. They reported that most of the participants that switched roles, unlike the group that didn't take a new perspective, were able to recall more in their second recall and report items related to the new perspective that were not included in their initial recall. This finding suggested that participants successfully updated their recall. It also revealed that participants have encoded the schema

irrelevant information and have access to them during the recall, if needed (Anderson & Picheret 1978).

One of the primary remaining questions is how the behaviorally observed phenomenon of memory update is represented in the brain. My goal was to tap into this question using a naturalistic set up, with an emphasis on the adaptive reconstructive nature of memory (Figure 3.1). Participants (N=21) watched a movie which contained a dramatic twist in the final scene (post viewing exposure to the twist) while being recorded in an fMRI scanner. The twist information drastically changed the interpretation of the previously encoded movie. Two control groups (N=21 each) were recruited, one group ("spoiled" condition) watched a version that gave the viewer immediate exposure to the movie's plot twist, whilst the other group ("no-twist" condition) never saw the ending twist scene. The experimental group viewed the movie as it was designed, with the plot twist at the end ("twist" condition). I then instructed participants to recall each scene in the movie out loud in their own words while still in the fMRI. I asked the participants to recall each scene of the movie in a way that reflects their current and most accurate (updated) information about the movie. That way I could see if the participants in the experimental twist condition could adaptively update their recall to reflect their current knowledge of the twist into scenes that had been originally viewed and encoded with no knowledge of the plot twist. Most participants (17 out of 21 in the twist condition) talked about the new interpretation when describing the encoded events of the movie. At a neural level, I evaluated whether the twist knowledge can transform the memory representations and make them distinct from the representation during encoding. It was shown that the knowledge of the twist makes the brain representations dissimilar to the encoding or recall representations of the participants naïve to the twist. I also asked whether the reactivation of brain representations during the recall is affected by updating memory (using twist information in recall

behavior). It was observed that the similarity between the patterns of brain response during encoding and recall was attenuated in participants who better updated their recall to accommodate the new information (negative correlation). These results suggest that update during retrieval has changed the event representations and made them dissimilar to the representation of the previously encoded events.

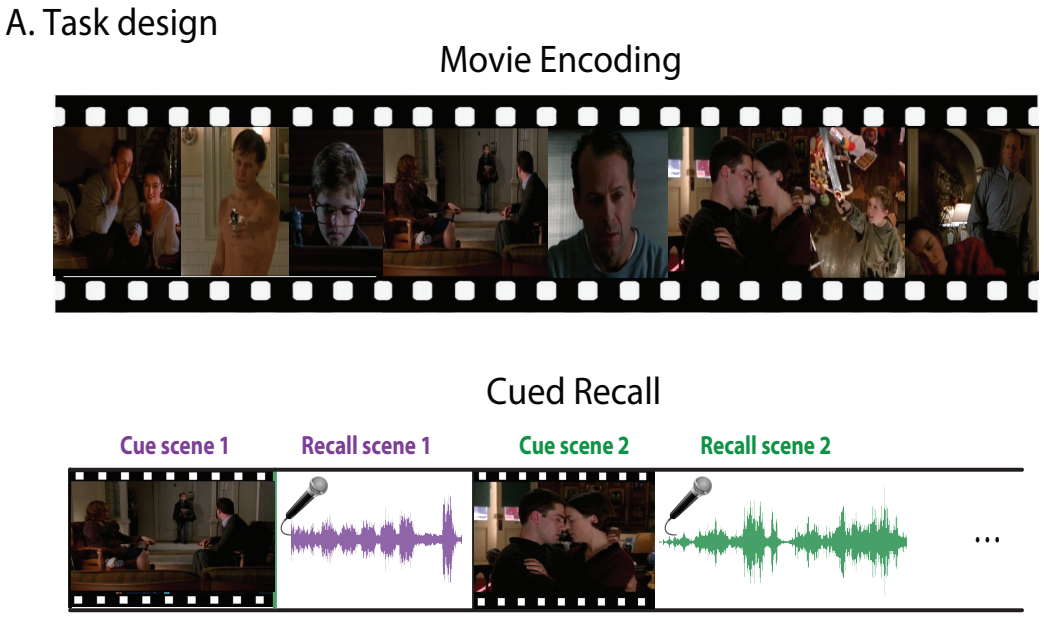


Figure 3.1: Depiction of experimental design and study groups. A) The task consisted of watching a 60 min version of The Sixth Sense. After the encoding, a cued recall task was performed in the scanner, which included the recall of 18 main scenes of the movie. B) Sixty-three participants divided in three completed the task. The main group (*twist*), watched the movie including the twist scene at the end and performed the recall. The first control group (*spoiled*) was informed about the twist information at the beginning of the movie. The third group (*no-twist*), never watched the ending and were naïve to the twist information. The control groups performed the recall task in the similar way as the *twist* group. Conditions with red frame are naïve to the twist information and conditions with green frame are aware of the twist information.

3.2 Material and Methods

To study the neural underlying of the memory update, I took two main paths. First, I compared the neural data when encoding and recall runs are influenced by the twist information vs. when there is no twist information. I use a classification method, support vector machine algorithms, to see if brain representations become dissimilar due to the effect of the twist information (Cortes & Vapnik, 1995). Second, to check if the reinstatement of the neural representations during the recall is influenced by the extent to which people update their memory, I use representational pattern similarity (Kriegeskorte et al., 2008), as well as a linear mixed model. The process of applying these methods of analysis, as well as the details about the stimuli and paradigm, are further explained in this section.

Stimuli

In selecting a stimulus, my goal was to have an entertaining narrative with a big reveal (twist) at the end, which would then have a significant effect on all of the previous scenes in the movie. I therefore used "The Sixth Sense" (M. Night Shyamalan, 1999) as the stimulus. In brief, "The Sixth Sense" depicts the story of a (presumably live human) psychologist treating a young boy who claims that he can see and speak with dead people. At the film's end, it is revealed that the psychologist was one of the ghosts the boy could see all along.

Three different versions of the stimuli were created. In the "twist" version, subjects watched a 60-min shortened version of "The Sixth Sense," with the twist appearing at the end of the film. In the "spoiled" version, subjects first read a statement describing the movie's twist, and then watched the same shorted version of the film. Finally, in the "no-twist" version, the ending scene which revealed the twist was removed. For fairness, I told subjects in the no-twist condition about the twist in the movie at the end of the experiment after they had exited the scanner.

All participants reported the movie as engaging in a post-scan questionnaire. Eighteen short clips from the beginning of the targeted movie scenes were sampled from the movie and were presented during the recall task as cues, and participants recalled the corresponding scene after the movie cue. In some of the scenes there were direct hints that the psychologist was not alive, for example, no one other than the kid ever talked to him in the movie. These hints were presumably discrete enough to not give away the twist. Furthermore, this lent itself especially well to the twist condition recall where subjects could solve out these hints.

Participants

Sixty-six right-handed, native-English speakers (ages 18-24, average = 20, 21 males) participated in the experiment. All participants were right-handed, with normal or corrected to normal vision. None of the participants had previously seen The Sixth Sense in full or in part (subjects were screened through an online questionnaire before coming into the lab). However, because the movie is well-known and frequently referenced in popular culture, participants with minimal knowledge about the twist (e.g. knowing that this is a movie about ghosts and the main character is actually dead) were admitted to participate in the spoiled condition (first control group) in order to facilitate data collection. In the post-scan questionnaire, two participants reported guessing the twist in the middle of watching the movie and their data was excluded. One participant did not understand the twist even after watching the final scene and their data was omitted as well. Four participants in the twist group and two participants in the spoiled group did not mention the twist or any interpretation related to the twist in their recall and did not follow the instruction to talk about the new interpretation of the story while doing the recall task. To isolate the neural representations related to memory recall, data of subjects who did not update their recall was dropped from the analysis. In addition, six participants with the poorest overall recall performance (recall score at least 1 SD below average) were excluded. The data of the remaining fifty-one participants (17 per group) were used in most of the analyses. For the analyses, where including the participants who did not alter their recall was required, the full dataset (N=63 total, N=21 per group) was used. This will be further explained in the relevant sections of methods and results. All participants handed in written informed consent prior to the experiment and received information about the conditions of the experiment and their rights. The experiment protocol and the consent forms were approved by Institutional Review Board of Princeton University.

Procedure

Experimental Design

Participants were pseudo-randomly divided into three groups: the main group (twist condition, N=21) watched a 60-min audio-visual edition of The Sixth Sense movie, including the twist at the end while undergoing fMRI scanning (Figure 3.1B). The first control group (spoiled condition, N=21), watched the same movie version in the scanner, but the movie twist was revealed in text that appeared on the screen after the first movie scene (Figure 3.1B). The second control group (no-twist condition, N=21) watched a 55 min version of the movie with no twist scene whatsoever. (Figure 3.1B).

Prior to the stimulus, participants were instructed to watch the movie attentively and were told there would be a task related to the movie content after the movie. The details of the recall task were not revealed before watching the movie as to not change movie-watching behavior, but subjects were assured that the task is easy, and that they should watch the movie as they usually would and do not need to try to memorize the movie. An anatomical scan was performed at the beginning of the scan session before watching the movie. After the movie, participants performed a cued recall task. During the cued recall task, subjects watched a short clip from different scenes in the movie to cue recall of the scene. They were instructed to freely describe the events of the scene, after the cue clip, to the end of the scene, using all information they gathered throughout the movie. The instructions were identical for all the three groups. During recall, participants watched the scene cues, waited for a count down on the screen (14 seconds long), and then recalled the appropriate scene. The recall task was self-paced, and subjects indicated when they had

finished describing each scene by pressing a key on a button box. Participants were instructed to stop the recall where each scene ended but were not given specific instructions on what constitutes "a full scene," allowing subjects to segment the film events individually.

After scanning, participants filled out a questionnaire about their experience in the scanner, including information about the movie and recall tasks. They were asked if they guessed the twist in the middle of the movie and if yes in which scene. All participants scored the movie as engaging.

Scanning Procedure

Participants were instructed to lie down comfortably on the scanner bed (in an effort to minimize their head motion), to watch the movie attentively, and to carefully follow the instructions for the post-movie task. Participants watched the movie and the instructions through a mirror mounted to the head coil which reflected a rear screen. The main screen was located at the back of the magnet bore and the movie was projected on the screen via an LCD projector. MR-safe, in-ear headphones were used for the movie audio. Eye-tracking was set up using to monitor the participants during the scans and ensure they are awake and attentive during the tasks. The movie and recall instruction were presented using Psychophysics Toolbox [http://psychtoolbox.org] in MATLAB (Mathworks), which enabled coordinating the onset of the stimuli (movie and recall cues) and data acquisition. Sound level of the movie was adjusted separately for each participant to assure a clear and comfortable audio signal. Recall speech was recorded during the fMRI scan using a customized MR-compatible recording system (FOMRI II; Optoacoustics Ltd). The MR recording system used 2 orthogonally-oriented optical microphones. The reference microphone captures the background noise, and the source microphone captures both background noise and the speaker's speech (signal). A dual-adaptive filter subtracted the reference input from the source channel using a least mean square approach. To achieve an optimal subtraction, the reference signal was

adaptively filtered so the filter gains are learned continuously from the residual signal and the reference input. To prevent divergence of the filter when speech is present, a voice activity detector was integrated into the algorithm. A speech enhancement spectral filtering algorithm further preprocessed the speech output to achieve a real-time speech enhancement. Audio recordings were further cleaned using noise removal software (Adobe Audition). The output recall recordings were fully comprehensible. A response box was used to collect the participants' commands during the recall task. Participants were instructed to press a button when they finished the recall of a scene to continue the task.

MRI Acquisition

MRI data was collected on a 3T full-body scanner (Siemens Prisma) with a 64-channel head coil. Functional images were acquired using an interleaved multiband EPI sequence (TR= 1500 ms, TE 33 ms, flip angle 80 degrees , whole-brain coverage slices of 2 mm thickness, FOV 192 mm^2, iPAT = 4). Anatomical images were acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) pulse sequence (1 mm^3 resolution). Anatomical images were acquired in a 6-min scan before the functional scans with no stimulus on the screen. Field maps were collected for B0 correction at the end of the recall run.

Data Analysis

Preprocessing

Preprocessing was performed using fMRIprep package (Esteban et al., 2019), including anatomy alignment, slice time correction, motion parameter estimation, B0 correction, co-registration, and transformation of the functional volumes to a template brain (MNI). Then, the datasets were

smoothed using Analysis of Functional NeuroImages – AFNI (Cox, 1996) – 3dBlurToFWHM to reach 7mm real smoothing (2mm default smoothing in the raw image taken from the scanner + 5 mm smoothing applied). Using AFNI, nuisance regressors were removed from the data which included head motion (12 degrees of freedom motion estimates), low-order drift polymonials (order of 2), and high-pass filtering (140 second cutoff). De-spiking the timecourses and all the subsequent analyses were conducted using custom MATLAB (Mathworks) scripts. Each run in the movie was z-scored across time (mean= 0, SD=1). In the recall data, the signal intensity of movie cues and recall were dramatically different. Thus, the recall sections were extracted from the timecourse (including, movie cues, countdown, and recall data) and z-scored after concatenation. All the analysis was performed in volume space. The results were projected on the surface for illustration using NeuroElf (http://neuroelf.net).

Masks and ROI definitions

Whole brain ROI analysis was performed on a set of 61 functionally defined ROIs distributed in 6 networks (visual, attention, auditory/language, executive, DMN |, and DMN ||) adapted from a separate dataset (Regev et al., 2018). In that study, clustering was performed using intrinsic connectivity networks (ICNs) and only voxels that reliably responded to the audiovisual story were included in the parcellation. Because the original data was collected at 3-mm resolution, ROIs were resampled to 2-mm resolution using FSL. Ultimately, the set of 61 ROIs covered 76695 cortical voxels that is depicted as non-shadowed regions in Figures and . Hippocampus ROI was adapted from Harvard-Oxford anatomical atlas (distributed with the FSL software package; (Esteban et al., 2019)).

Rating of Recall Behavior

Audio recordings during the recall task were comprehensible even without denoising; however, the recordings were further denoised using Adobe Audition to decrease the scanner noise and facilitate transcription. Following manual transcription of the recordings, two independent raters naïve to the purpose of the experiment rated the quality of each scene recall from 1-5 (1=extremely poor, 5=extremely good; 0 if not recalled), as well as whether or not the person recalling the scene knew the movie twist (1=knew twist, 2=did not know twist). Raters watched the movie before rating recalls. Recalls per scene for each subject were presented individually and in random order to prevent raters from inferring subject condition or from making inferences across scene recalls. For judging the quality of recall, raters were given minimal instructions and were allowed to choose and report their own rating criteria. Elements such as dialogues between characters, amount of details provided, and keywords were among the reported criteria. For rating whether subjects knew the twist, raters could use any element in the recall speech. This process resulted in a "recall-score" and a "twist-score" for each scene. These scores were summed up in each individual to create overall recall scores and twist score for each participant. Across all participants, seven scenes were rated as high on twist-score (scenes in which many participants talked about the twist and the new interpretation) which were later used as "critical scenes" in the classification analysis. Critical scenes were expected to show the highest amount of dissimilarity between the twist and no-twist conditions in the brain, as they contained the highest amount of twist relevant information.

Classification analysis

The aim of this analysis was to directly measure whether our manipulation (twist) influences the brain patterns and makes them distinct from the patterns without the twist.

First neural responses during critical scenes between the spoiled condition and the twist condition were compared. These critical scenes were previously identified using behavior (recalled scenes with an above average twist-score for all scenes). This allowed me to better isolate the representations related to the research question. The question was whether the plot twist information changes the corresponding brain representation. To test this the brain representations related to the no-twist condition needed to be compared with representations of the twist condition. To classify the movie data across conditions, timecourses of brain response related to critical scenes could be directly compared. However, this method could not be applied for the recall data since the recall data of each participants has unique length and content. To keep the analysis method identical for movie and recall comparisons, timecourses of brain response in each critical scene were averaged across time, resulting in one pattern of brain response per critical scene in each ROI. Then the pattern for each scene was transposed and concatenated with the other transposed scene patterns. This procedure created a vector of features for each ROI that contained spatial patterns of all scenes and could be used in classification (Figure 3.2A). Importantly, the within-scene averaging procedure resulted in equal size recall patterns in all participants in both movie and recall conditions.

The same method, movie and recall data were prepared for spatial classification as well. To classify any two groups, SVM classification (linear kernel) was performed on their corresponding vectors. A 17-fold classification was conducted by leaving out one participant from each group at each fold, training the classifier on the remaining 16 pairs, and testing the left-out pair. Repeating this procedure for all the pairs and averaging the results resulted in a classification accuracy for a given ROI.

Significance for all classification analyses was computed by shuffling the group labels in a way that each new groups contains a balanced half of each real groups. 1000 iterations of shuffling labels and a same 17 fold classification was performed and null distribution was created. Individual *P* values were generated for each comparison in each ROI by measuring the number of values on the null distribution that exceed the real classification accuracy. In whole brain analysis (61 ROIs), p values were corrected for multiple comparisons using false discovery rate (FDR; Benjamini and Hochberg 1995) at $q = 0.05$.

SVM classification was applied on movie-spoiled vs. movie-twist, movie-no-twist vs. movie-twist, recall-spoiled vs. recall-twist, and recall-no-twist vs. recall-twist (more in results section). Movie-no-twist vs. movie-spoiled and recall-no-twist vs. recall-spoiled were not included in the analysis as the comparisons would be redundant to twist vs. spoiled classifications.

To evaluate a possible interaction between conditions, the difference between classification accuracies was measured as:

"conditions difference" = (movie-spoiled vs. movie-twist) – (movie-no-twist vs. movie-twist) – [(recall-spoiled vs. recall-twist) – (recall-spoiled vs. recall-twist)]

This subtraction was applied to the output of the 1000 iterations of classifying the shuffled group labels which created a distribution (1000 values) of null "conditions-difference". Measuring the proportion of the values on this distribution that were bigger than the actual conditions-difference resulted on a significance measure (*p*-value) for the interaction.

Pattern Similarity Analysis

Linear classification was used to measure whether there are discriminable patterns in brain activity across all scenes as a function of subject's interpretation (twist, no-twist). To evaluate the relationship of the individual scenes in the movie and recall, RSA was performed. Therefore, pattern similarity analysis was implemented to compare the scene specific patterns of brain activity across different conditions (Figure 3.2B). Movie data was timestamped based on 18 scenes, which were defined according to the major shifts in the narrative (e.g. location, topic, time; as defined by an independent rater). The same scenes were recalled during the cued-recall paradigm. The recall data was manually timestamped based on the beginning and end of the speech for each scene using the recorded speech audio files. Calculating the arithmetic mean of time points within each scene provided a single pattern of brain response for each scene during movie viewing. The Pearson correlation between patterns of brain responses for movie scenes vs. recall scenes was computed and the average of correlation values for matching scenes created the representation similarity value. Statistical significance was only reached if matching scenes in the movie and recall could be differentiated from nonmatching scenes. Significance thresholds were calculated using a permutation test (Kriegeskorte et al. 2008), shuffling the scene labels and calculating random scene-to-scene correlations for each participant, then averaging across participants to create a null distribution of r values; the *P*-value was calculated as the proportion of the null distribution that was above the observed level of pattern similarity for matching scenes. Pattern similarity analysis was performed on all the participants (not just the ones with twist scores > 0) and all scenes (not just the seven critical twist scenes), which will be further explained in the results section. This procedure was implemented for all ROI pattern similarity analyses.

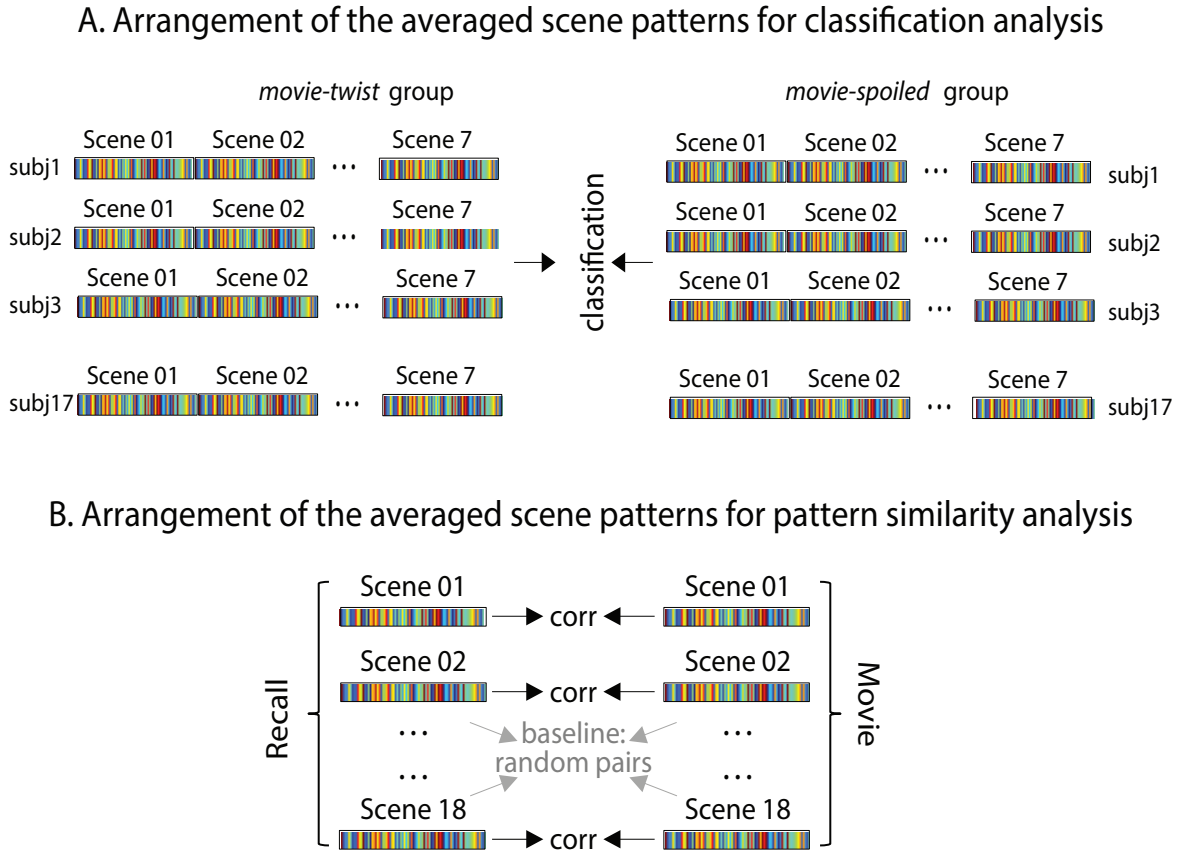


Figure 3.2: Arrangement of data for pattern analysis A) spatiotemporal pattern classification was performed by averaging timecourses of brain response of each critical scene (N=7) in a given ROI and creating a single vector of spatioemporal patterns for each participant. *movie-twist* and *movie-spoiled* conditions are given as examples. Classification across all conditions was performed using the same method. B) Pattern similarity analysis was performed by averaging timecourses of brain response of all scenes (N=18) in a given ROI and correlating matching scenes.

3.3 Results

The analysis served three main questions. First, does knowing the twist change how the movie is encoded? More specifically, were the brain responses of subjects who knew the twist before watching the movie (spoiler, twist first) discriminable from those who don't know the twist during movie-watching (no-twist vs. twist)? Second, does realizing the twist at the end of the movie affect the subsequent recall behavior and the corresponding brain representations? In other words, does "getting" the twist make the recall representations different from those of participants who never saw the twist scene (twist vs. no-twist)? Third, what is the particular effect of twist knowledge on reinstatement? Does it update the representations? After learning the twist, are the reinstated representations of the movie similar to those of subjects for whom the twist was spoiled early on? To tackle these questions, intersubject classification and pattern similarity analyses were used to relate response patterns across twist conditions (and encoding–recall), along with the behavioral data.

Recall behavior

Participants generally completed the recall task without challenges. Although they were not given explicit instructions about where to terminate recall for each scene, their recollections only intruded into the next scene in very few scenes. There were also a very few instances where recall was terminated earlier than the end of the scene as defined by the experimenter, or context from the previous scene was recalled before starting to recall the current scene. In each of these situations, recall timestamps were shifted appropriately to isolate the recall of the current scene as accurately as possible. The number of overall incorrect or forgotten scenes was: 29 out of 378, (7%) in recall-spoiled, 12 (6%) in recall-twist, and 26 (3%) in recall-no-twist; (N=63). After

removing the participants who did not alter their recall according to the twist and the ones with poor recall behavior (more than one SD below average), the number of correct scenes was 288 (94%), 296 (96%), and 295 (96%) in the recall-spoiled, recall-twist, and recall-no-twist respectively of the remaining 51 participants. The recall instructions were kept similar for all the groups to capture the natural behavior of recall update and to minimize any differences between the groups other than the twist information. As hypothesized, most participants were able to incorporate the twist information in their recall, in at least one scene. Participants who did not update their recalls (6 out of 42 participants in spoiled and twist groups) claimed after the scan that they had understood the twist and were aware of its effect on the interpretation. However, they decided to not talk about the twist since they misinterpreted the instructions as "talk about the original movie". The data of these participants was removed for classification analysis since it could not be confirmed that they had actually thought about the new interpretation when they were talking about the original interpretation of the scenes. For the reinstatement analysis (section 3.3.5), however, it was desired to show the relationship between the entire spectrum of behavior (including the participants who did not update their recall) and the brain encoding-recall pattern similarity and therefore, all the participants were included.

To evaluate the quality of the participants' recall, and the degree to which their recall reflected the twist, two independent raters evaluated the verbal recalls. Raters reported a recall-score and a twist-score for each scene. Raters were naïve as to which participant or group each recalled scene belonged to (see methods). Although the rating criteria were somewhat subjective to each rater, the two ratings for recall-scores were highly correlated (88% Pearson correlation, p-value 1.7×10^{-10}). The twist-scores were very similar between the two raters as well (93% Pearson correlation, p-value 9.7×10^{-29}). Recall-scores and twist-scores were averaged across the two raters

to create final scores. The final recall-score and twist-scores in all participants are depicted in Figure 3.3. Recall-scores and twist-scores were not significantly correlated to each other (7% Pearson correlation, p-value 0.56).

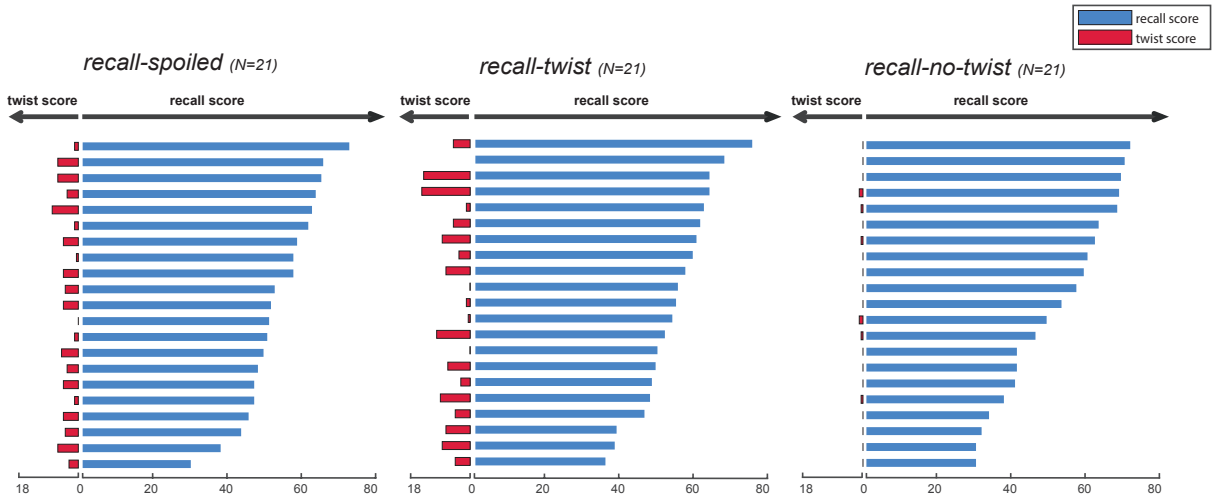


Figure 3.3: Summary of behavior in experimental groups. Each bar depicts the behavioral scores for one subject. Subjects in each group are sorted based on the recall scores. Blue bars indicate the recall score and red bard depict the twist scores. As evident by blue and red bars, recall-scores and twist-scores were not correlated (7% Pearson correlation, p-value 0.56).

Contrary to my initial prediction, participants did not talk solely about the updated interpretation based on the twist in their recall. Most of the participants in the twist condition (17 out of 21) talked about both the new interpretation as well as the older interpretation of the scene. For example, a typical recall of the "restaurant scene" would consist of explaining the situation and what was happening according to their original interpretation, e.g. "Malcom enters the restaurant, his wife ignores him. It seems like she's angry because he is late for their anniversary dinner." And then they would proceed to explain how the twist changed this interpretation "…but now we know that he was actually dead so she couldn't see him." While this recall behavior

captures an interesting natural aspect of the memory updating process, it suggests that the neural representations supporting recall will include both original and updated interpretations. To better differentiate neural representations of the fully updated interpretation, one would have had to design the recall task so as to isolate the updated content from the original encoded content. Due to the presence of both the original and updated content, it is expected that decoding the update-specific recall representations will be challenging (further addressed in the Discussion section).

Effect of the twist on movie viewing

As we have discussed, previous research has shown that perceiving the same stimulus after receiving different context information leads to distinguishable brain representations (Yeshurun et al., 2017). In the context of this study, knowing vs. not-knowing the twist could be considered as distinct contexts and thus I asked whether the knowledge of the twist before viewing alters the brain representations during encoding. Participants began the movie in the same way, however, after the initial scene, the spoiled condition group received a passage of text on the screen revealing the twist (thus spoiling) the movie. After the reveal, the spoiled group watched the sections of the movie that were related to the twist with different prior information relative to twist and no-twist condition groups. I hypothesized that the twist: knowing that the main character was a ghost for the duration of the movie rather than thinking he was alive and living his normal life, would dramatically change the interpretation of some scenes in the movie. I further hypothesized that the difference in story perception would be reflected in participants' brain representations during encoding. This would build upon previous related research in two ways. First, spatiotemporal classification instead of temporal classification was implemented. Rather than taking the brain data across all time points in each voxel as features for the classification, the timepoints within scenes were averaged to create spatially-distributed response patterns across voxels in each ROI. I then

concatenated the patterns for different scenes to create a single vector that included the spatial information of all scenes across time. Linear SVM classifiers were applied (see methods) to evaluate the separability of response patterns for the two conditions. This analysis was intended to replicate the previous findings (Yeshurun et al., 2017) in the spatial domain. In addition, this method can accommodate stimuli with different lengths which have the same event structure (Chen et al., 2017), thus allowing me to apply it to the recall data and directly compare the effect of the plot twist on movie and recall data. Secondly, I introduced a control comparison and asked whether the movie data for the twist group could be differentiated from the no-twist group.

I first conducted the analysis at the ROI level. Precuneus, angular gyrus (AG), medial prefrontal cortex, and hippocampal ROIs were selected as targets based on previous studies highlighting their role in the DMN, and in memory encoding and retrieval, as well as perspective taking (Addis, Barense, & Duarte, 2015; R. Buckner et al., 2008; Chen et al., 2017; Rugg & Vilberg, 2013; Yeshurun et al., 2017). Early auditory cortex was used as a control ROI since it is hypothesized that differences between patterns due to twist information will emerge at areas with long integration timescales, which is needed for the integration of the contextual information, and thus will not be apparent at sensory areas, as early auditory cortex, with short processing timescale. As depicted in the bars on the left side of Figure 3.4, spoiled and twist conditions were classifiable in DMN ROIs (except for hippocampus) but not auditory cortex ROI, with the classification accuracies of precuneus acc=73%. P=0.01, mPFC acc=0.73% p=0.008, ANG acc=71%, p=0.01. P-values were calculated by a permutation test shuffling group belongings. Moreover, twist and no-twist conditions appear indistinguishable in precuneus (46%, p=0.61), mPFC (55%, p=0.26), but classifiable in AG (68%, p=0.03). The same pattern of results was observed in the whole brain ROI analysis. Figure 3.5 depicts the classification conditions with p-value lower than 0.05 obtained

using a permutation test in which group belongings were shuffled to create the null distribution. Areas colored as yellow pass correction for multiple comparisons (FDR). As shown in the Figure 3.5, in spoiled vs. twist comparison, areas such as precuneus, medial frontal regions, angular gyrus, anterior temporal gyrus, and some parts of visual cortex appear to respond differently in spoiled and twist conditions. Differentiation of patterns of brain response in the precuneus, angular gyrus and mPFC implies the role of twist knowledge in transformation of representations during encoding and is in agreement with previous literature (Euston, Gruber, & McNaughton, 2012; Ghosh et al., 2014; Spalding, Jones, Duff, Tranel, & Warren, 2015; Tse et al., 2011; Yeshurun et al., 2017).

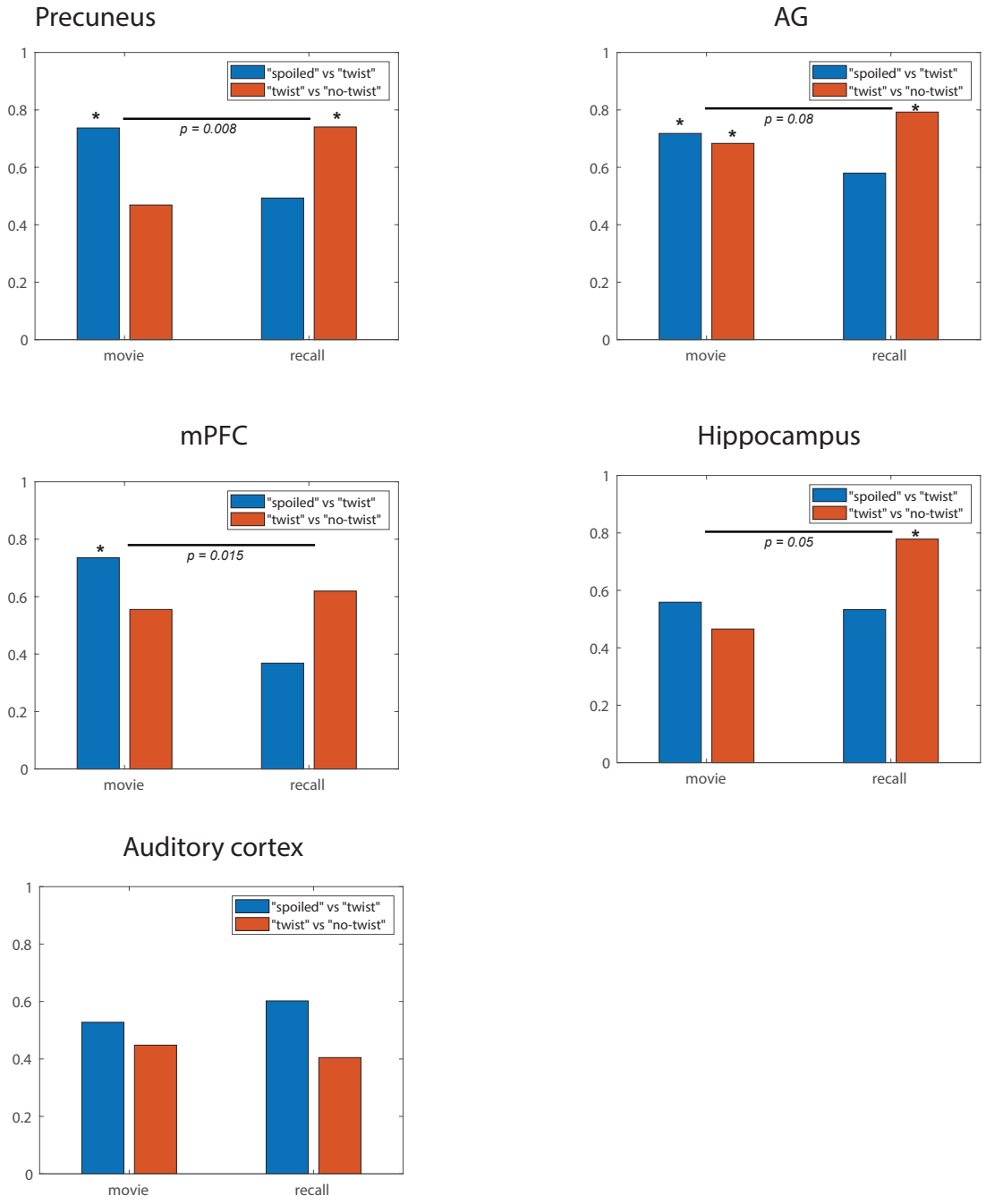


Figure 3.4: ROI classification analysis in 5 regions. Chance level is 0.5. Real chance was calculated using a permutation test for each comparison separately. Comparisons that passed significance test are marked by asterisk.

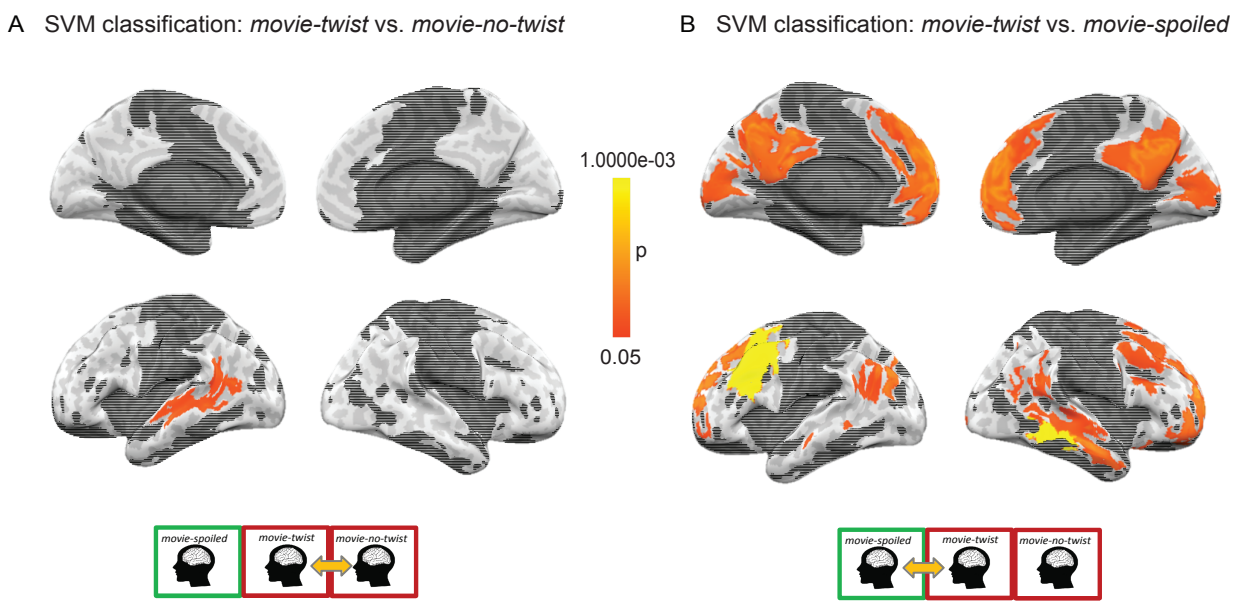


Figure 3.5: Classification of movie representations across 61 ROIs. Shaded areas were not included in the analysis (no functionally defined ROIs in those regions). Maps have been threshold at p value < 0.5 (obtained from permutation analysis). Yellow areas pass FDR correction. Green-framed icons depict the conditions that participants are aware of the twist. Red-framed icons depict the conditions that subjects are naïve to the twist.

Effect of the twist on recall representations

Next, I tested whether learning about the twist after-viewing the movie in the twist group altered the neural responses during recall. The analysis of the recall transcript assured me that behaviorally most participants in the twist condition updated their recall for at least some of the scenes. I wanted to see whether this memory update was reflected in the brain during recall. In previous analyses, I showed that response patterns for participants in the twist group were different from those in the spoiled group while watching the movie. After watching the final plot twisting scene and during the recall, however, the twist condition group and the spoiled group had similar knowledge about

the movie, including its plot twist. Thus, if memory update occurred, their recall representations should be similar, as they can both incorporate the twist into their recall and explain how it changes the story. With this in mind, I hypothesized that these response patterns in these groups would largely be indistinguishable. In agreement with such a prediction, I observed a nonsignificant classification in precuneus (49%, p=0.5), mPFC (36%, p=0.9), angular gyrus (57%, p=0.15), and hippocampus (53%, p=0.27) ROIs (Figure 3.4). On the other hand, the recall of the twist group might be differentiable from the no-twist group because participants in the latter group were never informed about the twist and their recall followed the original interpretation (average twist condition score ~0). As expected, response patterns for these two groups during recall were classifiable across DMN ROIs (except for mPFC) with accuracies of: precuneus (74%, p=0.003), mPFC (62%, p=0.11), angular gyrus (79%, p=0), hippocampus (78%, 0.003) (Figure 3.4). Likewise, the whole-brain ROI analysis revealed high classification accuracies in higher order regions in recall-twist vs. recall-no-twist but not recall-twist vs. recall-spoiled (Figure 3.6).

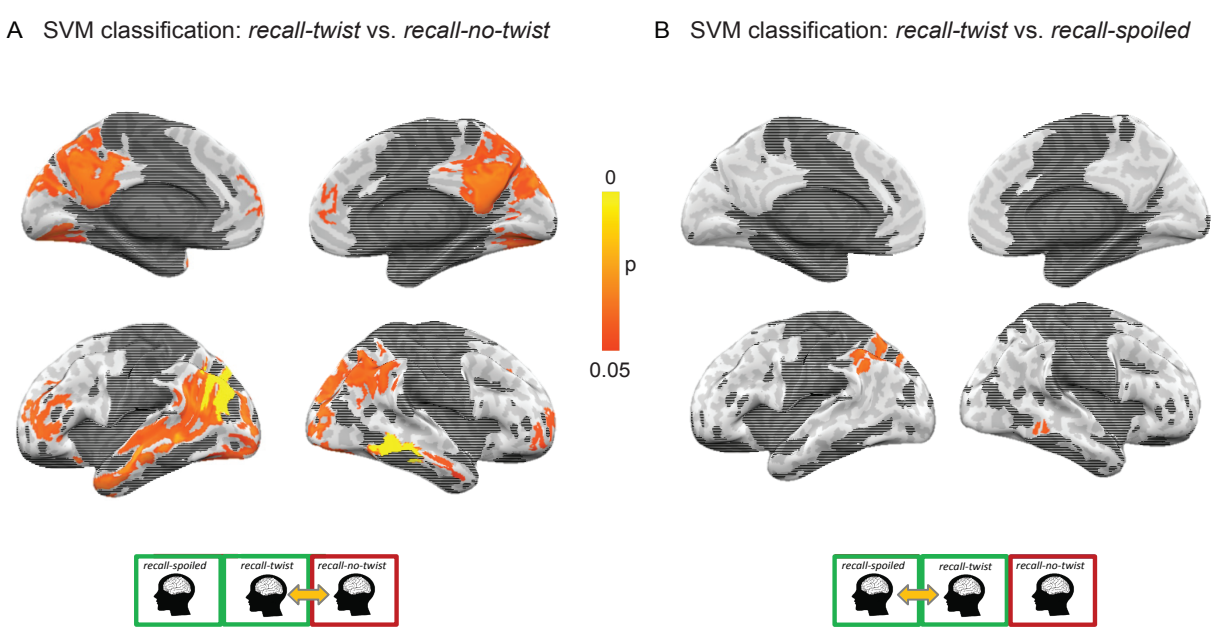


Figure 3.6: Classification of recall representations across 61 ROIs. Shaded areas were not included in the analysis (no functionally defined ROIs in those regions). Maps have been threshold at p value < 0.5 (obtained from permutation analysis). Yellow areas pass FDR correction. Green-framed icons depict the conditions that subjects are aware of the twist. Red-framed icons depict the conditions that subjects are naïve to the twist (note that twist condition, the icon in the middle, has changed color relative to figure 3.5)

Interaction between movie encoding and movie recall

While recall classification results suggest that plot twist information influences the brain representations, they cannot alone provide sufficient evidence supporting the statement that memories are being updated. For such a claim I would need to interpret the results based on the possible connection between the movie and recall data. For instance, if memory representations were modified after exposure to the twist, the recall representations of the twist group may be more

similar to the movie representations of the spoiled group rather than to their own (twist condition) group. Therefore, if the classifier is trained using the movie-spoiled and movie-twist data, then it could be tested whether each recall-twist belongs to movie-spoiled or movie-twist group. Based on the hypothesis those recall patterns should contain twist information, the recall-twist pattern may now be more similar to the movie-spoiled rather than their own movie-twist group. I was unable to reliably classify the recall-twist as similar to movie-spoiled and dissimilar to movie-twist (accuracy 42%). A possible explanation for this result could be the presence of competing *twist-related* and *twist-unrelated* representations in both movie and recall of the spoiled and twist conditions that make the classification more challenging (see Discussion).

However, looking further into the previous movie and recall classification results provides indirect evidence for a plot twist-mediated relationship between the patterns of movie and recall. In each ROI, I calculated the interaction of the movie and recall conditions by creating a difference index as followed:

"difference index" = (*movie-spoiled* vs. *movie-twist acc**) – (*movie-no-twist* vs. *movie-twist acc*) – [(*recall-spoiled* vs. *recall-twist acc*) – (*recall-no-twist* vs. *recall-twist acc*)]

* *acc = classification accuracy*

The difference value is an indirect measure of how the same patterns transformed across difference conditions in the hypothesized direction. A positive index implies that movie and recall patterns changed in the opposite direction. Movie-spoiled and movie-twist were distinct with regard to which resulted in a big positive value but recall-spoiled and recall-twist became more similar, under with the plot twist. Similarly, recall-no-twist vs. recall-twist become differentiable,

resulting in a large negative value. Subtracting this negative value from the movie patterns positive value yielded a positive index. Accordingly, a positive index was observed in all four DMN ROIs. I statistically assessed this effect by shuffling the group labels in a permutation test and subtracting the classification accuracies (see Methods). As Figure 3.4 shows, the interaction was significant in precuneus (p=0.005), mPFC (0.01), and hippocampus (0.05), and close to significant in angular gyrus (p=0.08), which could imply their possible involvement in the process of updating memory. Running the same analysis in all ROIs (Figure 3.7), significant interaction was observed in anterior and ventromedial prefrontal cortex (p=0.001, FDR corrected at q<0.05).

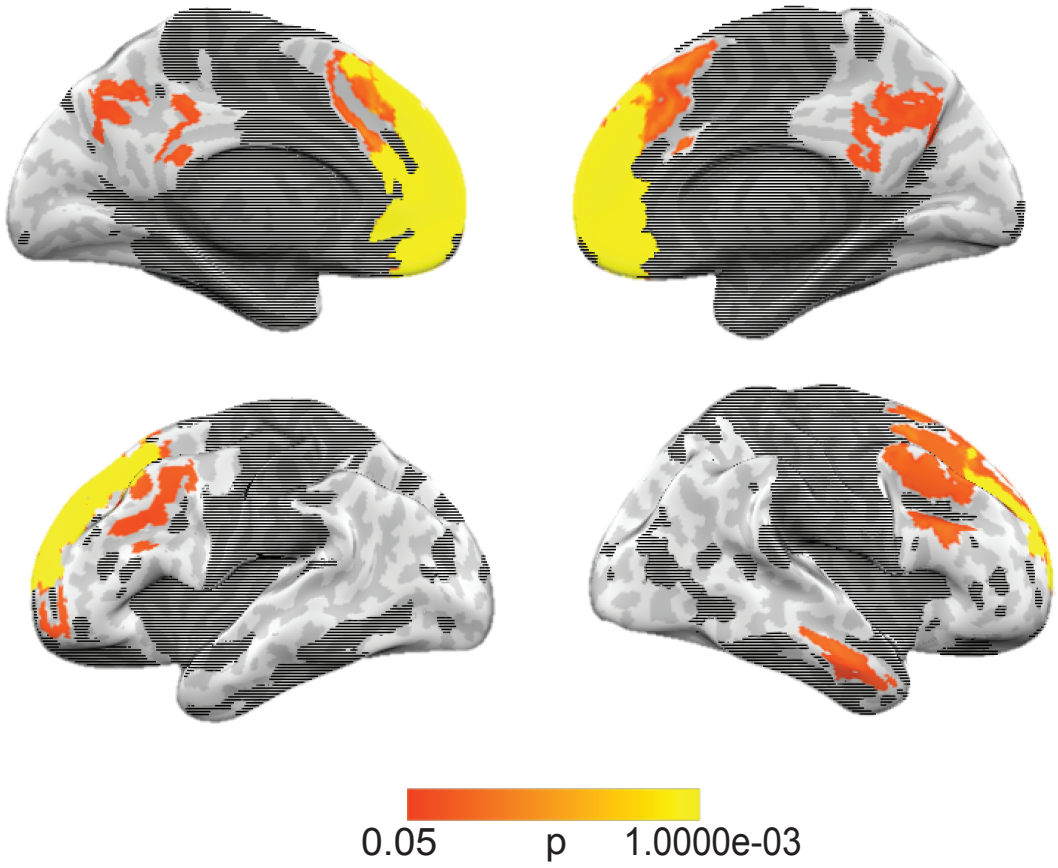


Figure 3.7: The interaction of movie classifications (Figure 3.5) and recall classifications (Figure 3.6). Map is thresholded at $p<0.05$ using a null bootstrap test. Yellow regions pass FDR correction ($q<0.05$).

Effect of the twist on memory reinstatement

To directly study the phenomenon of memory update, I looked for changes in response patterns during the recall of information in the groups that learned about the twist after viewing the movie (twist condition). I hypothesized that if the encoded representations of a scene are altered during memory update to incorporate twist information, the patterns which are reinstated during the recall should be less similar to their original encoding patterns. In other words, participants have encoded certain representation during the movie viewing; but they do not reinstate the exact same patterns during the recall, because the memories are being reconstructed in a different way to incorporate the critical new twist information. The plot twist forces viewers to retroactively reinterpret the movie. Reinstatement of encoded memory patterns during spoken recall has been measured using pattern similarity between encoding and recall (Chen et al., 2017); a higher similarity between encoding and recall patterns suggest a stronger reactivation of memory during recall. Altered reinstatement due to memory update should result in diminished encoding-to-recall pattern similarity for scenes requiring large updates due to the twist. In the same vein, I hypothesized that if a particular participant better incorporates the twist information into their recall, there will be a larger decrease in their encoding-to-recall pattern similarity.

Each participant received a twist-score which represented the number of scenes in which they talked about the plot twist or resulting new interpretation from the plot twist in their recall. If participants with high twist-score had lower encoding-to-recall pattern similarity, I would have expected to see a negative correlation between these two measures. This hypothesis was tested by correlating within-participant encoding-to-recall pattern similarities in participants of the twist condition group with their corresponding twist-scores. It was important to include the entire spectrum of behavior in the analysis when assessing the relationship between behavior and changes in the brain patterns. Therefore, the entire set of participants in the twist group was used, including the ones who did not alter their recall based on the twist. Likewise, this analysis was not restricted to the seven critical scenes because some of the non-critical scenes had positive twist scores (see Methods). Twist-scores ranged from 0 to 14 (mean ~ 6) and all scenes were included in the analysis.

As expected, a significant negative correlation between pattern similarity values and twist scores were observed in precuneus and angular gyrus, meanwhile there was no relationship between these two measures in the auditory cortex control (Figure 3.8). The interaction analysis using the classification values suggested a possible interaction of movie and result patterns due to the influence of the twist information. This is concurrent with the analysis based on the pattern similarity between encoding and recall and its relationship to behavior. Together, these results provide evidence that the patterns captured may be involved in the memory update process, since they change in relation to behavior.

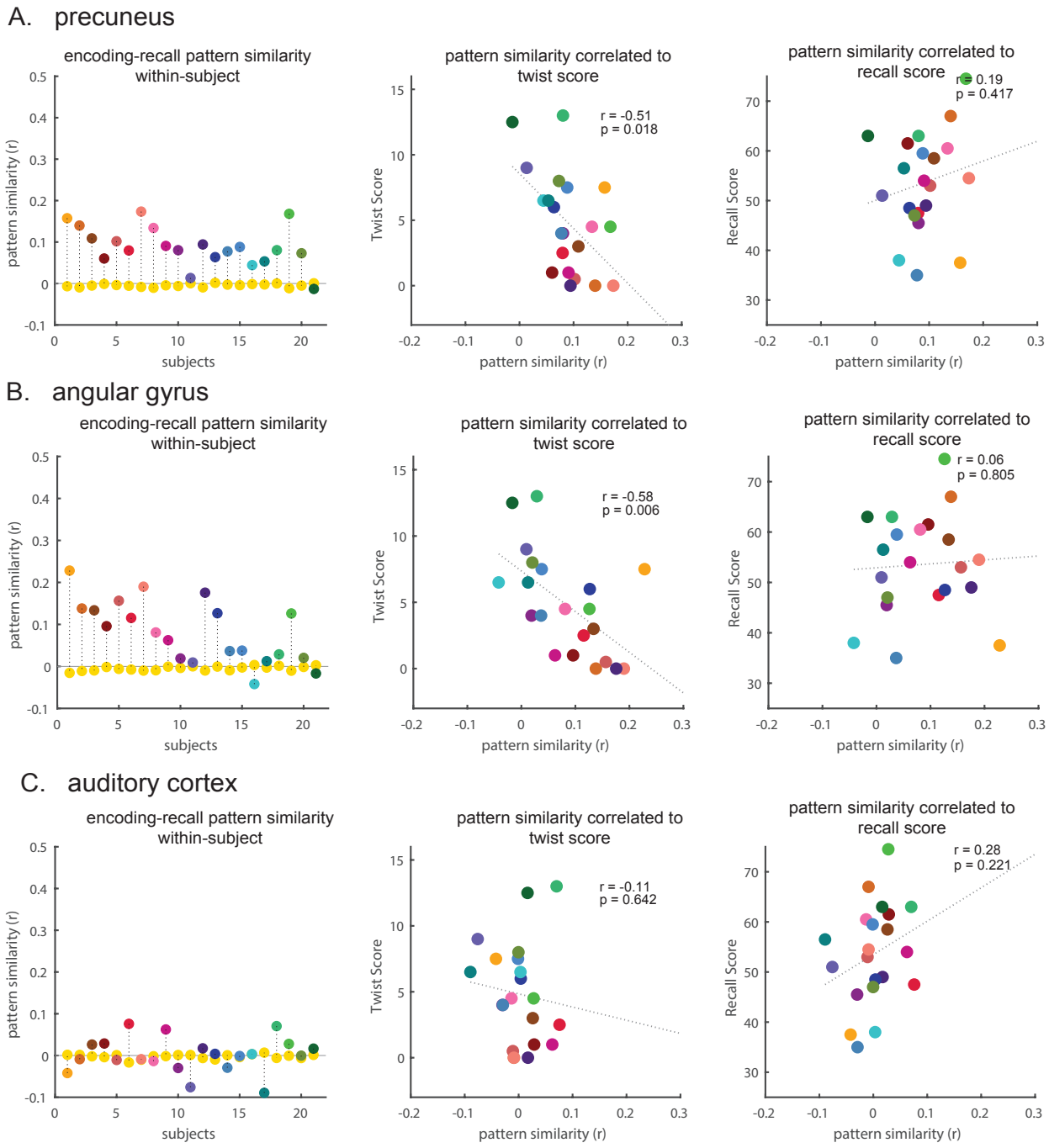


Figure 3.8: Correlation of *within* subject pattern similarity between encoding and recall to memory update behavior in *twist* group (twist score). Left panel depicts pattern similarity between matching scenes in each subject (each color), as opposed to the pattern similarity between non-matching scenes (yellow colors, baseline). The middle panels depict the relationship between encoding-recall pattern similarity of matching scenes in each subject with their corresponding twist scores.

The panels on the right depict the relationship between encoding-recall pattern similarity of matching scenes in each subject with their corresponding recall scores. The same color for each participant has been used across all panels.

Analysis was further extended to test whether the negative correlation between encoding-recall pattern similarity and twist scores is driven by subject-specific mutual movie-recall patterns or by more general across-subject similarity between the patterns of encoding and recall. That is to say, I questioned whether similar results would be observed if the recall representations are correlated with the average of patterns in the other participants, based on the negative correlation of twist score with encoding and recall pattern similarity *within* each subject was observed. I measured the similarity between the brain patterns of each participant in the recall-twist group with the average of patterns in the rest of the subjects in the movie-twist, and a similar negative correlation with behavior (twist-score) was observed (Figure 3.9 A)

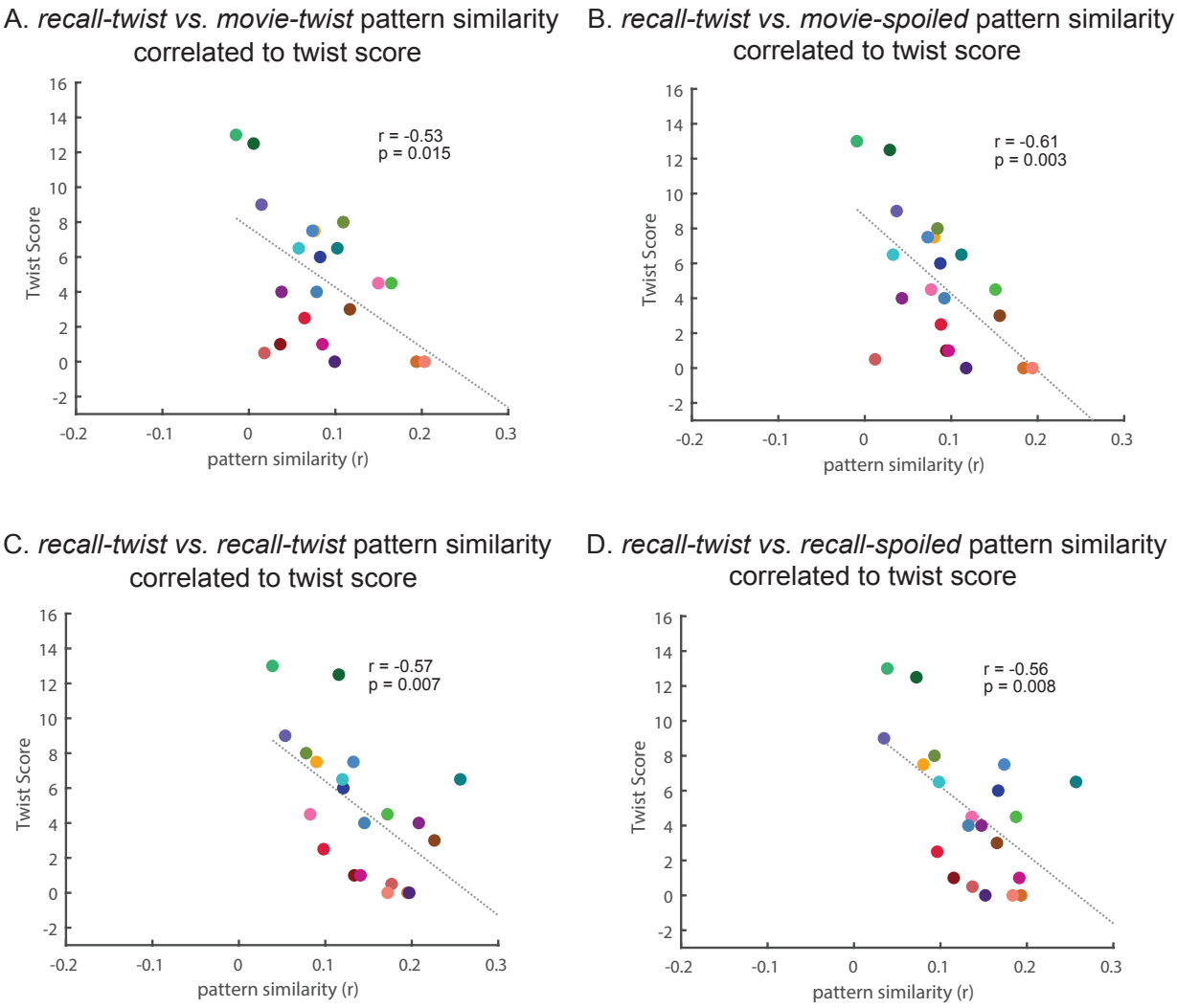


Figure 3.9: Correlation of *between* subject pattern similarity between recall of *twist* group and encoding or recall of other subjects in their group or subjects in the other groups, to memory update behavior (twist score)

The same negative correlation pattern of results was observed while comparing the recall-twist group to the movie-twist group, and even the recall-spoiled and movie-spoiled groups (Figure 3.9 B-D). One explanation for this result could be the schematic nature of the patterns captured by temporally-averaged scene representations. Regions in DMN network have been shown to be

involved in processing the schema (Baldassano et al., 2018). Moreover, while reinstatement is often described as reactivation of the encoding representations during retrieval (Bird et al., 2015; Buchsbaum et al., 2012; J. D. Johnson et al., 2009; Polyn et al., 2005; Wing et al., 2014), retrieval patterns of a given participant are reliably similar to the averaged movie patterns of the other participants during *between* subject encoding-recall pattern similarity (Chen et al., 2017). Since scene-specific pattern similarity mostly represents the shared aspect of the events, it was predicted that if a certain recall is closer to the shared schematic recall pattern for that event, the encoding-to-recall pattern similarity would be higher. Perhaps in this experiment, the shared pattern is the dominant storyline, and what *seems* to be happening in each scene, the locations, characters and dialogues, and is shared between movie and recall. When a participant updates their recall and talks about the new interpretation, their brain representation deviates from the shared pattern and becomes more dissimilar to her or his previous movie representation and also to the average of others' representations. At the same time, an updated recall will obtain a high twist-score and therefore contribute to the negative correlations I observed above. While the shared-schematic patterns can be an interesting explanation for these results, having only 21 recall twist-scores to input into the correlations one should be cautious at making strong interpretations.

3.4 Discussion

In this study, the effects of two different perspectives on participants' encoding and recall of the same movie was examined. The two perspectives were having the knowledge of the movie's plot twist (knowing that the main character was dead throughout the movie) vs. not knowing the twist (maintaining the assumption that the main character is alive throughout). The manipulation in the experimental (twist) group was having the participants watch the movie (encoding) without

revealing the plot twist until the end. Then they were asked to recall the events and express their (updated) interpretation of the movie. It was hypothesized that they would change their perspective, at least in the critical scenes (scenes that their interpretation highly depend on the twist information), during the recall task and talk about the new interpretation of the movie. In line with this prediction, most participants incorporated the effect of the plot twist information in their recall, in at least one scene. These results are in accordance with previous behavioral results, such as Anderson and Picheret 's home-buyer/burglar experiment (Anderson and Picheret 1978) in which participants were able to change their recall using the information they were given only after the encoding. In that study, however, it was not possible to disentangle whether new information *modified* the recall representations. The paradigm did not involve dramatic changes in the encoded events and the additional recalled data during the second round could be explained by attending to the previously encoded but ignored information that became important only after changing the perspective. Moreover, brain imaging was not available at the time of that study and brain representations during the recall could not be directly measured or compared with non-updated recall or encoding representations.

The novelty of the current study was to evaluate the neural correlates underlying the recall update behavior using a paradigm that enforced updating the encoded representations. It was observed that the recall representation of the participants in the twist condition group could be differentiated from the recall of the participants in no-twist condition group, but not the participants in the spoiled condition group. Interestingly, main ROIs in the DMN including: precuneus, angular gyrus, ventromedial PFC, and hippocampus, showed significant classification accuracy between recall-twist vs. recall-no-twist, but not recall-twist vs. recall-spoiled groups. These regions are part of the retrieval network, which is also implicated in tasks that require the construction of future or

imagined events (Hassabis & Maguire, 2007; Rugg & Vilberg, 2013; Daniel L Schacter & Addis, 2007; Daniel L. Schacter et al., 2007).

These results suggest that participants in the twist condition modified their representation of the movie at the time of recall. Such that encoded patterns of movie scenes that were reactivated during the recall were also modified to incorporate new information revealed in the plot twist. This matches very well Bartlett's reconstructive memory notion (Bartlett 1932). I will henceforth refer to this hypothesis as the *pattern-modification hypothesis*.

Differentiation of updated and not-updated recalls provides some evidence in support of the pattern-modification hypothesis. However, a stronger support for modifications in memory reinstatement can be only established if the recall patterns are evaluated in connection to movie patterns. Thus, to study a possible connection, first the movie representations were assessed and compared. It was hypothesized that during the encoding of the movie, *spoiled* vs. *twist* conditions would be differentiable since the latter group does not have the knowledge of the twist *yet*. On the other hand, twist vs. no-twist movie encoding conditions might be not differentiable since they both lack the twist information. This hypothesis was confirmed, showing significant classification accuracies in precuneus, mPFC and angular gyrus. The results in the angular gyrus were less reliable since the twist vs. no-twist conditions were also classifiable, which could indicate the noise in the data. This data replicates previous findings (Yeshurun et al., 2017), which showed that timecourses of brain response between two groups of participants listening to the same story with different perspectives can be distinguished using Euclidean distance measure. My study supported those results by classifying brain representations spatially (not just temporally). The results, however, had relatively high significance thresholds (~65% acquired using permutation methods)

which suggests that patterns could not be isolated ideally. This could be due to competing representations, such as the presence of both old and new interpretations in the signal. To support such a claim and to show that real behavior-relevant patterns have been captured (rather than false positive classification of noise), a control comparison was added (twist vs. no-twist conditions) that showed reassuring chance level classification accuracy in most DMN ROIs (expect for angular gyrus).

The interaction of movie and recall classification results may indicate modification of patterns supporting the pattern-modification hypothesis. The patterns of brain response in corresponding voxels were distinguishable between twist and spoiled condition groups during movie encoding, became indistinguishable when compared during recall. Accordingly, the recall patterns were then classifiable between twist and no-twist condition, which were not classifiable while watching the movie. Interactions of movie-watching and recall classifications were significant in all DMN areas (except for angular gyrus). In the whole brain ROI analysis, voxels in vmPFC showed significant interaction after correcting for multiple comparisons. These results could suggest the involvement of these regions in processing incoming information in relation to prior knowledge (plot twist based information) and updating the patterns of representations during recall.

An interesting and unexpected observation in this study, as stated in the behavioral results section, was that participants in the twist condition mostly recalled scenes through the lens of their original perspective (that the main character was alive). The recall updates typically occurred before or after this baseline perspective recall with phrases updating their perspective such as, “But now I know that…”. Surprisingly, even some participants in the spoiled group took the same

approach, by explaining what appeared to be happening in the scene and adding: "But I knew that…" I presume that this behavior highlights the importance and dominance of the old perspective (main character alive). The plot of the movie intentionally implemented this perspective to deceive the audience and shock them with the plot twist at the end. Thus, it is not counterintuitive that even with a more complete set of information, one might still consider the partial-information perspective alongside it – of how it's "supposed" to be perceived. Furthermore, The Sixth Sense compared to audio signal used by Yesherun et al., has less room for imagination and construction. When a scene is being created in one's mind using only a short verbal description, there is room to use one's imagination and construct the mental representation to incorporate a certain creative perspective using new information. However, watching a movie leaves less room for imagination of the individual and depends more on processing the scene the way it is presented by the director, thus the effect of the "correct perspective", may be more subtle. Especially in a movie such as The Sixth Sense, that the director intentionally keep the audience away from the real story, ignoring this dominant narrative may be more challenging for the audience (the subjects in this experiment).

To study whole brain in the current experiment, ROI analysis using functionally defined ROIs from a previous study (Regev et al. 2018), was preferred to searchlight analysis (Kriegeskorte, Goebel, & Bandettini, 2006; Kriegeskorte, Mur, & Bandettini, 2008). ROI analysis is less computationally demanding than a searchlight and has more power due to less number of comparisons. Moreover, in classification or pattern similarity analyses, the number of voxels that create patterns influence the results. Selection of the size of the pattern in a searchlight cube appears arbitrary while voxels in the ROI analysis are functionally related. It is not clear how the activity related to subtle differences in perspectives is spread out across the voxels or what would

be an ideal window size to capture this effect. Additionally, such an ideal window size may be different in different regions of the brain, as regions across the brain process the stimuli differently. Therefore, an ROI analysis appeared more suitable for this study.

The pattern-modification hypothesis was more directly tested using pattern similarity between encoding and recall of the same events and correlating that to behavior. It was shown that *within* and also *between* subject pattern similarity of the matching recall and movie scenes negatively correlated with twist-score. That is to say participants who had higher twist-scores (which indicated more update in the recall) showed lower similarity between their movie and recall patterns. This finding suggests that the change in recall may have influenced the corresponding brain patterns thereby making them more dissimilar to the movie or other recall patterns.

To the best of my knowledge, this study was the first to examine the effect of memory update on brain representations using a naturalistic stimulus. Specifically, the memory modification, which is often studied using memory distortions, was examined in an adaptive real-life situation, in which the update occurred successfully to support behavioral goals. While this study cannot disentangle the possible differences between the mechanisms of memory update in normal functioning versus distorted situations, it may be a fundamental step in navigating this question for future studies. One possible future direction for this line of work could be studying the process of memory update naturalistically over time to assess how congruent or incongruent new information effects the consolidation (and reconsolidation) of memories, and when or how distortions occur. Another question that remains salient for future studies is how successful memory update in adaptive naturalistic tasks influence the connectivity of different brain regions, which could provide more mechanistic insight into the process.

4. Concluding remarks

In this dissertation, I took a two-fold approach to study the reconstructive nature of memory. First, memory encoding and retrieval were evaluated in relation to the process of construction, and it was shown that scene specific patterns of brain response were shared across all these three conditions (encoding, recall, and imagination). Mutual activity between memory and construction were shown at the level of brain response patterns, not just overlapping areas of activity. This work advocates a more holistic approach in studying memory, in which memory is not studied in isolation but in relation to other cognitive functions that entail *construction*. Second, memory update was studied in a naturalistic context in which modifications of memory occurred to achieve behavioral goals (a more accurate recall and sharing movie information via spoken recall). It was shown that participants successfully updated their story recall, and that the updated and not-updated recalls were evident by dissimilarity of the brain representation. Moreover, the analysis suggested a possible link between memory modification and recall behavior. These results validate and encourage studying reconstructive memory using an adaptive approach.

The brain is constantly involved in constructive processes such as in imagining hypothetical situations, predicting future events, memory update, and memory distortions, and visual illusions. A more holistic approach to study construction in the brain may enable a better insight into the neural underlying of these processes. There are many practical applications surrounding the questions that could be answered through a study of constructive processes. For example, how do humans judge the source of an event, whether it is internally or externally generated? How do they know whether it is a fantasy, a dream, or a previous real episode? The

inability to differentiate reality from imagination is present not only in memory related phenomenon but in other forms of reality monitoring as well, such as hallucinations (M. K. Johnson & Johnson, 1993; Simons et al., 2017). Therefore, studying construction in the brain would not only be helpful to memory researchers but would also be of significant benefit to a wide range of question in health and pathology.

I began this dissertation by discussing false memories. I want to conclude it with a discussion of truth. The differences between the interpretations of the same episode begin at the level of encoding. As I have discussed, incoming information is processed in light of prior knowledge. Since all humans have different prior experiences and knowledge, it is not a stretch to imagine they also all may have different interpretations of the same event; different truths. The modification of event memory based on prior knowledge lives beyond encoding and continues during the process of reconsolidation and recollection, as discussed in section 1.1. Furthermore, at the time of sharing a memory with others, recalled memories may be further altered, even unintentionally, to fit the audience or current situation. These shared memories can contribute to a collective memory of that event. At the end of this process, what remains from the truth of the original event is unknown and perhaps varies across different situations and among different individuals. This process highlights that studying reconstructive memory is important not only at individual level, but also it could benefit our understanding of the social and cultural aspect of the human experience.

5. References:

Addis, D. R., Barense, M., & Duarte, A. (2015). *Wiley Handbook on The Cognitive Neuroscience of Memory*. Somerset, UNITED STATES: Wiley. Retrieved from http://ebookcentral.proquest.com/lib/princeton/detail.action?docID=3059091

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

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences*, *5*(12), 533–539.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2016). Discovering event structure in continuous narrative perception and memory. *BioRxiv*, 081018. https://doi.org/10.1101/081018

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

Bartlett, F. C., & Burt, C. (1933). Remembering: A Study in Experimental and Social Psychology. *British Journal of Educational Psychology*, *3*(2), 187–192. https://doi.org/10.1111/j.2044-8279.1933.tb02913.x

Bartlett, S. F. C., & Bartlett, F. C. (1995). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Berkers, R. M. W. J., van der Linden, M., de Almeida, R. F., Müller, N. C. J., Bovy, L., Dresler, M., … Fernández, G. (2017). Transient medial prefrontal perturbation reduces false memory formation. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *88*, 42–52. https://doi.org/10.1016/j.cortex.2016.12.015

Besnard, A., Caboche, J., & Laroche, S. (2012). Reconsolidation of memory: a decade of debate. *Progress in Neurobiology*, *99*(1), 61–80. https://doi.org/10.1016/j.pneurobio.2012.07.002

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex (New York, N.Y.: 1991)*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bird, C. M., Keidel, J. L., Ing, L. P., Horner, A. J., & Burgess, N. (2015). Consolidation of Complex Events via Reinstatement in Posterior Cingulate Cortex. *The Journal of Neuroscience*, *35*(43), 14426–14434. https://doi.org/10.1523/JNEUROSCI.1774-15.2015

Bonasia, K., Sekeres, M. J., Gilboa, A., Grady, C. L., Winocur, G., & Moscovitch, M. (2018). Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic events at short and long delays. *Neurobiology of Learning and Memory*, *153*, 26–39. https://doi.org/10.1016/j.nlm.2018.02.017

Buchsbaum, B. R., Lemire-Rodger, S., Fang, C., & Abdi, H. (2012). The Neural Basis of Vivid Memory Is Patterned on Perception. *Journal of Cognitive Neuroscience*, *24*(9), 1867–1883. https://doi.org/10.1162/jocn_a_00253

Buckner, R., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The Brain's Default Network: Anatomy, Function, and Relevance to Disease. *Annals of the New York Academy of Sciences*, (1124), 1–38.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Burgess, P. W., & Shallice, T. (1996). Confabulation and the control of recollection. *Memory (Hove, England)*, *4*(4), 359–411. https://doi.org/10.1080/096582196388906

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*(5), 219–227. https://doi.org/10.1016/j.tics.2007.02.005

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Chow, H. M., Mar, R. A., Xu, Y., Liu, S., Wagage, S., & Braun, A. R. (2014). Embodied comprehension of stories: interactions between language regions and modality-specific neural systems. *Journal of Cognitive Neuroscience*, *26*(2), 279–295. https://doi.org/10.1162/jocn_a_00487

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2012). Imagery and perception share cortical representations of content and location. *Cerebral Cortex (New York, N.Y.: 1991)*, *22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

Clark, H. H. (2006). Context and Common Ground. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 105–108). Elsevier.

Clark, H. H., & Krych, M. A. (2004). Speaking while Monitoring Addressees for Understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261–288.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Dennis, N. A., Bowman, C. R., & Turney, I. C. (2015). Functional Neuroimaging of False Memories. In *The Wiley Handbook on the Cognitive Neuroscience of Memory* (pp. 150–171). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118332634.ch8

Dikker, S., Silbert, L. J., Hasson, U., & Zevin, J. D. (2014). On the Same Wavelength: Predictable Language Enhances Speaker–Listener Brain-to-Brain Synchrony in Posterior Superior Temporal Gyrus. *The Journal of Neuroscience*, *34*(18), 6267–6272. https://doi.org/10.1523/JNEUROSCI.3796-13.2014

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

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111. https://doi.org/10.1038/s41592-018-0235-4

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The Role of Medial Prefrontal Cortex in Memory and Decision Making. *Neuron*, *76*(6), 1057–1070. https://doi.org/10.1016/j.neuron.2012.12.002

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*(3), 316–355. https://doi.org/10.1037/0096-3445.108.3.316

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, *3*(2), 208–214. https://doi.org/10.3758/BF03212420

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114. https://doi.org/10.1016/j.neuropsychologia.2013.11.010

Ghosh, V. E., Moscovitch, M., Colella, B. M., & Gilboa, A. (2014). Schema Representation in Patients with Ventromedial PFC Lesions. *Journal of Neuroscience*, *34*(36), 12057–12070. https://doi.org/10.1523/JNEUROSCI.0740-14.2014

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the Future. *Science*, *317*(5843), 1351–1354. https://doi.org/10.1126/science.1144161

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Hardt, O., & Nadel, L. (2018). Systems consolidation revisited, but not revised: The promise and limits of optogenetics in the study of memory. *Neuroscience Letters*, *680*, 54–59. https://doi.org/10.1016/j.neulet.2017.11.062

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using Imagination to Understand the Neural Basis of Episodic Memory. *The Journal of Neuroscience*, *27*(52), 14365–14374. https://doi.org/10.1523/JNEUROSCI.4549-07.2007

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 of the United States of America*, *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

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1263–1271. https://doi.org/10.1098/rstb.2008.0296

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., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A Hierarchy of Temporal Receptive Windows in Human Cortex. *The Journal of Neuroscience*, *28*(10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008

Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review*, *16*(1), 80–87. https://doi.org/10.3758/PBR.16.1.80

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: neural responses shared across languages. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(44), 15277–15283. https://doi.org/10.1523/JNEUROSCI.1800-12.2012

Hupbach, A., Gomez, R., & Nadel, L. (2015). Memory Reconsolidation. In *The Wiley Handbook on the Cognitive Neuroscience of Memory* (pp. 244–264). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118332634.ch12

John D. Bransford, M. K. J. (1972). “Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall.” *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. *Neuron*, *63*(5), 697–708. https://doi.org/10.1016/j.neuron.2009.08.011

Johnson, M. K., & Johnson, M. K. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Frontiers in Human Neuroscience*, *8*, 59. https://doi.org/10.3389/fnhum.2014.00059

Kern, R. S., Van Gorp, W. G., Cummings, J. L., Brown, W. S., & Osato, S. S. (1992). Confabulation in Alzheimer’s disease. *Brain and Cognition*, *19*(2), 172–182.

Kesteren, M. T. R. van, Rijpkema, M., Ruiter, D. J., & Fernández, G. (2010). Retrieval of Associative Information Congruent with Prior Knowledge Is Related to Increased Medial Prefrontal Activity and Connectivity. *Journal of Neuroscience*, *30*(47), 15888–15894. https://doi.org/10.1523/JNEUROSCI.2674-10.2010

Konkle, T., & Oliva, A. (2007). Normative representation of objects and scenes: Evidence from predictable biases in visual perception and memory. *Journal of Vision*, *7*(9), 1049–1049. https://doi.org/10.1167/7.9.1049

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642. https://doi.org/10.1038/35090055

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.004.2008

Lacy, J. W., & Stark, C. E. L. (2013). The Neuroscience of Memory: Implications for the Courtroom. *Nature Reviews. Neuroscience*, *14*(9), 649–658. https://doi.org/10.1038/nrn3563

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *The Journal of Neuroscience*, *31*(8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Lindsay, D. S., & Read, J. D. (1995). "Memory work" and recovered memories of childhood sexual abuse: Scientific evidence and public, professional, and personal issues. *Psychology, Public Policy, and Law*, *1*(4), 846–908. https://doi.org/10.1037/1076-8971.1.4.846

Loftus, E. F. (1997). Creating childhood memories. *Applied Cognitive Psychology*, *11*(7), S75–S86. https://doi.org/10.1002/(SICI)1099-0720(199712)11:7<S75::AID-ACP514>3.0.CO;2-F

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366. https://doi.org/10.1101/lm.94705

Loftus, E. F., & Pickrell, J. E. (1995). The Formation of False Memories. *Psychiatric Annals*, *25*(12), 720–725. https://doi.org/10.3928/0048-5713-19951201-07

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, *104*(3), 667–676.

Maguire, E. A. (2001). Neuroimaging studies of autobiographical event memory. *Philosophical Transactions of the Royal Society of London. Series B*, *356*(1413), 1441–1451. https://doi.org/10.1098/rstb.2001.0944

Mar, R. A. (2004). The neuropsychology of narrative: story comprehension, story production and their interrelation. *Neuropsychologia*, *42*(10), 1414–1434. https://doi.org/10.1016/j.neuropsychologia.2003.12.016

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., … Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, 201608282. https://doi.org/10.1073/pnas.1608282113

McGaugh, J. L. (2000). Memory--a Century of Consolidation. *Science*, *287*(5451), 248–251. https://doi.org/10.1126/science.287.5451.248

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal Representation of Related and Opposing Memories Develop within Distinct, Hierarchically Organized Neural Schemas. *Neuron*, *83*(1), 202–215. https://doi.org/10.1016/j.neuron.2014.05.019

Melo, B., Winocur, G., & Moscovitch, M. (1999). False recall and false recognition: An examination of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures. *Cognitive Neuropsychology*, *16*(3–5), 343–359. https://doi.org/10.1080/026432999380825

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726. https://doi.org/10.1038/35021052

Nader, Karim, & Einarsson, E. Ö. (2010). Memory reconsolidation: an update. *Annals of the New York Academy of Sciences*, *1191*(1), 27–41. https://doi.org/10.1111/j.1749-6632.2010.05443.x

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY, US: W W Norton & Co. https://doi.org/10.1037/11494-000

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, *27*(2), 169–190; discussion 190-226.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science (New York, N.Y.)*, *310*(5756), 1963–1966. https://doi.org/10.1126/science.1117645

Poppenk, J., McIntosh, A. R., Craik, F. I. M., & Moscovitch, M. (2010). Past Experience Modulates the Neural Mechanisms of Episodic Memory Formation. *Journal of Neuroscience*, *30*(13), 4707–4716. https://doi.org/10.1523/JNEUROSCI.5466-09.2010

Przybyslawski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation. *Behavioural Brain Research*, *84*(1–2), 241–246.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, *13*(10), 713–726. https://doi.org/10.1038/nrn3338

Regev, M., Honey, C. J., Simony, E., & Hasson, U. (2013). Selective and Invariant Neural Responses to Spoken and Written Narratives. *The Journal of Neuroscience*, *33*(40), 15978–15988. https://doi.org/10.1523/JNEUROSCI.1580-13.2013

Regev, M., Simony, E., Lee, K., Tan, K. M., Chen, J., & Hasson, U. (n.d.). Propagation of Information Along the Cortical Hierarchy as a Function of Attention While Reading and Listening to Stories. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhy282

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407. https://doi.org/10.3758/BF03196177

Romero, K., & Moscovitch, M. (2012). Episodic memory and event construction in aging and amnesia. *Journal of Memory and Language*, *67*(2), 270–284. https://doi.org/10.1016/j.jml.2012.05.002

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*(2), 255–260. https://doi.org/10.1016/j.conb.2012.11.005

Schacter, D. L., Koutstaal, W., & Norman, K. A. (1996). Can cognitive neuroscience illuminate the nature of traumatic childhood memories? *Current Opinion in Neurobiology*, *6*(2), 207–214.

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, *49*, 289–318. https://doi.org/10.1146/annurev.psych.49.1.289

Schacter, Daniel L. (2012). Constructive memory: past and future. *Dialogues in Clinical Neuroscience*, *14*(1), 7–18.

Schacter, Daniel 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, Daniel 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, Daniel L., Addis, D. R., & Buckner, R. L. (2008). Episodic Simulation of Future Events. *Annals of the New York Academy of Sciences*, *1124*(1), 39–60. https://doi.org/10.1196/annals.1440.001

Schacter, Daniel L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic Future Thinking: Mechanisms and Functions. *Current Opinion in Behavioral Sciences*, *17*, 41–50. https://doi.org/10.1016/j.cobeha.2017.06.002

Schacter, D.L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*(3), 182–203. https://doi.org/10.1037/0003-066X.54.3.182

Schnider, A. (2017). *The Confabulating Mind: How the Brain Creates Reality*. Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/oso/9780198789680.001.0001/oso-9780198789680

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687–E4696. https://doi.org/10.1073/pnas.1323812111

Simons, J. S., Garrison, J. R., & Johnson, M. K. (2017). Brain Mechanisms of Reality Monitoring. *Trends in Cognitive Sciences*, *21*(6), 462–473. https://doi.org/10.1016/j.tics.2017.03.012

Sommer, T. (2017). The Emergence of Knowledge and How it Supports the Memory for Novel Related Information. *Cerebral Cortex*, *27*(3), 1906–1921. https://doi.org/10.1093/cercor/bhw031

Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D., & Warren, D. E. (2015). Investigating the Neural Correlates of Schemas: Ventromedial Prefrontal Cortex Is Necessary for Normal Schematic Influence on Memory. *The Journal of Neuroscience*, *35*(47), 15746–15751. https://doi.org/10.1523/JNEUROSCI.2767-15.2015

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-

analysis. *Journal of Cognitive Neuroscience*, *21*(3), 489–510. https://doi.org/10.1162/jocn.2008.21029

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

St-Laurent, M., Abdi, H., Bondad, A., & Buchsbaum, B. R. (2014). Memory Reactivation in Healthy Aging: Evidence of Stimulus-Specific Dedifferentiation. *The Journal of Neuroscience*, *34*(12), 4175–4186. https://doi.org/10.1523/JNEUROSCI.3054-13.2014

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(5), 1565–1572. https://doi.org/10.1523/JNEUROSCI.4657-08.2009

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia*, *44*(12), 2189–2208. https://doi.org/10.1016/j.neuropsychologia.2006.05.023

Szpunar, K. K., Spreng, R. N., & Schacter, D. L. (2014). A taxonomy of prospection: introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18414–18421. https://doi.org/10.1073/pnas.1417144111

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences*, *104*(2), 642–647. https://doi.org/10.1073/pnas.0610082104

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007). Schemas and Memory Consolidation. *Science*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., … 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

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding Sound and Imagery Content in Early Visual Cortex. *Current Biology*, *24*(11), 1256–1262. https://doi.org/10.1016/j.cub.2014.04.020

Wagoner, B. (2017). *The constructive mind: Bartlett's psychology in reconstruction*. New York, NY, US: Cambridge University Press. https://doi.org/10.1017/9780511920219

Warren, D. E., Jones, S. H., Duff, M. C., & Tranel, D. (2014). False recall is reduced by damage to the ventromedial prefrontal cortex: implications for understanding the neural correlates of schematic memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(22), 7677–7682. https://doi.org/10.1523/JNEUROSCI.0119-14.2014

Wing, E. A., Ritchey, M., & Cabeza, R. (2014). Reinstatement of Individual Past Events Revealed by the Similarity of Distributed Activation Patterns during Encoding and Retrieval. *Journal of Cognitive Neuroscience*, 1–13. https://doi.org/10.1162/jocn_a_00740

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same Story, Different Story: The Neural Representation of Interpretive Frameworks. *Psychological Science*, *28*(3), 307–319. https://doi.org/10.1177/0956797616682029

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event Perception: A Mind/Brain Perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185.