Sophie Siestrup

The Influence of Mnemonic Prediction Errors on Brain Activity and Episodic Memory - a Perspective on Memory Modification

- 2023 -

Psychologie

The Influence of Mnemonic Prediction Errors on Brain Activity and Episodic Memory - a Perspective on Memory Modification

Inaugural-Dissertation

Zur Erlangung des Doktorgrades

im Fachbereich Psychologie und Sportwissenschaft

der Westfälischen Wilhelms-Universität in Münster

Vorgelegt von

Sophie Siestrup

aus Steinfurt

- 2023 -

Dekan:	Prof. Dr. Gerald Echterhoff
Erste Gutachterin:	Prof. Dr. Ricarda I. Schubotz
Zweite Gutachterin:	Prof. Dr. Pienie Zwitserlood
Datum der mündlichen Prüfung:	_____
Datum der Promotion:	_____

Acknowledgements

First, I would like to thank my supervisor Ricarda I. Schubotz for giving me the opportunity to pursue a doctorate in her lab. I am highly grateful for her continuous support and encouragement, all the critical discussions and her constant openness. I really appreciate how we overcame all challenges together. I would also like to thank Pienie Zwitserlood for taking over the role of my co-supervisor, including her invariable interest in the project despite all the difficulties the COVID pandemic posed for us. Furthermore, I am grateful for the great organization of the research group FOR 2812 by Sen Cheng and Vinita Samarasinghe. I am very happy that I had the chance to be a part of this wonderful interdisciplinary group.

Furthermore, I would like to thank Benjamin Jainta, for working on this project together with me and sharing all burdens, responsibilities and successes; Ima Trempler for the guidance and encouragement whenever needed; Falko Mecklenbrauck, for all the helpful advice and fun conversations; and Nina Heins for the persisting (emotional) support. Of course, I am grateful to every current and former member of the Biological Psychology Group at the University of Münster who were always supportive and made me enjoy working in the group: Nadiya El-Sourani, Jennifer Pomp, Lena Leeners, Amelie Hübner, Marlen Roehe, Rosari Naveena Selvan, Nina Liedtke, Marius Boeltzig, and Anoushiravan Zahedi. I would also like to express my gratitude to Monika Mertens, who always did whatever possible to help me with data collection and enriched my time in this group with her understanding, helpful nature. Warm thanks also to all student assistants who were not only a great help, but also wonderful people to spend time with: Helena Sydlik, Annika Garlichs, Christin Schwarzer, Simon Wieczorek, Lena Puder, Lana Steuernagel and Alea Bexten.

Finally, I would like to thank my family; my parents, Dagmar and Jürgen, my sister, Johanna, and Maxi for always supporting me and being there for me whenever I needed it.

Table of Contents

Summary 1

List of Original Publications 4

1 Theoretical and Empirical Background 5

1.1 Episodic Memories – Features and Functions 6

1.2 Neural Bases of Episodic Memory 7

1.2.1 Encoding 7
1.2.2 Consolidation 9
1.2.3 Retrieval 11

1.3 False Memories and Memory Modification 13

1.3.1 The Constructive Nature of Memory 15
1.3.2 Labilization and Reconsolidation 16
1.3.3 Interference Between Memory Traces 18

1.4. Memory Modification – Flaw or Favorable? 19

1.5 Predictive Coding and Episodic Memory 22

1.5.1 The Predictive Coding Framework 22
1.5.2 The Influence of Prediction Errors on Episodic Memory 25
1.5.3 Neural Underpinnings of Mnemonic Prediction Errors 28

2 Research Questions 30

3 Research Articles 33

3.1 Study 1: What Happened When? Cerebral Processing of Modified Structure and Content in Episodic Cueing 33

3.2 Study 2: Solidity Meets Surprise: Cerebral and Behavioral Effects of Learning from Episodic Prediction Errors 53

3.3 Study 3: Minor Changes Change Memories: FMRI and Behavioral Reflections of Episodic Prediction Errors 77

4 General Discussion and Future Directions 129

4.1 Summary of the Presented Studies ____ 129
4.2 Mnemonic Prediction Errors in the Brain ____ 131
4.2.1 Neural Responses to Mnemonic Prediction Errors ... 131
4.2.2 The Role of Hippocampus in Mnemonic Prediction Error Processing ... 138
4.2.3 The Influence of Consolidation ... 140
4.3 The Role of Mnemonic Prediction Errors in Memory Modification ___ 143
4.3.1 Behavioral Evidence ... 143
4.3.2 Neural Evidence ... 146
4.3.3 Mechanisms of Modification ... 151
4.4 Critical Evaluation and Methodological Considerations ____ 153
4.4.1 Basic Assumptions and Interpretations ... 153
4.4.2 Detailed Methodological Decisions ... 155
4.5 Outlook ____ 158
4.5.1 General Ideas for Future Research ... 158
4.5.2 Tracking Mnemonic Representations After Prediction Error: A Proposed Future Study ... 160
5 Conclusion ____ 164
References ____ 165
Appendix ____ 184
Abbreviations ____ 187
List of Figures ____ 188
Curriculum Vitae ____ 189
Declarations ____ 191

Summary

Episodic memories allow us to vividly relive our personal past. However, they are not exact recordings of our experiences, but can be modified on the basis of new relevant information. Recently, mnemonic prediction errors (PEs) have been identified as a potential driving force for such memory changes. Mnemonic PEs arise when there is a mismatch between what we expect on the basis of memories and true situational input. According to the predictive coding framework, PEs serve as learning signals to the brain to update internal predictive models. To that end, memory modifications supposedly allow us to maintain valid predictions in an ever-changing environment. However, little is known about how episodic PEs are processed by the brain, how they impact memory and which conditions might influence these responses. For this reason, the aim of the current thesis was to broaden the understanding of mnemonic PEs in episodic memory by analyzing signatures of their neural processing and their influence on episodic memories.

For this purpose, a new episodic modification paradigm was established. Participants encoded complex, unique episodes from demo videos in the form of toy stories. After further consolidation through retrieval practice (**Studies 1** and **2**), participants went through a functional magnetic resonance imaging (fMRI) session during which original but also slightly modified episodes were presented to elicit PEs. Afterwards, participants conducted a post-fMRI memory test during which memory for originally encoded episodes was probed. Again, original but also modified episodes were presented, and participants had to judge after every video whether it belonged to the original episode repertoire. **Study 1** focused on the difference between PEs based on structure or content episode information. To that end, episodes were either modified by switching two subsequent action steps (structure modification) or a single object (content modification). In **Study 2**, which was based on the same data as **Study 1**, the influence of different retrieval practice protocols was analyzed to evaluate the impact of

memory solidity on PE processing. Lastly, **Study 3** focused on the differential processing of content PEs that either challenged the gist of episodes (gist modifications) or left it intact (surface modifications).

One major finding was that mnemonic PEs commonly activated superior parietal areas as well as (right) ventrolateral prefrontal cortex. Consequently, these areas seem to be involved in the processing of episodic mismatch and mnemonic selection processes in general. Furthermore, as expected, different types of episodic modifications elicited different brain responses. In **Study 1**, content modifications yielded increased brain activation in temporo-occipital, parietal and parahippocampal areas, reflective of new object information processing. Structure modifications were characterized by activation in right dorsal premotor cortex, as well as posterior temporal and parietal areas, reflecting the processing of new sequence information. In **Study 3**, surface and gist modifications elicited activation in brain regions involved in the processing of content modifications. Except for anterior cingulate cortex (ACC), midcingulate cortex and posterior hippocampus, neural responses were stronger for gist than surface modifications, corroborating the idea that they induced a stronger PE signal.

In **Study 2**, a behaviorally relevant increase in memory solidity was achieved by increasing the frequency of retrieval practice from two to eight retrievals. This was evidenced by the participants' decreased tendency to endorse modified episodes as originally encoded, also after PE. Interestingly, activation in pre- and subgenual ACC increased for modified episodes after fewer compared to more previous retrieval opportunities. This finding may be attributed to increased neural processing demands during episodic retrieval and/or the role of ACC in learning from PE. Additionally, it was demonstrated in **Study 2** that activation in the episodic memory network, including ACC and hippocampus, specifically increased over time when false memories were formed. Furthermore, neural mismatch responses were enhanced for later correct rejections compared to false alarms.

In all studies, PEs influenced episodic memories, further corroborating the idea that mnemonic PEs can lead to memory modification. Memory changes were characterized by the participants' increased tendency to endorse modified episodes as originally encoded in the post-fMRI memory test after the experience of PEs during the preceding fMRI session. Memory for original episodes remained largely unaffected by PEs. Interestingly, only PEs based on gist modifications did not lead to the formation of false memories, potentially because of the highly impactful mismatch signal due to their high episodic relevance.

Taken together, this thesis provides valuable new insights into the neural processing of mnemonic PEs and false memory formation. The neural and behavioral findings suggest that memory modifications after PE were characterized by the additional encoding of an alternative episode version, competing with the original one. In the case of strong mismatch signals, the formation of false memories was avoided, likely through a clear mnemonic distinction of episodes.

List of Original Publications

This thesis is based on the following original research articles:

Study 1 Siestrup, S., Jainta., B., El-Sourani, N., Trempler, I., Wurm, M. F., Wolf, O. T., Cheng, S., & Schubotz, R. I. (2022). What happened when? Cerebral processing of modified structure and content in episodic cueing. *Journal of Cognitive Neuroscience*, *34*, 1287–1305. https://doi.org/10.1162/jocn_a_01862

Study 2 Siestrup S., Jainta B., Cheng S., & Schubotz R. I. (2022). Solidity meets surprise: Cerebral and behavioral effects of learning from episodic prediction errors. *Journal of Cognitive Neuroscience*, *35*, 291–313. https://doi.org/10.1162/jocn_a_01948

Study 3 Siestrup, S., & Schubotz, R. I. (2023). Minor changes change memories: FMRI and behavioral reflections of episodic prediction errors. *Manuscript submitted for publication/Manuscript under review*

[Please note that additional empirical findings which were not reported in the listed original publications, but are based on the same acquired data, are included in the Appendix. They will be referred to in the General Discussion and Future Directions section (section 4).]

1 Theoretical and Empirical Background

Can I trust my memories? This certainly is a question that most of us have asked ourselves before. From the perspective of episodic memory research, the disillusioning answer is: no, you cannot; at least not unconditionally. However, we should not be frustrated by this observation, but rather see it as a chance: only the fact that memory is not perfect allows us to study it (Schacter & Slotnick, 2004). But if memory is not veridical, why do we remember at all? A surprisingly straightforward answer to this question has been brought forth by decades of episodic memory research: we remember the past to predict the future (Schacter, 2012; Schacter & Addis, 2007a). In the highly dynamic environment we call our home, successful predictions are only possible if we allow for a certain flexibility in the templates that we use for forming these predictions, that is, our memories (Alberini & Ledoux, 2013; Finn, 2017).

Recent theoretical considerations as well as empirical findings highlight the role of mnemonic prediction errors (PEs) in memory modification. Such PEs arise when our predictions, derived from memories, do not match the current situational inputs. Mnemonic PEs might act as a signal as to when a memory needs to be modified to remain a reliable basis for future predictions (Exton-McGuinness, Lee, & Reichelt, 2015; R. S. Fernández, Boccia, & Pedreira, 2016). The aim of the current thesis is to broaden our understanding of mnemonic PEs in episodic memory by characterizing signatures of their neural processing and their impact on the fate of episodic memories.

In the following sections, this thesis will introduce the concept of episodic memories and their neural foundations. It will become clear that episodic memories are not exact copies of past experiences, as false memories and memory modifications occur. Some suggested mechanisms underlying memory modifications will be presented, and the proposed adaptive value of memory modifications will be highlighted. In line with the predictive coding

framework, mnemonic PEs will be introduced as a possible trigger of memory modifications, giving rise to the research questions addressed by this thesis.

1.1 Episodic Memories – Features and Functions

Episodic memory, a form of declarative memory, (Squire, 1992a; Suddendorf & Corballis, 2007; Tulving, 2005) refers to memory for personally experienced events (Tulving, 2005; but see Pillemer, Steiner, Kuwabara, Thomsen, & Svob, 2015 for a discussion on vicarious memories). However, even though (or maybe *because*) the field of episodic memory research has been extensively pursued for many decades, the exact details which define episodic memory are still not completely agreed upon (Suddendorf & Corballis, 2008).

Following an early suggestion by Tulving, memories can be classified as episodic when they include information about *what* happened *where* and *when* (Tulving, 1972). Another common approach for defining episodic memory is by focusing on the characteristics of its retrieval. In that sense, it was postulated that episodic memory retrieval is accompanied by autonoetic consciousness, which means being aware that the remembered episode was experienced personally (Tulving, 1985). However, autonoetic consciousness can be complicated to study empirically, since it relies on subjective experiences (Suddendorf & Corballis, 2008). It is also possible to define episodic memories through several key characteristics. For example, Conway (2008) describes episodic memories as representing sensory and perceptual details, most often in the form of visual imagery, with a clear perspective (observer or field perspective).

Episodic memory is essential for learning. It provides a base for deriving general knowledge and abstract schemas (Conway, 2008; Conway, Gardiner, Perfect, Anderson, & Cohen, 1997). It has been proposed that a key function of episodic memory is to maintain records of short-term goal processing, as episodic memories allow us to track our (recent) experiences so that we can prepare to execute a certain planned behavior when the right moment

comes (Conway, 2008). Suddendorf and Corballis (2007) proposed that episodic memory is the past-directed component of *mental time travel*, which potentially evolved due to the fitness benefit of mentally simulating the future to plan ahead. Thus, episodic memory is highly relevant for predicting and planning future actions and pursuing short- as well as long-term goals. Furthermore, for understanding the nature of episodic memories and, in turn, their malleability, it is crucial to illuminate how mnemonic representations of episodes are formed, processed, and recalled.

1.2 Neural Bases of Episodic Memory

1.2.1 Encoding

To enable us to later remember an experience, it must first be *encoded* into our episodic memory. During encoding, event features, such as sensory inputs, are processed and linked together (Shimamura, 2014), so that durable memories can be formed (Rugg, Johnson, & Uncapher, 2015). Over the years, lesion studies in amnestic patients and animals as well as neuroimaging work have identified several brain regions involved in the encoding of episodic memories. These include structures of the medial temporal lobe (MTL), like hippocampal and parahippocampal areas, prefrontal cortex (PFC), retrosplenial cortex, and parietal cortex (Hasselmo, 2012; Manns & Squire, 2002; Rugg, 2002; Shimamura, 2014).

Generally, different cortical regions can be involved in episodic encoding, depending on what exactly is being encoded. In that sense, it has been observed that activation during encoding occurs in the same cortical regions that are involved in current situational processing (Rugg et al., 2015). Aside from that, the contribution of MTL is crucial for successful encoding (Rugg et al., 2015). For example, a large amount of neuroimaging studies has shown that increased activation in hippocampus and parahippocampal cortex during episodic encoding correlates with later successful remembering (e.g., Alkire, Haier, Fallon, & Cahill, 1998; Hasselmo, 2012; Paller & Wagner, 2002; Ranganath, 2010; Rugg, 2002; Shimamura, 2014). It

has been suggested that the key function of hippocampus in episodic memory encoding is linking together separate event features that constitute an episode (Davachi, 2006; Ranganath, 2010; Rugg, 2002; Shimamura, 2014), and to then store this episode's pattern of cortical activations (Rugg et al., 2015). Hippocampus also mediates so-called *pattern separation*, so that cortical activation patterns representing individual (similar) episodes can be differentiated (Dodson & Schacter, 2002; Kirwan & Stark, 2007; Ngo, Michelmann, Olson, & Newcombe, 2021).

Another crucial brain region in episodic memory encoding is PFC. To form an episodic memory for a particular event, some event features must be encoded from working memory into long-term memory (Shimamura, 2014). This selective control process is likely supported by PFC (Miller & Cohen, 2001; Shimamura, 2008, 2014), and some evidence suggests that medial PFC (mPFC) regions are involved in such strategic executive functions during memory encoding (Himmer, Schönauer, Heib, Schabus, & Gais, 2019). Furthermore, left PFC was suggested to have a key role in episodic encoding (Tulving, Kapur, Craik, Moscovitch, & Houle, 1994), potentially by integrating semantic features of items into episode representations (Rugg, 2002). Interestingly, also reduced activation in several brain areas has been linked to efficient encoding, namely in ventral posterior parietal cortex (vPPC) and posteromedial cortex, including posterior cingulate cortex (PCC; Daselaar, Prince, & Cabeza, 2004; Shimamura, 2014; Uncapher & Wagner, 2009). Shimamura (2014) proposed that successful encoding requires a suppression of these brain regions, potentially to foster a prioritized encoding of sensory stimulus features.

It is noteworthy that not all aspects of an episode are equally likely to be encoded successfully (Rugg et al., 2015). For example, aspects of episodes that receive more attention during encoding are known to be remembered better (Chun & Turk-Browne, 2007; Rugg et al., 2015), and it has been suggested that only the gist of an episode is stored, i.e., its most relevant

features (Cheng, Werning, & Suddendorf, 2016). In line with behavioral findings, it has been demonstrated that cortical-hippocampal interactions during encoding can be modulated by attention as well (Rugg et al., 2015).

1.2.2 Consolidation

From behavioral evidence it is known that some time after encoding, memories are less easily disrupted compared to directly after encoding (Hasselmo, 2012). That is because memories are further stabilized, or *consolidated* (Dudai, 2004; Rugg et al., 2015; Shimamura, 2014). Generally speaking, consolidation is accomplished through the replay or reactivation of memory content so that mnemonic representations are strengthened (Shimamura, 2014; Tranel & Damasio, 2002). More specifically, it has been suggested that during this process, hippocampal-neocortical connections are revised, allowing for the memory to be integrated with existing knowledge structures and/or to be differentiated from competing representations (Antony, Ferreira, Norman, & Wimber, 2017). Replay of memory content can occur off-line during sleep (Inostroza & Born, 2013; Poe, Walsh, & Bjorness, 2010), especially so-called slow-wave sleep (Cheng & Werning, 2013), but also through the active retrieval of memories, for example during retrieval practice (Antony et al., 2017; Hasselmo, 2012; Rowland, 2014).

At the beginning of the consolidation process stand changes of synaptic efficiency, managed through neurochemical mechanisms that act on a molecular and cellular level. This first phase is quickly completed, usually lasting few minutes to days, and is referred to as cellular or *synaptic consolidation*. Afterwards follows what is called *systems consolidation*, i.e., the adaptation of neural networks that support memory storage and/or retrieval (Dudai, 2004; Moscovitch, 2003; Moscovitch, Winocur, Ryan, & Nadel, 2008; Nadel, Hupbach, Hardt, & Gomez, 2008). The traditional view of systems memory consolidation is that eventually, repeated replay enables neocortex to support the memory trace without hippocampus. According to this *standard theory of consolidation* (Squire, 1992b), neocortical areas will then

interact on their own to retrieve a fully consolidated memory, which occurs completely independent of the hippocampus (Moscovitch et al., 2008; Ryan, Hoscheidt, & Nadel, 2008). Within this framework, mPFC is believed to take over the organization of reinstatement of episodic memories from hippocampus, making it a central hub of memory consolidation (Euston, Gruber, & McNaughton, 2012). This theory was mainly inspired by the finding that amnestic patients with hippocampal lesions could still recall remote personal events that had occurred before the brain damage (Squire, 1992b). However, recent evidence also provides support for an alternative account, the so-called *multiple trace theory* (Moscovitch et al., 2008). According to this theory, hippocampal and neocortical sites will always act together to retrieve an episodic memory. Since the retrieval context differs each time a memory is reactivated, modified memory traces will be created and re-encoded. The older a memory is, the more often it will be reactivated, leading to more widespread and/or stronger memory traces in hippocampus and thereby enhanced resilience (Moscovitch et al., 2008). In line with this framework, many neuroimaging studies have demonstrated that remote episodic memories still activate hippocampus (e.g., Maguire, Henson, Mummery, & Frith, 2001; Rekkas & Constable, 2005; Ryan et al., 2001; Ryan et al., 2008).[1]

Interestingly, consolidation does not necessarily mean that a memory is permanently stabilized. For example, it has been proposed that when a memory is reactivated, it will be returned to a labile state and become open for alterations (Nader, 2015; Nader, Schafe, & Doux, 2000). However, when memories have been extensively consolidated, such alterations become more unlikely (Exton-McGuinness et al., 2015; R. S. Fernández et al., 2016).

[1] Please note that there is a multitude of theories about memory consolidation and only two seminal theories are presented as examples in this thesis.

1.2.3 Retrieval

Remembering the personal past not only requires successful encoding, but also *retrieval* of episodes (Shimamura, 2014)[2]. Retrieval is triggered by an internal or external retrieval cue which leads to "the emergence of a consciously accessible representation of a specific past episode" (Rugg et al., 2015, p. 85). It has been proposed that a specific cognitive state, called *retrieval mode*, must be adopted so that stimuli can serve as retrieval cues (Rugg et al., 2015; Tulving, 1983). Furthermore, a "core recollection network" (Rugg et al., 2015, p. 95) has been identified for episodic memory retrieval. This network includes hippocampal and parahippocampal areas, (medial) PFC, PCC and retrosplenial cortex, as well as areas in parietal cortex (Hayama, Vilberg, & Rugg, 2012; J. D. Johnson & Rugg, 2007; Rugg & Vilberg, 2013; Rugg et al., 2015).

An especially robust connection has been found between episodic retrieval and PFC activation (for reviews, see Fletcher & Henson, 2001; Rugg, 2002). Retrieval related activation has been detected in anterior parts of PFC (Brodmann area [BA] 10), but also in dorsolateral locations (Rugg, 2002). During early stages of retrieval, (right) PFC is involved in executive control processes crucial for selective episode feature reactivation (Cabeza, Locantore, & Anderson, 2003; Wagner, 2002; Shimamura, 2014), which Schacter and colleagues referred to as *focusing* the retrieval (e.g., Schacter, Norman, & Koutstaal, 1998). It has also been suggested that right PFC supports the aforementioned retrieval mode, allowing episodic memories to be experienced autonoetically (Nyberg et al., 1995; Rugg, 2002; Tulving, 1983; Wheeler, Stuss, & Tulving, 1997). Further, frontal regions are involved in monitoring and evaluating retrieval outputs (Rugg, Fletcher, Frith, Frackowiak, & Dolan, 1996; Schacter et al., 1998), for example during source monitoring, i.e., determining the origin of a memory (M. K. Johnson, Hashtroudi,

[2] Please note that retrieval is usually accompanied by new (incidental) encoding (e.g., Finn, 2017), so it can be difficult to attribute observed brain activation during retrieval tasks to one or the other process with absolute certainty (e.g., Okado & Stark, 2005).

& Lindsay, 1993; Schacter et al., 1998). MPFC involvement (including anterior cingulate cortex [ACC]) is often observed during the retrieval of remote rather than recent memories (Euston et al., 2012), and has been linked to retrieval related processing of context information (Kveraga et al., 2011; Rugg et al., 2015) and strategic memory search (Himmer et al., 2019). Additionally, as outlined above, it has been suggested that after consolidation, mPFC instead of hippocampus mediates the reinstatement of episodic memories (Euston et al., 2012).

Aside from PFC, medial parietal cortex is reliably activated during episodic memory retrieval (Rugg, 2002), including, precuneus and PCC (Henson, Rugg, Shallice, Josephs, & Dolan, 1999; Kapur et al., 1995; Rugg, 2002). Presumably, these areas subserve visual imagery during the reactivation of experienced episodes (Fletcher et al., 1995; Rugg, 2002). Additionally, lateral parietal cortex, specifically angular gyrus, is active during episodic retrieval. It was proposed that this region contributes to the reorientation of attention from the retrieval cue to retrieved content (Cabeza, Ciaramelli, & Moscovitch, 2012; Rugg et al., 2015). Aside from that, it was suggested that vPPC is involved in forming connections between cortical areas which facilitates their coactivation during retrieval (Shimamura, 2011, 2014).

Additionally, MTL structures are involved in episodic memory retrieval. When a retrieval cue triggers cortical activation that (partially) overlaps with the one of a past episode, the hippocampal representation of this episode is reactivated, finally leading to the reinstatement of the episode's cortical activation pattern at encoding (Rugg et al., 2015). This process is called *pattern completion* (e.g., Sugar & Moser, 2019). Accordingly, brain activation at encoding and retrieval is often highly similar (e.g., Danker & Anderson, 2010; Rugg et al., 2015; Shimamura, 2014), lending support to the *cortical reinstatement hypothesis* of episodic retrieval (Abe, 2012; Rugg, Johnson, Park, & Uncapher, 2008).

Importantly, brain activation observed at retrieval not only mirrors encoding-related activation. The reason for this is the constructive, or generative, nature of episodic memory

retrieval, so that brain activity during retrieval of episodic memories also reflects constructive processes (Rugg et al., 2015; see section 1.3.1 for more details). This generative mechanism can lead to a change of memory content, as illuminated in the next paragraph.

1.3 False Memories and Memory Modification

Our episodic memories are far from perfect, most obviously since we tend to forget. Aside from that, memories often differ from an originally experienced event, and can sometimes appear completely false, i.e., not even grounded on something that happened to us (Schacter, 1999).

False memories have received much attention since it was discovered that some patients started to remember non-veridical episodes of sexual abuse as a result of therapy that included strategies like suggestive questioning or hypnosis (M. K. Johnson, Raye, Mitchell, & Ankudowich, 2012; Loftus, 1997; Schacter, 2012). In legal settings, false memories pose an important challenge, for example in the case of eyewitness testimony. Furthermore, innocent people can falsely remember (and confess) committing crimes when confronted with fake evidence or suggestive interrogation techniques (Loftus, 1997; Schacter, 1999, 2012). These observations inspired the work of Elizabeth Loftus and her colleagues. They demonstrated that participants' memory can be manipulated so that they are convinced to remember an event, for example being lost in the mall as a child, that never actually happened to them via specific suggestion techniques (Loftus, 1996, 1997). Further, it was observed that when people witness an event and later are presented with new or conflicting information about it, their memory can be modified. As a consequence, they misremember that said new information was part of the originally experienced episode and may even reject veridical episode details (Brainerd & Reyna, 1998). This phenomenon is called *misinformation effect* (Loftus, 1996, 1997). Related to this are the so-called *memory intrusions*: when participants study two sets of items (e.g., word lists or groups of objects), they sometimes erroneously remember items belonging to the

wrong set; usually, items from the second set intrude memory for the first set (Hupbach, Gomez, & Nadel, 2009; Klingmüller, Caplan, & Sommer, 2017). Together, these instances of false remembering can be grouped under the umbrella term *suggestibility* (Schacter, 1999; Schacter & Slotnick, 2004).

Suggestibility can be seen as one example of *misattribution*, i.e., the false recollection of the original source of an event or item (Schacter, 1999; Schacter & Slotnick, 2004). Importantly, misattribution can also occur without suggestion from external sources, for example when a new event is spontaneously classified as previously experienced (Schacter, 1999). This type of memory error is called *false recognition* (Dodson & Schacter, 2002; Underwood, 1965). False recognition is often observed for lures, e.g., words or sentences, that are highly similar to and/or semantically related with encoded material (Bransford & Franks, 1971; Dodson & Schacter, 2002; Franks & Bransford, 1970; Underwood, 1965). Similarly, participants also tend to produce *false recall* (Reyna, Corbin, Weldon, & Brainerd, 2016; Roediger & McDermott, 1995) by naming words or sentences that were not studied, but are closely related to encoded material. Additionally, memories are subject to *bias,* which means that the encoding and retrieval of personal experiences is influenced by knowledge, expectations, personal beliefs and emotions (Schacter, 1999; Schacter & Slotnick, 2004).

A multitude of studies suggest that false memories, or memory modifications, can result from processes that come into play during encoding and/or retrieval of episodic memories (Dodson & Schacter, 2002). The next paragraphs will introduce a selection of important theories of false memory formation and memory modification. Importantly, these explanatory accounts are not necessarily mutually exclusive (e.g., Sinclair & Barense, 2019) and might rather be interpreted as complementary.

1.3.1 The Constructive Nature of Memory

Early theories of memory assumed that a memory is just passively stored in the brain, ready to be retrieved again, much like pulling a file from a drawer (Schacter, 1995). Today, however, there is large agreement among researchers that remembering is an active and constructive process (Schacter, 2019). In fact, the constructive nature of memory is often seen as a basic requirement for its malleability. Conway and Loveday (2015) even argued that, to some degree, all episodic memories are false as they are a product of said constructive process. The current understanding of constructive memory encoding and retrieval has been described previously in this thesis (sections 1.2.1 and 1.2.3). Briefly, the main characteristics of constructive memory are the binding of separate episode features, pattern separation during encoding and pattern completion during retrieval in the MTL. Further important processes are selective retrieval and the monitoring of retrieval outputs, likely mediated by PFC (Schacter et al., 1998). In summary, the constructive memory framework assumes that a memory is pieced together from individual parts each time it is recalled (Schacter & Addis, 2007b).

During all stages of the constructive process, certain challenges can arise that can influence veridical reconstruction. For example, when pattern separation fails or the retrieval focus is unspecific, features from different episodes can be reactivated, leading to competition and potentially source confusion (M. K. Johnson et al., 2012; Schacter et al., 1998). Furthermore, recognition may then be based on information that is similar between representations, i.e., the gist, and not their distinct features (Schacter et al., 1998). When retrieval monitoring is accomplished on the basis of inadequate episodicity criteria, internally generated information (i.e., imagined scenarios) may be erroneously attributed to experienced episodes (M. K. Johnson et al., 2012; Schacter et al., 1998). Aside from the challenges outlined above, the constructive process was shown to be heavily influenced by the current context, semantic knowledge, and beliefs and attitudes. Like that, the final memory becomes tinted by

these aspects (Schacter, 2012). Hence, the constructive memory framework highlights that the episodic memory system does not act independently of other cognitive and memory systems (Cheng et al., 2016).

While memories might be modified during the process of retrieval itself, as the constructive memory framework postulates, there is also evidence that memories can be changed after retrieval, or reactivation. One explanatory account for this is provided by the *reconsolidation hypothesis*, as outlined in the next section.

1.3.2 Labilization and Reconsolidation

Past studies showed that the consolidation of newly acquired memories can be disrupted shortly after encoding by inhibiting new protein synthesis, leading to amnesia (Davis & Squire, 1984; Nader et al., 2000). Similar observations were made when already consolidated fear memories in rats were reactivated. When disruptive treatments like electroconvulsive shock or protein synthesis inhibitors were applied in a short time window after reactivation, weakening or loss of before established fear memories were induced (Judge & Ouartermain, 1982; Misanin, Miller, & Lewis, 1968; Nader et al., 2000; Sara, 2000). It was concluded that when a memory is reactivated, it becomes labile again and needs to be reconsolidated, a process that requires new protein synthesis, just like initial consolidation (Nader et al., 2000). When the memory is labile, it can be modified, i.e., strengthened, weakened, or altered to include new information (Elsey, Van Ast, & Kindt, 2018; Finn, 2017). As the effects of protein synthesis take some time to unfold, usually several hours, the final outcome of reconsolidation can only be observed after a delay between reactivation and test (Elsey et al., 2018; Kiley & Parks, 2022). Since its introduction, evidence for the reconsolidation hypothesis has been found for different types of memories, including implicit and explicit ones, and in a multitude of species, including humans (Alberini & Ledoux, 2013; Elsey et al., 2018; Hupbach, Gomez, Hardt, & Nadel, 2007; Sinclair & Barense, 2018).

Usually, reconsolidation studies in humans apply so-called interference paradigms. In a first step, participants learn a set of words or objects. In a second session, they are reminded of the original experience, which is believed to render the memory labile, and shortly after get to learn a new set of items. During a memory test in a third session, participants often believe that new items as encountered in session two were experienced in session one, but not vice versa. This finding has been interpreted as successful reconsolidation, so that new information was incorporated into the original memory trace (Finn, 2017; Hupbach et al., 2007; Kiley & Parks, 2022). As in non-human animals, this effect was demonstrated to be time-dependent, i.e., there need to be temporal gaps between all experimental sessions in order for initial consolidation and reconsolidation to occur (Hupbach et al., 2007; Sinclair & Barense, 2018). Reactivated memories can not only be influenced by interfering information, but also electroconvulsive shock, transcranial magnetic stimulation (TMS) and certain pharmacological interventions in humans (Elsey et al., 2018; Sinclair & Barense, 2019). Importantly, the reconsolidation account assumes that through these interventions, the original memory trace is changed permanently (Brewin, 2015).

However, some criticism has been raised regarding evidence for reconsolidation in humans. Aside from mixed findings and replication problems concerning the reconsolidation effect (Gisquet-Verrier & Riccio, 2018), it has recently been argued that most reconsolidation studies in humans lack an ideal design that would allow results to be exclusively interpreted in light of the reconsolidation hypothesis (Elsey et al., 2018). In fact, alternative explanations could account for many findings (Brewin, 2015; Gisquet-Verrier & Riccio, 2018; Kiley & Parks, 2022; Klingmüller et al., 2017; Sederberg, Gershman, Polyn, & Norman, 2011), and it is likely that reconsolidation is not the only mechanism that allows for modification in human episodic memories (Sinclair & Barense, 2018). Another mechanism that could influence

veridical recall in humans, the interference between competing memory traces, will be presented in the following.

1.3.3 Interference Between Memory Traces

False memories and memory modifications might come about through the confusion or competition between different memory traces, i.e., their interference (Klingmüller et al., 2017; Loftus & Pickrell, 1995; Yassa & Reagh, 2013). Memories can be influenced by previous experiences (proactive interference) or subsequent experiences (retroactive interference), while the focus of interest in memory modification research is often on the latter, like in the misinformation paradigm (Loftus & Pickrell, 1995). Accordingly, interference may also be the underlying source of the aforementioned memory intrusions (Klingmüller et al., 2017). Importantly, interference accounts usually assume that the original memory trace is left unaltered and that false memories or memory modifications arise due to the confusion of different traces (Klingmüller et al., 2017; Reyna, 1995; Sederberg et al., 2011).

Several theories have been developed to explain how exactly interference influences memory. One of those is *fuzzy-trace theory*. The basic assumption of fuzzy-trace theory is that memories comprise two different kinds of representations. Verbatim representations code for surface details of episodes, while gist representations capture their overall meaning (Brainerd & Reyna, 1990; Reyna & Brainerd, 1995; Reyna, 1995). These representations can interfere with each other, which can cause memory errors. In general, gist representations are less vulnerable to interference than verbatim ones, which are believed to be more difficult to access, especially after longer time intervals. Suggestibility effects might arise due to the inaccessibility of verbatim memory, for example the exact source of a piece of information. In that case, competing verbatim information from external sources or gist-congruent misinformation might interfere with the original memory and consequently be endorsed as originally experienced (Reyna, 1995). False recognitions without externally suggested information might occur due to

the interference of gist memory with verbatim memory, so that (semantically) related lures are falsely recognized as previously encountered (Reyna et al., 2016; Reyna, 1995). According to a related notion, interference of memory traces might result from missing distinctiveness. Following this idea, false memories are always based on existing memory content. When memories are not distinct from each other from the beginning (i.e., due to undetailed encoding), or they lose their distinctiveness over time, retroactive interference can more easily occur between similar memory traces. For example, distinct information about the source of a memory might not be available (anymore), giving rise to source confusion (Howe, 1998).

Memories might also become permanently associated with each other, giving rise to interference between them, and, in turn, false memories (e.g., Otgaar, Muris, Howe, & Merckelbach, 2017). This suggestion is for example taken up by the so-called *temporal context model*. According to this model, new and old memories are both represented by different memory traces. However, when both traces share some contextual features, for example because some old context elements were reinstated during the encoding of the new memory, the traces become associated with each other. Consequently, they become prone to interference which can, for example, result in source confusion and other memory errors (Sederberg et al., 2011; Sinclair & Barense, 2018).

1.4. Memory Modification – Flaw or Favorable?

What becomes clear is that episodic memories can be changed, even though the exact mechanism underlying this modification might not be completely clear yet. Often, memory modification is viewed as a shortcoming of our memory system, resulting from disturbances during encoding, storage, or retrieval (e.g., Schacter, 1999). In some cases, for example when false memories arise from an underlying pathology, like brain damage or mental illness (Conway & Loveday, 2015; Dodson & Schacter, 2002; El Haj, Colombel, Kapogiannis, &

Gallouj, 2020), this classification appears justified. But is this also always true for other types of false memories, like false recognition or intrusion of new information into a memory?

In fact, it has been argued that just because someone might experience unveridical memories, this does not mean the person's episodic memory system is defective (Cheng & Werning, 2016). Interestingly, patients experiencing amnesia due to temporal lobe damage make fewer false recognition errors than healthy individuals, indicating that to some extent, the occurrence of such false memories underlies a healthy memory system (Schacter, 2012; Schacter & Addis, 2007a). It has been proposed that many instances of memory distortions might actually be by-products of adaptive features of the memory system (Schacter, 1999, 2012), and some memory modifications themselves might have an adaptive value (Schacter & Addis, 2007b).

For example, one fundamental function of the memory system is the storage and retrieval of general similarity and gist information and the formation of schemas (Schacter & Addis, 2007b). This allows us to comprehend and categorize information, to transfer knowledge to different tasks, to generalize among contexts, and to interpret new information on the basis of previous experiences (Schacter, 1999, 2012). However, at the same time, memory errors like false recognition and false recall of gist-coherent items can arise as a side-effect of such an efficient generalization process (Schacter, 1999). Furthermore, modifying memories through existing schemas can even make memories more instead of less reliable, for example when the initial perception of an event was incomplete or distorted (Zacks, Bezdek, & Cunningham, 2021). Additionally, it seems to be a much more economically valuable solution to retain only the most important details of our experiences. As a side effect, this may lead to source confusions, as the exact context during which a memory was acquired might not be accessible (Schacter & Addis, 2007b). Aside from that, modifying memories in light of own beliefs and

attitudes might benefit the maintenance of a coherent and positive self-concept (Conway & Loveday, 2015), which can contribute to life satisfaction (Schacter, 1999).

Another fundamental function of episodic memories is to allow us to mentally simulate future events, so that we can prepare and plan ahead (Schacter, 2012; Schacter & Addis, 2007a). However, future events will never be exact copies of past events. Consequently, we need to be able to flexibly recombine available information to create imaginary simulations of the future (Schacter & Addis, 2007a, 2007b). As outlined above, remembering a past experience draws on the same constructive mechanism, just in a retrograde manner. Thus, the constructive flexibility that is needed for simulating the future can give rise to memory errors as a side effect (Schacter, 2012; Schacter & Addis, 2007a). These assumptions are further substantiated by the finding that common brain regions, for example MTL, subserve both, remembering the past and imagining the future (Schacter, 2012; Schacter & Addis, 2007b; Schacter, Addis, & Buckner, 2007).

Furthermore, it has been suggested that the only merit of retaining information about past experiences in the first place is to serve the anticipation of future events (Atance & O'Neill, 2005; Buckner & Carroll, 2006; Schacter & Addis, 2007b). In other words, memories are templates we can use to derive predictions from. Accordingly, the integration of new information into a memory can be interpreted in two ways: while incorporating misinformation is undesirable and can lead to confusion, including new information can be beneficial when it is relevant for successful future planning (Finn, 2017; Schacter, Guerin, & St. Jacques, 2011). In fact, many researchers argue that a key function of memory modifications, for example through reconsolidation, is to refine memory representations (De Oliveira Alvares et al., 2013), so that we can maintain valid predictions in an ever-changing environment (Alberini & Ledoux, 2013; Finn, 2017). Such refinements can occur in the form of appropriate strengthening of essential memory traces, or by incorporating new relevant information. The latter is often

referred to as *memory updating*[3] (e.g., De Oliveira Alvares et al., 2013), highlighting its proposed adaptive role.

This concept of modifying a memory to maintain its predictive relevance is highly reminiscent of another prominent framework in cognitive neuroscience: the *predictive coding* framework, which postulates the tuning of predictions through updating of internal generative models. To understand how the predictive coding framework might relate to memory modifications, the basics of this framework will be outlined in the following.

1.5 Predictive Coding and Episodic Memory

1.5.1 The Predictive Coding Framework

Traditional models of neural processing assume that the brain passively awaits sensory input which then travels bottom-up from lower to higher cortical areas in a feedforward manner (for reviews, see Friston, 2018; Nave, Deane, Miller, & Clark, 2020). In contrast, according to the predictive coding framework, the brain constantly predicts the most likely incoming sensory signals and their causes (Clark, 2015; Reichardt, Polner, & Simor, 2020), based on internal *generative models* (Clark, 2013). These generative models are probabilistic as they capture statistical regularities extracted from the environment (Clark, 2013; Nave et al., 2020; Sayood, 2018) through Bayesian inference (Friston, 2003). Using these models, the brain can compare predicted input with actual sensory stimulations, allowing it to focus on unpredicted, i.e., informative input, while fully predicted, i.e., non-informative, input can be neglected. Thus, the

[3] Many researchers use the term 'memory updating'. However, the term is used differently throughout literature, sometimes referring to the incorporation of new information into an old memory (e.g., Sinclair & Barense, 2018), the replacement of old memorized information (e.g., Ye, Shi, Li, Chen, & Xue, 2020), the weakening of old memory content (e.g., Kim, Lewis-Peacock, Norman, & Turk-Browne, 2014), or mnemonic benefits of newly acquired information, without interference of old and new memories (e.g., Bein, Plotkin, & Davachi, 2021; Wahlheim, Smith, & Delaney, 2019). Due to this ambiguity, I will preferentially use the term 'memory modification' instead of 'memory updating' as a broader description that a memory was influenced to avoid misunderstandings. Please note that the term 'memory modification', as it is used in this thesis, does not necessarily imply a change of original memory traces, but rather describes the measurable output of remembering.

use of limited computational and physiological resources in neural processing can be optimized (Huang & Rao, 2011; Sayood, 2018).

These comparative computations are accomplished by a hierarchical system of neuron subpopulations in the brain called *representation units* and *error units*. Representation units code for predictions, while error units produce a mismatch signal when these predictions are not met by incoming sensory inputs (Clark, 2013). This mismatch signal is known as *prediction error* (PE), i.e., the residual error of what could not be predicted by the current internal model (Clark, 2013; Huang & Rao, 2011; Nave et al., 2020). In this hierarchical processing cascade, error units receive inputs from representation units of higher (and their own) levels, while representation units are informed by error units of (their own and) lower levels (Clark, 2013; Stefanics, Kremláček, & Czigler, 2014). Error signals are conveyed via forward connections, so that PEs propagate from lower sensory to superior processing areas, while predictions are delivered through backward connections in a top-down manner (Clark, 2013). To render neuronal processing as resource-efficient as possible, the overall aim of generative models is to minimize prediction error signals, or, in other words, to optimally predict neuronal activity at the next lower level in the hierarchy. To achieve this, PE signals shape predictions at each level of the hierarchy until a good match of (sensory) input and predictions can be achieved. Thus, PEs serve as a quality measure for internal models, signaling the need for updates when predictions are not yet optimal. In summary, updating internal models via PEs allows for more accurate predictions in the future (Clark, 2013; den Ouden, Kok, & de Lange, 2012; Friston, 2005; Lupyan & Clark, 2015; Reichardt et al., 2020).

Within generative models, PE signals are further modulated by *second-order predictions* that predict the precision of the PE, corresponding to their reliability. Only precise PE signals should be used to update internal models (Friston, 2018; Nave et al., 2020). Thus, the reliability of PEs is estimated at each level of the processing hierarchy given the current

context. The outcome of this *precision weighting* process is used to flexibly assign more or less weight to PE signals from sensory inputs versus prior predictions in a way that suits the current situation (Friston, 2018; Lupyan & Clark, 2015; Nave et al., 2020). When the environment is characterized by high levels of noise or uncertainty, for example when navigating through our house in the dark, it is likely most useful to rely more on prior predictions, or, in this example, the knowledge about the house's architecture (Lupyan & Clark, 2015).

Importantly, not all PEs are alike. In reinforcement learning and motivational control, so-called *signed* PEs play a dominant role. These PEs signal the valence of an outcome, meaning whether it was better or worse than expected. In contrast, *unsigned* PEs, as they predominantly occur in perceptual inference, convey information about how *surprising* the presence or absence of an event is (den Ouden et al., 2012). To avoid unintended interpretations through the connotation of *surprise* as a subjective emotional reaction, one might rather use the information theoretical term *surprisal* which quantifies the informativeness of unexpected input in the context of PE (Barto, Mirolli, & Baldassarre, 2013; Clark, 2013; Tribus, 1961)[4]. Notably, PEs not only signal surprisal, but also carry representational content as they are linked to specific predictions (den Ouden et al., 2012). In other words, PEs not just convey the information *that* something was unpredicted, but also *what* was unpredicted. For this reason, it is possible to track PE signatures in the brain, using techniques like functional magnetic resonance imaging (fMRI) or electroencephalography (EEG). In fact, it has been demonstrated that different types of PEs elicit different brain activation, depending on the predictive task and incoming (sensory) information that needs to be processed (Bubic, von Cramon, Jacobsen, Schröger, & Schubotz, 2009; den Ouden et al., 2012; Gläscher, Daw, Dayan, & O'Doherty, 2010; Heins et al., 2020; Rao & Ballard, 1999).

[4] However, please note that both terms are often used interchangeably.

Even though the majority of evidence concerning the described mechanisms of predictive coding exists for visual processing (Friston, 2018), it has been proposed that predictive coding constitutes a "general computational strategy employed by the brain" (Huang & Rao, 2011, p. 591). While the classical view of predictive coding refers to current situational input, the framework can be extended to account for anticipated states, allowing for the prediction of unfolding sequences of actions (Schiffer, Ahlheim, Wurm, & Schubotz, 2012). In that sense, predictive coding provides a unifying framework for understanding the hierarchical organization of brain networks and a multitude of empirically observed neural phenomena (Huang & Rao, 2011). The influence of PEs has been documented in many domains, such as attention, motivation, and other cognitive processes, like learning and memory formation (den Ouden et al., 2012). In fact, generative internal models might be seen as equal to memories (Reichardt et al., 2020), which infers that updating internal models is synonymous to updating, or modifying, memories (Schubotz, 2015). This suggestion is supported by the proposed functions of episodic memory earlier in this thesis, namely predicting the future. In summary, this implication postulates a vital role of PEs in memory modification. Therefore, the next section of this thesis will focus on illuminating the influence of PEs on episodic memory.

1.5.2 The Influence of Prediction Errors on Episodic Memory

As can be assumed from the above section, PEs can lead to the formation of new memories but also the modification of old ones, suggesting some kind of dual role of PEs (Sinclair & Barense, 2019). Importantly, in both cases PEs drive the encoding of new information; either as a new memory or part of an old one (Krawczyk, Fernández, Pedreira, & Boccia, 2017).

Much research has focused on the role of PEs in forming new memories. In line with the proposed role of PEs in learning, a multitude of studies has demonstrated a subsequent mnemonic benefit for unexpected experiences (Bein, Plotkin, & Davachi, 2021; Brod,

Hasselhorn, & Bunge, 2018; Greve, Cooper, Kaula, Anderson, & Henson, 2017; Kafkas & Montaldi, 2018). Overall, there is much evidence that especially strong PEs lead to the formation of new memories (Bein et al., 2021; Brod et al., 2018; Greve et al., 2017; but see also Ortiz-Tudela et al., 2023; Turan, Ehrlich, Shing, & Nolden, 2023). Some also report an additional memory enhancement as a result of weak PE, which can be interpreted as a memory congruency effect (Brod, Werkle-Bergner, & Shing, 2013; Greve, Cooper, Tibon, & Henson, 2019; Quent, Greve, & Henson, 2022). It has been proposed that the encoding benefit following PE might be due to increased attention to the unexpected new information (Bein et al., 2021; Brod et al., 2018; Greve et al., 2017). In line with this suggestion, it was recently demonstrated that PE only benefits retention when participants become consciously aware of the unexpected change and can later recall it (Wahlheim & Zacks, 2019).

In contrast to strong (and potentially weak) PE, it has been suggested that moderate levels of PE might weaken or destabilize an original memory, potentially opening a door for memory modification (Milton, Das, & Merlo, 2023; Sevenster, Beckers, & Kindt, 2014). These observations would be in line with the so-called *non-monotonic plasticity* hypothesis, which assumes that while low and high levels of reactivation (e.g., via PE) promote better memory, moderate reactivation has the reverse effect (Ritvo, Turk-Browne, & Norman, 2019; Sinclair & Barense, 2019). However, others have reported that strong PEs are most potent in modifying memory, e.g., by the weakening of original memory associations (Kim, Lewis-Peacock, Norman, & Turk-Browne, 2014) or by promoting memory intrusions (Sinclair & Barense, 2018). Irrespective of their exact strength, it is well established that in order to trigger memory modification, PEs must overcome a certain threshold (Exton-McGuinness et al., 2015). Furthermore, the degree to which memories can be modified following PE might be influenced by additional conditions, such as the initial memory strength (Exton-McGuinness et al., 2015; R. S. Fernández et al., 2016).

What several explanatory accounts of false memory formation as presented above (section 1.3) have in common is the observation that the reactivation of an old memory can cause memory modification, for example in a sense that new information becomes part of the original memory (Finn, 2017; Xue, 2022). It has been proposed that whether or not such reactivations lead to memory changes critically depends on the presence of PE (e.g., in the form of new information presented during reactivation), making it a necessary prerequisite for memory destabilization (Exton-McGuinness et al., 2015; R. S. Fernández et al., 2016). Accordingly, R. S. Fernández and colleagues state that "an effective reminder capable of labilizing memory is one which elicits a PE signal" (R. S. Fernández et al., 2016, p. 436), and several studies showed that PE is a requirement for initiating reconsolidation (e.g., Exton-McGuinness et al., 2015; R. S. Fernández et al., 2016; Pedreira, Pérez-Cuesta, & Maldonado, 2004; Sinclair & Barense, 2018). In fact, much traditional work on memory modification and false memories seems to be in accordance with this proposed role of PE as well: misinformation items or new associations are often delivered in a way so that they conflict before encoded material (Finn, 2017), i.e., the new piece of information violates before established predictions. Interestingly, the two roles of PE do not need to operate in a mutually exclusive manner: while PE can foster the retention of correct details, it can at the same time increase the introduction of false details in the same memory (Sinclair, Manalili, Brunec, Adcock, & Barense, 2021).

Taken together, there is much evidence that PEs can trigger memory modification[5]. However, findings on how exactly memory is influenced remain mixed and investigations about naturalistic episodic memory in humans are sparse. For example, a straightforward analysis of different types of mnemonic PEs and their impact on memory is, to my knowledge, still missing, as PEs are usually operationalized in only one way within the same study or paradigm. Also,

[5] Importantly, this thesis does not postulate that PE is the driving force of *all* types of incorrect remembering. Rather, it is suggested that in some situations, PE might trigger the modification of episodic memories by serving as a signal when to incorporate new information of predictive relevance.

insights into further modulating factors, like the influence of memory solidity, are still lacking. Nevertheless, some research has started to explore potential neural underpinnings of mnemonic PEs, as outlined in the following.

1.5.3 Neural Underpinnings of Mnemonic Prediction Errors

Even though there seems to be general agreement that PEs can induce memory modification, little is known about neural correlates of such mnemonic PEs. While several studies demonstrated that the hippocampal complex responds to mnemonic PEs (Bein, Duncan, & Davachi, 2020; Chen, Olsen, Preston, Glover, & Wagner, 2011; Duncan, Ketz, Inati, & Davachi, 2012; Kumaran & Maguire, 2006, 2007; Long, Lee, & Kuhl, 2016), establishing its role as a mismatch detector, the possible contribution of other brain areas has only sparsely been illuminated.

Another interesting candidate area is PFC, which might provide predictive signals in the context of episodic memories (Schubotz, 2015), and is believed to mediate PE in scenarios with higher cognitive demands (den Ouden et al., 2012). Evidence for this has been delivered by Schiffer and colleagues, who, additionally to PFC, identified PE signals in several parietal areas and caudate nucleus (Schiffer et al., 2012; Schiffer & Schubotz, 2011).

Furthermore, it was observed that, in line with the predictive coding framework, mnemonic PEs are processed dynamically by the brain. For example, it was demonstrated that when the same unexpected input was repeated, brain activation in hippocampus as well as other PE-sensitive brain areas decreased, demonstrating the updating or adaptation of the internal model due to accumulating evidence for an alternative (Schiffer, Ahlheim, Ulrichs, & Schubotz, 2013; Schiffer et al., 2012). Additionally, when the internal model was especially strong due to much pre-exposition with encoded episodes, PE signals in parahippocampal cortex were more pronounced compared to weaker models (Schiffer, Ahlheim, et al., 2013).

It is striking that many of these results were achieved using simple associative mismatch paradigms like learning of word or image pairs (e.g., Chen et al., 2011; Long et al., 2016), or even memories of static images (Bein et al., 2020; Duncan et al., 2012), which most likely do not capture the naturalistic complexity of episodic memories. The studies by Schiffer and colleagues (Schiffer, Ahlheim, et al., 2013; Schiffer et al., 2012) provide a promising starting point for the development of a naturalistic PE-dependent episodic modification paradigm based on their use of complex, authentic everyday-life episodes. However, Schiffer and coworkers did not test for a behavioral impact of PEs on episodic memory and the dynamic brain responses they reported did not differentiate between the establishment of true and false memories. This distinction seems to be highly relevant, since, for example, it was recently found that increased hippocampal activation in response to mnemonic PE can predict the formation of false memories (Sinclair et al., 2021). Furthermore, none of the existing research has done justice to the multifaceted nature of episodic memories which likely inform different types of predictions. For example, based on the predictive coding framework, it would be expected that different types of mnemonic PEs in the context of episodic memory elicit different brain responses, which can potentially be further modulated by factors like memory (or model) solidity. To this end, the present thesis aimed to provide further evidence on the behavioral and neural implications of mnemonic PEs by targeting the research questions presented in the following.

2 Research Questions

Episodic memories have been studied for many decades, and we already know much about how the brain accomplishes their encoding, consolidation and retrieval. Interestingly, episodic memories are usually not exact, veridical reconstructions of past experiences, and several mechanisms have been proposed to underlie memory changes. Understanding mechanisms by which memories can be modified is a topic with important implications for our everyday-life, for example in legal and clinical settings, and it can help us understand how our memory systems operate. While some memory distortions may be undesirable, it was postulated that modifying memories has an important adaptive value: allowing us to maintain valid predictions in an everchanging environment. In line with the predictive coding framework, such memory modifications might be triggered by mnemonic PEs that arise when there is a mismatch between predictions that we derive from memories, and actual sensory input.

From the evidence presented above, it becomes clear that research on neural processing of PEs in episodic memory is still in its infancy. While some findings suggest that mnemonic PEs recruit hippocampal complex and PFC, two core structures in episodic memory, it remains to be elucidated how different types of mnemonic PEs are processed by the brain and which conditions further modulate the influence of PEs on brain activity and episodic memory. To deepen the understanding of PEs in episodic memory, the thesis targeted the following research questions:

1. How are different types of episodic PEs (either depending on the type of episodic information or their relevance within the episode) processed by the brain and how do they influence episodic memory?

2. How do different memory consolidation approaches influence neural processing of episodic memories, especially in the case of mnemonic PEs, and memory performance?

3. Which (dynamic) brain activations in response to mnemonic PEs characterize memory modification?

Using the experiments conducted by Schiffer and colleagues (Schiffer, Ahlheim, et al., 2013; Schiffer et al., 2012) as a starting point, a naturalistic episodic modification paradigm was developed. Two initial pilot studies were conducted to establish a set of episodes and to define an experimental protocol to allow for their successful encoding. In the final paradigm, participants encoded complex, unique episodes in the form of toy stories from demo videos. Thus, episodes were minimally influenced by common everyday activities which allowed for controlled encoding in the laboratory. During encoding, a high level of self-involvement by the participants was ensured to approach the natural circumstances under which memories for personally experienced events are formed. In **Study 1** and **Study 2**, the encoding phase was followed by a retrieval phase to promote additional memory consolidation. During a subsequent fMRI session, participants were presented subtly modified versions of before encoded episodes to elicit mnemonic PEs. Afterwards, they completed a memory test during which memory for originally encoded episodes was probed.

In **Study 1**, the focus was on characterizing brain responses for mnemonic PEs that challenged different types of episodic information, namely their structure and content. While it was expected that both types of PEs elicit some common neural activation, it was also hypothesized that PEs should elicit brain activation in areas that are responsible for specifically processing either type of incoming information, in line with the predictive coding framework.

Further, it was expected that the repeated experience of mnemonic PEs would trigger the modification of episodic memories.

In **Study 2**, which was based on the same data as **Study 1**, the influence of different consolidation protocols on neural activation and episodic memory was probed. Further, it was tested which dynamic brain responses give rise to later false memories. For that, fMRI data was analyzed based on later false alarms or correct rejections for modified episodes. Specifically, it was tested in which brain areas activation increased over the course of multiple PE experiences to give rise to false alarms with a focus on the hippocampal complex.

Study 3 aimed for a more specific investigation of the aforementioned content-based PEs from **Study 1**. For that, a new set of episodes was established in a further pilot study so that content modifications differed regarding their meaningfulness for the overall storylines of episodes. Two conditions were established: modifications that did not challenge the storyline of episodes (so-called surface modifications), and modifications that did (so-called gist modifications). It was expected that the two types of modifications differ regarding PE-related brain activation and their influence on subsequent episodic memories.

3 Research Articles

3.1 Study 1: What Happened When? Cerebral Processing of Modified Structure and Content in Episodic Cueing

Running title: Modified Structure and Content in Episodic Cueing

Sophie Siestrup, Benjamin Jainta, Nadiya El-Sourani, Ima Trempler, Moritz F. Wurm, Oliver T. Wolf, Sen Cheng, & Ricarda I. Schubotz (2022)

Journal of Cognitive Neuroscience, *34*, 1287–1305

What Happened When? Cerebral Processing of Modified Structure and Content in Episodic Cueing

Sophie Siestrup[1,2], Benjamin Jainta[1], Nadiya El-Sourani[1], Ima Trempler[1,2], Moritz F. Wurm[3], Oliver T. Wolf[4], Sen Cheng[4], and Ricarda I. Schubotz[1,2]

Abstract

■ Episodic memories are not static but can change on the basis of new experiences, potentially allowing us to make valid predictions in the face of an ever-changing environment. Recent research has identified prediction errors during memory retrieval as a possible trigger for such changes. In this study, we used modified episodic cues to investigate whether different types of mnemonic prediction errors modulate brain activity and subsequent memory performance. Participants encoded episodes that consisted of short toy stories. During a subsequent fMRI session, participants were presented videos showing the original episodes, or slightly modified versions thereof. In modified videos, either the order of two subsequent action steps was changed or an object was exchanged for another. Content modifications recruited parietal, temporo-occipital, and parahippocampal areas reflecting the processing of the new object information. In contrast, structure modifications elicited activation in right dorsal premotor, posterior temporal, and parietal areas, reflecting the processing of new sequence information. In a post-fMRI memory test, the participants' tendency to accept modified episodes as originally encoded increased significantly when they had been presented modified versions already during the fMRI session. After experiencing modifications, especially those of the episodes' structure, the recognition of originally encoded episodes was impaired as well. Our study sheds light onto the neural processing of different types of episodic prediction errors and their influence on subsequent memory recall. ■

INTRODUCTION

Episodic memories enable us to vividly relive events that we experienced at some point in our personal life (Tulving, 2002). However, there is evidence that they are not always veridical reconstructions of our past (Lee, Nader, & Schiller, 2017; Scully, Napper, & Hupbach, 2017; Nader, 2015; Nader & Einarsson, 2010). Situations we encounter in everyday life are usually not exactly the same as those we experienced before. So, there is always a certain discrepancy between our expectations, which we derive from our memories, and the new events we experience. According to the predictive coding framework, this discrepancy leads to a prediction error (Reichardt, Polner, & Simor, 2020; Barto, Mirolli, & Baldassarre, 2013). Prediction errors serve as bottom–up learning signals that allow us to adapt our internal predictive models to an ever-changing environment to maintain valid predictions in the long run (Schubotz, 2015; Friston & Kiebel, 2009; Friston, 2005). According to this view, it is adaptive that memories are modified in favor of valid internal models informed and updated by later experiences (Fernández, Boccia, & Pedreira, 2016; Exton-McGuinness, Lee, & Reichelt, 2015). Evidence accumulates that mnemonic prediction errors are important drivers of memory change (Sinclair & Barense, 2019), and researchers recently begun to address the question how mnemonic prediction errors are processed by the brain (e.g., Bein, Duncan, & Davachi, 2020; Kim, Lewis-Peacock, Norman, & Turk-Browne, 2014).

The aim of this study was to characterize neural responses to different types of mnemonic prediction errors during episodic retrieval by targeting two basic types of episodic memory information: either their content ("what") or their structure ("when"; cf. Griffiths, Dickinson, & Clayton, 1999). To do so, we adapted a previously developed episodic cueing paradigm (Jainta et al., 2022; Schiffer, Ahlheim, Ulrichs, & Schubotz, 2013; Schiffer, Ahlheim, Wurm, & Schubotz, 2012). After encoding short episodes from videos and consolidating memories in two further retrieval sessions, participants went through a fMRI session and were either presented original episode videos or slightly modified versions thereof. To create the latter, a subset of videos was manipulated with regard to the occurrence of an object (content modification) or the order of two consecutive action steps (structure modification) to elicit different types of mnemonic prediction errors (see Figure 1 for an example). In a post-fMRI memory test, participants' memory for original and modified episodes was probed.

[1]University of Münster, Germany, [2]Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany, [3]University of Trento, Italy, [4]Ruhr University Bochum, Germany

Journal of Cognitive Neuroscience 34:7, pp. 1287–1305
https://doi.org/10.1162/jocn_a_01862

Although today many agree that prediction errors drive memory modification during episodic retrieval (Barron, Auksztulewicz, & Friston, 2020; Fernández et al., 2016; Kim et al., 2014), there is no unifying model from which neuroanatomical hypotheses can be derived. Based on existing studies, we expected that some regions might be engaged in episodic prediction errors in general, whereas others would be engaged only in content or structure prediction errors.

As to the former, the medial frontal cortex may serve more general control over consolidation and retrieval of long-term memories (Peters, David, Marcus, & Smith, 2013; Euston, Gruber, & McNaughton, 2012). Furthermore, the hippocampus is regarded a core structure of episodic memory (Horner & Doeller, 2017; Stachenfeld, Botvinick, & Gershman, 2017; Maguire, Intraub, & Mullally, 2016) and responds to mnemonic prediction errors (Bein et al., 2020; Long, Lee, & Kuhl, 2016). In addition to these common neural responses to episodic surprise, structure and content episodic modifications were expected to engage different brain regions. Structure modifications should elevate activity in premotor areas because of their central role in sequential order processing (Schubotz, 2004). More specifically, dorsal premotor and adjacent prefrontal sites along the superior frontal sulcus (SFS; dorsal premotor cortex [PMd]) were found for stepwise ordinal linking of individual action or event steps, as required in different predictive tasks (Pomp et al., 2021; Hrkać, Wurm, & Schubotz, 2014; Schubotz, Korb, Schiffer, Stadler, & von Cramon, 2012; Stadler et al., 2011; Tamber-Rosenau, Esterman, Chiu, & Yantis, 2011; Kurby & Zacks, 2008). By contrast, content modifications were expected to engage areas related to object processing, including lateral occipitotemporal cortex (OTC; Lingnau & Downing, 2015), anterior intraparietal sulcus (IPS; Schubotz, Wurm, Wittmann, & von Cramon, 2014; Creem-Regehr, 2009), and fusiform gyrus (FG; Reber, Gitelman, Parrish, & Mesulam, 2005).

If new content and/or structure information induced updating of the original predictive model during fMRI, as expected, this should also reduce memory accuracy in a post-fMRI memory test (Jainta et al., 2022; Schiffer et al., 2012, 2013). We thus expected a weakening of the original episodic memory, that is, false rejections of original videos as new, and/or the creation of alternative episode representations, that is, false acceptances of modified videos as originals.

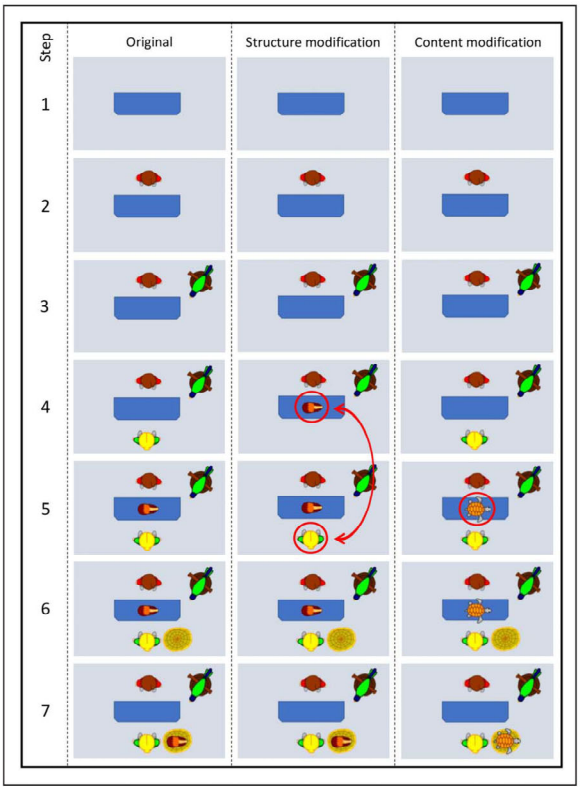


Figure 1. Example of an original episode and its modified versions, shown by the sequence of the main event steps. Twenty-four stories existed in three different versions each: an original, a structure modification, and a content modification. For the structure modifications, two adjacent action steps were switched compared with the original. In this example, the original shows the blonde woman join the scene before the guinea pig is positioned on the sales counter; in the structure modification, the guinea pig appears before the blonde woman (red circles). For the content modifications, an object was exchanged compared with the original (here: tortoise instead of guinea pig on the sales counter in Step 5). Note that in the fMRI experiment, each participant was only presented with one of the three versions of a story. We do not reproduce photos of our stimulus material because it is copyrighted material (PLAYMOBIL figures); instead, we provide schematic images.

METHODS

Participants

Forty-five women took part in the study. Participants had (corrected-to-) normal vision, were native German speakers, and were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). As in our previous study (Jainta et al., 2022), participants were all female to achieve a good match between the hands in the videos and the hands of the participants. They reported no history of neurological or psychiatric disorders or substance abuse. Four participants started the experiment but did not finish, either because of technical problems during the second retrieval session (three participants) or personal reasons (one participant). Data from five additional participants were excluded from analyses because of the incorrect presentation of video stimuli during the fMRI session (one participant) and increased movement during the fMRI session (four participants, approximately 5-mm movement). Consequently, 36 participants were part of the final sample (M = 22 years, SD = 2.78 years, range = 18–30 years). Similar sample

sizes have yielded stable results in our previous work (e.g., Jainta et al., 2022; Pomp et al., 2021; El-Sourani, Trempler, Wurm, Fink, & Schubotz, 2019). Participants received course credits or money for their participation and gave written informed consent to participate in this study. The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the University of Münster.

Stimuli

We used the same set of videos as previously reported (Jainta et al., 2022; available upon request at https://www.uni-muenster.de/IVV5PSY/AvicomSrv/). These were 78 short videos (duration = 8.80–17.88 sec, M = 12.71 sec) of stories that were played with PLAYMOBIL toys, showing only the toys and hands and underarms of an actress. Stories comprised six to nine action steps (M = 7.4 steps) and 4–14 separable objects (M = 6.93 objects), such as characters, animals, vehicles, and tools. The same object appeared in only one of the stories.

Stories were filmed from above with a digital single-lens reflex camera (Nikon D5300), which was centrally mounted above the table and faced straight down. Matte white paper served as a base. A frame of 47.5 cm × 28 cm was taped on the paper, congruent with the section captured by the camera (in the following referred to as camera frame). Objects that were needed for a particular story were positioned next to the camera frame and were only moved into view in the moment at which they appeared in the story. During filming, the actress wore a black pullover and black rubber gloves. To facilitate future imitation from demo videos, the back of the right hand was marked with a yellow dot (Franz, Ford, & Werner, 2007). Video material was edited using Adobe Premiere Pro CC (Adobe Systems Software, Version 12.1.2). All videos had a frame of size 1920 × 1080 pixels and a frame rate of 25 frames per second. Videos started with seven frames showing only background and ended after seven frames showing the final toy constellation. Throughout the experiment, videos were presented at a visual angle of approximately 7.3° × 13° using Presentation software (Version 20.3 02.25.19, NeuroBehavioral Systems).

On the basis of two pilot studies, we chose 24 out of originally 30 stories for our stimulus set. Stories were excluded when they were particularly difficult to imitate or describe. One of the 30 stories was excluded because of low memorability as indicated by low performance in a signal detection task.

The 24 final stories existed in three different versions each: (1) an original version as encoded by the participants, (2) a version in which two adjacent action steps were switched (structure modification), and (3) another variation of the original video in which one object was exchanged (content modification). Story scripts were created by five experimenters who all had to agree that the original story and modifications thereof were semantically valid (within a toy world) and that modifications did not change the overall outcome of the story. For creating videos with modifications, the respective stories were played and filmed again exactly the way as for the original video. The only aspect that differed between original and modified versions was a single change of either the order of two action steps (i.e., one transition out of 7.33 transitions, on average, for structure modifications) or one object (i.e., one object out of 6.95 objects, on average, for content modifications).

Modifications were never introduced in the first two action steps so that the beginning of a video served as a cue for prediction. Furthermore, no modifications were introduced in the last two action steps, either. The exact time point of the modification in each video was determined by identifying the video frame that diverged from the original version. For an example of an episode and its modified versions, see Figure 1.

Six other stories were used in one version only. Four of them were presented for the first time in the fMRI session, we refer to them as novel episodes in the following. The two remaining videos were only used for practice and did not appear in the fMRI experiment and memory test.

Procedure

Encoding

Encoding sessions were conducted in a computer laboratory at the Department of Psychology at the University of Münster and followed our previously reported protocol, with some modifications (Jainta et al., 2022). The encoding consisted of two sessions that took place on two consecutive days and lasted about 2 and 1.5 hr, respectively. During each of the two sessions, participants encoded half of the episodes. We chose to split the training over 2 days to avoid fatigue or a decrease in motivation because of the relatively long duration of the task.

The 24 demo videos were organized in four subsets, containing six videos each, balanced for the number of action steps (A1, A2, B1, B2). On each day, participants encoded one A and one B subset. This means that each participant encoded each video either during Session 1 or during Session 2; the same video was not encoded on both days. Which subsets were trained in which session was balanced over participants. The order in which episodes were encoded was randomized for each participant. The first session started with two practice videos to familiarize participants with the task.

During encoding, participants sat at the same setup that had been used for filming the stimulus material and likewise wore a black pullover and gloves with a yellow dot on the right hand, so that they could be filmed while playing the stories themselves. The experimenter sat opposite of the participant, supervising the performance. For each story, the toys were positioned next to the camera frame,

following the same arrangement as used while creating the stimulus material (Figure 2). Each episode video was presented 3 times from the first-person perspective. Then, participants had to imitate each story correctly 3 times. After imitation, participants had to deliver a detailed description of the story to ensure that they understood it correctly and had paid attention to all objects involved. If participants made a mistake during an imitation or description trial, they were immediately interrupted by the experimenter to avoid encoding of incorrect scripts. They would then start over with a new imitation/description attempt. On average, participants only performed one incorrect imitation attempt (M = 1.04, SD = 0.47).

Retrieval Sessions

To further consolidate episodic memories, participants went through two additional sessions during which they completed an active retrieval task of before encoded episodes. Active retrieval is known to aid memory consolidation and improve retention (Rowland, 2014). The first consolidation session took place on the day after the second encoding session. The second session was conducted approximately 1 week later (range = 4–8 days; M = 6.36 days, SD = 0.93 days).

Participants always watched the first two steps of a demo video. Then, the video stopped and a question was displayed below the still video frame, which either read "Left?" or "Right?". The participants' task was to visualize the rest of the story from memory the way they had performed it and then answer how many steps of the entire story had been performed with the left or right hand, respectively. They answered by pressing a number key (0–9) on their keyboard. Upon response delivery, the video played until the end and participants were instructed to carefully watch the video to self-check their response. Afterward, written feedback ("correct," "incorrect") was provided for 1.5 sec. When feedback is included in retrieval tasks, consolidation has been shown to occur independent of initial retrieval success (Rowland, 2014; Roediger & Butler, 2011). We chose this task because we wanted to encourage active retrieval of the encoded episodes without laying a special focus on aspects that would be modified during the fMRI session. Importantly, the number of steps that were conducted with the left or right hand did not change in videos containing either type of modification. The task was self-paced and started with two practice videos during which the experimenter carefully checked whether the participant had understood the task correctly.

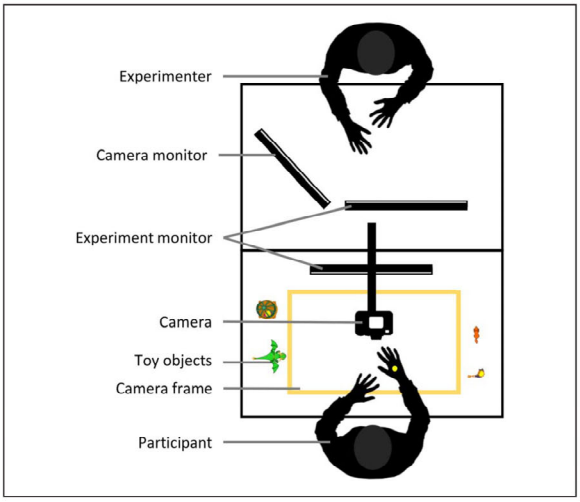


Figure 2. Encoding setup. During encoding, participants imitated toy stories from demo videos, while sitting at the filming setup. Their performance was monitored by the experimenter.

During the retrieval sessions, we established the two experimental factors consolidation Times and Schedule. To this end, half of all episodes were consolidated 2 times in total, the other half 8 times (factor Times). Furthermore, half of the stories were consolidated during both separate sessions, that is, in a spaced manner, while the other half of the stories were consolidated only in Session 2, that is, in a massed manner (factor Schedule). Participants were explicitly instructed to always visualize the story and not just remember previous responses they gave. Each type of question was presented equally often after each video and per session. At the end of Session 2, participants went through a short practice (four video trials, four question trials, one null event) of the task they would conduct during the fMRI scan.

The two described experimental factors consolidation Times (2, 8) and Schedule (spaced, massed) are not further addressed in this article, as they are central for a companion paper (Siestrup, Jainta, Trempler, Cheng, & Schubotz, in preparation) describing the influence of different consolidation strategies on brain activation during episodic recall. Both factors were balanced with respect to the factors reported here, so we can exclude any confounding effects (fully crossed design). Correct answer rates in the retrieval task did not differ significantly between episodes, which were later presented in the original (ori), structure modified (str), or content-modified (con) version (M_{ori} = .854 ± .013; M_{str} = .846 ± .012; M_{con} = .852 ± .008; $F(2, 70) = .314$, $p = .732$, $\eta p^2 = .009$), so that we can rule out that any confound was introduced through the retrieval sessions.

fMRI Session

The fMRI session took place approximately 1 week after the second retrieval session (range = 6–13 days; M = 7.69 days, SD = 1.31 days) and was conducted as previously described (Jainta et al., 2022). Participants were told

that videos of themselves playing the stories would be presented in the fMRI session. Although participants had actually been filmed during encoding, these videos were not used during the fMRI experiment. This was only a cover story to elevate personal identification with the videos to benefit episode reactivation. We previously confirmed in a pilot study that this cover story works as intended and already applied it successfully in our previous fMRI study (Jainta et al., 2022). Participants were fully debriefed after completion of the study.

During the fMRI session, participants were presented with original and modified videos reminiscent of the previously encoded episodes. Importantly, each video was only shown in the original or one divergent version. Following a previously used paradigm (Schiffer et al., 2012, 2013), modified and original episodes were presented repeatedly to simulate the natural circumstances that potentially foster memory modification, that is, updating of internal models because of increasing evidence for the validity of an alternative. Thus, eight videos were repeatedly presented in the original version; eight included a structure modification; and eight, a content modification. Which stories belonged to which conditions varied between participants. In addition, four novel stories were included in the fMRI session.

The fMRI experiment consisted of six blocks, each containing the 24 videos reminiscent of the previously encoded episodes. Consequently, each video was presented 6 times over the course of the session. Within blocks, videos were presented in pseudorandom order so that transition probabilities between conditions were balanced. In addition, each block contained three null events during which only a fixation cross was presented (duration: 7–10 sec). Furthermore, each novel video was presented once per block. Therefore, the whole experiment contained 18 null events and 24 novel video trials. Participants were not informed about the block structure of the experiment.

Participants were instructed to attentively watch the presented videos. They were told that after some videos, a short description would be presented (e.g., "Rescuing princess") that either matched or did not match the story shown in the video (question trials). The task was to either accept or reject the description by pressing one of two buttons on a response box with the right index or middle finger, respectively. This type of task has been used successfully before to focus participants' attention on complex video stimuli (Jainta et al., 2022; El-Sourani et al., 2019). Importantly, neither type of modification influenced the overall outcome of episodes so that all descriptions used as questions were valid for all episode versions. Questions never highlighted any type of modification (Figure 3). Throughout the entire experiment, each story was once followed by a matching description and once by a nonmatching description, resulting in a total number of 56 question trials in the experiment. Each block contained 9–10 question trials and, per block, approximately 50% of descriptions were to be accepted, and 50% were to be rejected. The question was presented for a maximum of 3 sec or until participants responded. Upon response delivery, participants received a 1-sec written feedback whether they answered correctly, incorrectly, or too late, in case no response was given. Participants were naive with regard to this distribution of question trials.

Between trials, a fixation cross was presented for a duration of 2 sec (1 sec after question trials) to serve as an interstimulus interval. Before each trial, a variable jitter of 0, 0.5, 1, or 1.5 sec of fixation was added for enhancement of the temporal resolution of the BOLD response (Figure 3). In total, the fMRI task had a duration of approximately 48 min.

Post-fMRI Memory Test

Immediately after the fMRI session, participants completed an explicit memory test as described previously (Jainta et al., 2022). Importantly, encoding occurred incidentally, as participants were not informed beforehand that their memory for episodes would be tested.

Participants were seated in a separate room in front of a laptop and instructed to remember their encoding sessions 2 weeks prior during which they had played the stories themselves. They were presented all stories that they had seen in the fMRI session in two different versions. More precisely, when modified videos had been presented during the fMRI session, these modified videos were presented again during the memory test and additionally each story was shown in the original version. When original episode videos had been presented during the fMRI scan, these original videos were presented again in the memory test and, additionally, each story was shown in a modified version, either containing a structure modification in half of the cases or a content modification.

The participants' task was to rate after each video whether they knew this exact episode from the encoding sessions, using a Likert scale including 1 (*yes*), 2 (*rather yes*), 3 (*rather no*), and 4 (*no*), by pressing one out of four marked keys on the laptop's keyboard. Similar rating schemes have previously been used in memory research (Jainta et al., 2022; Kim et al., 2014). Response time was not restricted, but participants were instructed to respond quickly and intuitively. Videos were presented in a pseudorandomized order, so that half of the stories (of each experimental condition) were first presented in their original version followed by a modified version and vice versa. Novel videos were shown twice in the same version, so that, in total, the memory test comprised 56 video trials. The completion of the task took approximately 15 min.

MRI Data Acquisition and Preprocessing

MRI scans were conducted with a 3-Tesla Siemens Magnetom Prisma MR tomograph using a 20-channel head coil. Participants lay supine on the scanner bed with their right

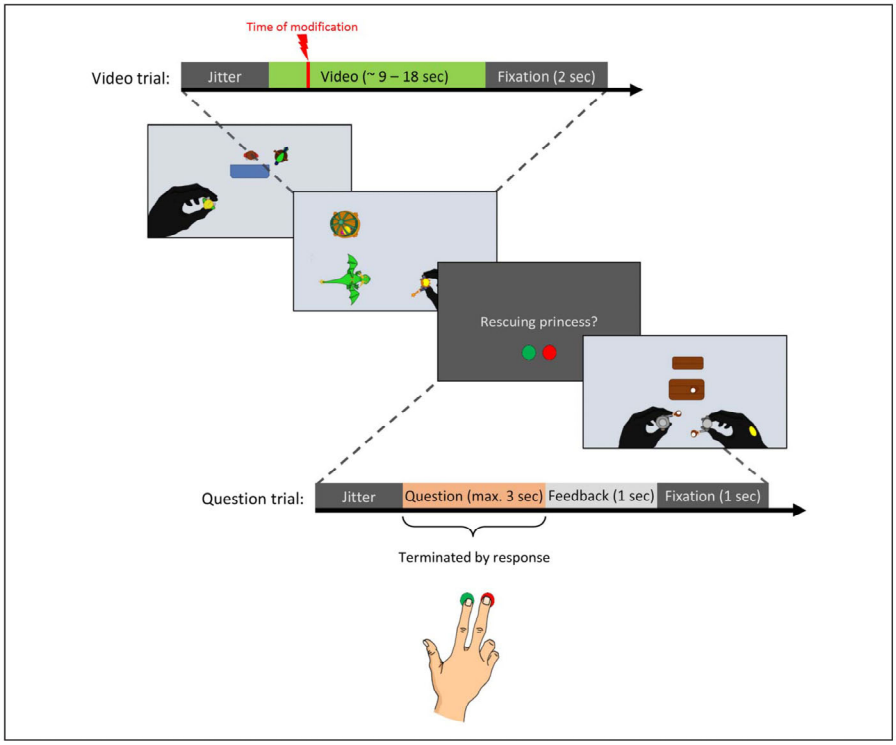


Figure 3. Schematic depiction of task during fMRI session. Video trials consisted of a variable jitter (0, 0.5, 1, or 1.5 sec of fixation), a video showing a toy story (ca. 9–18 sec) and a 2-sec interstimulus interval (fixation). Question trials included a variable jitter, a question regarding the story shown in the preceding video (maximally 3 sec long or terminated by response), and a 2-sec interstimulus interval. The interstimulus interval after question trials was divided into a 1-sec feedback ("correct," "incorrect," "too late") and a 1-sec fixation. Aside from the question, it was depicted which button should be pressed to accept (left, green) or reject (right, red) the description. For each modified video, we determined the exact video frame during which the modification occurred (time of modification), which we used to precisely model modification-related brain activation. For original and novel videos, comparable time points were chosen.

index and middle finger positioned on the two appropriate buttons on a response box. Head, arm, and hand movements were minimized by tight fixation with form-fitting cushions. Participants were provided with earplugs and headphones to attenuate scanner noise. Stimuli were projected on a screen that the participants saw via an individually adjusted mirror, which was mounted on the head coil.

High resolution T1-weighted anatomical images were obtained with a 3-D multiplanar rapidly acquired gradient echo sequence before functional imaging. One hundred ninety-two slices with a thickness of 1 mm were acquired, using a repetition time of 2130 msec, an echo time of 2.28 msec, a flip angle of 8°, and a field of view of 256 × 256 mm^2. Functional images of the whole brain were acquired in interleaved order along the anterior commissure–posterior commissure plane using a gradient-echo EPI sequence to measure BOLD contrast. Thirty-three axial slices with a thickness of 3 mm (voxel size 3 mm^3) were obtained, using a repetition time of 2000 msec, an echo time of 30 msec, a field of view of 192 × 192 mm^2, and a flip angle of 90°.

Processing of imaging data was conducted with SPM12 (Wellcome Trust) implemented in MATLAB (Version R2020b, The MathWorks Inc.). Data were preprocessed by slice time correction to the middle slice, movement correction and realignment to the mean image, coregistration of the functional data to individual structural scans, normalization of functional and structural images into the standard Montreal Neurological Institute (MNI) space on the basis of segmentation parameters, and spatial smoothing using a Gaussian kernel of FWHM of 8 mm. Furthermore, a 128-sec high-pass temporal filter was applied.

Statistical Data Analysis

fMRI Design Specifications

Statistical analyses of the fMRI data were conducted with SPM12. We used a general linear model (GLM) for serially autocorrelated observations (Worsley & Friston, 1995; Friston et al., 1994) and convolved regressors with the canonical hemodynamic response function. Regressors were original videos (ori), videos containing a structure modification (str), and videos containing a content modification (con), each comprising 48 trials. For str and con trials, the onsets of events were time-locked to the point in the video at which the modification occurred (time of modification). For ori trials, we calculated a hypothetical time of modification (mean of times that corresponded to points of structure and content modification in the non-modified video) to serve as a comparable onset. These conditions were modeled as events as we were interested in the phasic effect of the prediction violation at the precise moment it occurred. To each of those regressors, we added a parametric modulator to model the repeated

presentation of each video. The 24 novel videos were modeled as events as well, with onsets timed to the middle of the video. Two additional regressors modeled the 18 null events and the 56 question trials. The modeled activation of null events and questions was time-locked to their respective onsets. Null events were modeled as epochs, containing their full presentation time (7–10 sec), whereas questions were modeled as events. The six subject-specific rigid-body transformations obtained from realignment were included as regressors of no interest. Therefore, the GLM comprised 15 regressors in total.

As a first step, we calculated first-level *t*-contrasts for str > ori and con > ori as well as the direct contrasts str > con and con > str to analyze brain activity in response to the specific modification types. In addition, we calculated the first-level *t*-contrasts for each condition versus novel (nov) videos (ori > nov, str > nov, con > nov). We used this approach to demonstrate successful retrieval of encoded episodes (Jainta et al., 2022) and to validate that brain responses to episodic modifications were qualitatively different from novelty responses. A conjunction of str > ori and con > ori contrasts was calculated to detect shared effects of both modifications (Nichols, Brett, Andersson, Wager, & Poline, 2005). As an additional, more liberal approach to detect shared activation, we aggregated str and con modified (mod) videos to calculate the contrast mod > ori. Gray matter masking was applied on the first level of the analysis. For masking, we used the smoothed individual normalized gray matter image (8-mm FWHM), which was thresholded at .2 using ImCalc in SPM12 to create a binary mask. Second-level group analyses were performed by using one-sample *t* tests across participants. We applied a threshold of $p < .001$ on the whole-brain level and then used false discovery rate (FDR) correction at $p < .05$ on the cluster level to correct for multiple comparisons. Brain activation patterns were visualized with the software MRIcroGL (Version 1.2.20200331, McCausland Center for Brain Imaging, University of South Carolina).

To deepen our understanding of how prediction errors contribute to memory modification, we constructed a second GLM in which we split the str and con regressors into later false alarms and correct rejections in the post-fMRI memory test. The other regressors were the same as for the other GLM, but no parametric modulators were included. We contrasted false alarms with correct rejections, separately for each modification type, to investigate whether we can identify brain activation that predicts later false memories. However, this analysis did not yield any significant results.

Behavioral Data Analysis

The behavioral data analysis was conducted using RStudio (R Core Team, 2020; Version 1.3.1073).

To test our hypothesis that repeated presentations of modified videos in the fMRI session lead to a decrease in memory accuracy in the memory test in general, we considered the corrected hit rate (i.e., the discrimination index P_r, hit rate minus false alarm rate; Snodgrass & Corwin, 1988; ratings *yes* and *rather yes* were grouped as *acceptance*, and *no* and *rather no* as *rejection*). To better understand how memory for original and modified episodes was influenced in detail, we also analyzed hit rates and false alarm rates separately. Furthermore, we examined RTs in the memory test, which can serve as an indicator of how long it takes to retrieve information (correctly) from memory (Collins & Quillian, 1969). Longer RTs indicate increased difficulty of retrieval because of higher cognitive processing demands (Noppeney & Price, 2004; Larsen & Plunkett, 1987), which may also occur when competing versions of an episode are processed.

For the analysis of corrected hit rates as well as hit rates, false alarm rates (unmodified videos = targets, modified videos = distractors), and RTs for modified videos in the memory test ($\text{modified}_{\text{MT}}$), we applied a 2 × 2 within-subject factorial design with the factors $\text{Modification}_{\text{FMRI}}$ (yes, no) and $\text{Version}_{\text{MT}}$ (str, con). For analyzing RTs for original videos in the memory test ($\text{original}_{\text{MT}}$), we applied a within-subject design with the factor $\text{Version}_{\text{FMRI}}$ (ori, str, con). RTs were averaged over all trials of the same factorial combination. Several participants did not give any correct answers (i.e., rejection) in response to $\text{modified}_{\text{MT}}$ videos for one or more factorial combinations. For this reason, the number of datapoints included in this specific analysis was reduced to 23 per factorial combination.

We also conducted an explorative analysis on behavioral data from the fMRI session. We calculated the error rate and mean RT according to the within-subject factor $\text{Version}_{\text{FMRI}}$ (ori, str, con, nov) per participant. No response was given in only 0.2% of all question trials, and these trials were not further considered in the analysis.

For the choice of statistical tests, data were inspected for normal distribution using the Shapiro Wilk Test. Furthermore, data were checked for extreme outliers as defined as values above quartile 3 + 3 × interquartile range or lower than quartile 1–3 × interquartile range. When data were normally distributed or could be transformed to fit normal distribution (RTs; logarithmic transformation) and showed no extreme outliers, we used conventional repeated-measures ANOVA (rmANOVA). When the prerequisites for parametric analysis were not met, we used a nonparametric rmANOVA based on aligned rank-transformed data (package ARTool; Wobbrock, Findlater, Gergle, & Higgins, 2011; corrected hit rates, hit rates, false alarm rates, error rates). Post hoc pair-wise comparisons were conducted with paired *t* tests or Wilcoxon signed-ranks tests (one-tailed when comparing ori and str and ori and con, two-tailed when comparing str and con; always two-tailed for explorative analysis of fMRI task). In addition, we used one-sample Wilcoxon signed-ranks tests to test whether corrected hit rates were significantly larger than zero.

As descriptive statistics, we report mean values and standard errors of the mean. For all behavioral analyses, we

Table 1. Whole-Brain Activation for Shared Activation of Both Episodic Modifications at FDR $p < .05$ (Cluster Level)

			MNI Coordinates			
Localization	*H*	*Cluster Extent*	*x*	*y*	*z*	*t Value*
		(str > ori)∩(con > ori)				
pIPS extending into AG	R	147	33	−67	56	5.03
		mod > ori				
Superior parietal lobe/pIPS	L	152	−27	−61	50	5.59
mIPS	L	l.m.	−39	−43	38	4.03
pIPS	R	600	33	−67	35	6.21
mIPS	R	l.m.	45	−37	50	4.58
Precuneus	R	l.m.	6	−67	41	4.49
IFS	R	207	42	11	35	5.39
MFG	R	l.m.	36	14	53	3.89
OTC	R	119	54	−52	−10	5.87
Cerebellum	L	80	−6	−82	−37	5.11

H = Hemisphere; L = Left; R = Right; str = structure modification; con = content modification; ori = original; mod = modification (aggregated); l.m. = local maximum.

applied a significance level of $\alpha = .05$. p values were adjusted according to the Bonferroni correction for multiple comparisons (Bonferroni, 1936). If the assumption of sphericity was violated as assessed by Mauchly's test of sphericity, we report Greenhouse–Geisser-corrected degrees of freedom and p values.

RESULTS

fMRI Results

Behavioral Performance during fMRI Session

We calculated a nonparametric rmANOVA on error rates for the fMRI task with the factor $\text{VERSION}_{\text{FMRI}}$ (ori, str, con, nov). Descriptively, error rates were generally very low for all factor levels ($M_{ori} = .028 \pm .007$; $M_{str} = .036 \pm .008$; $M_{con} = .028 \pm .006$; $M_{nov} = .073 \pm .014$) and did not differ significantly, $F(3, 105) = 1.99$, $p = .12$, $\eta p^2 = .05$. There was a significant effect of $\text{VERSION}_{\text{FMRI}}$ on RTs, $F(3, 105) = 14.32$, $p < .001$, $\eta p^2 = .29$. Post hoc tests revealed that participants' took longer to respond after novel videos than after any other version (ori vs. nov: $t(35) = -5.95$, $p < .001$; str vs. nov: $t(35) = -3.79$, $p = .003$; con vs. nov: $t(35) = -4.22$, $p < .001$; $M_{ori} = 950.450$ msec ± 26.813 msec; $M_{str} = 978.695$ msec ± 30.695 msec; $M_{con} = 969.741$ msec ± 26.250 msec; $M_{nov} = 1034.822$ msec ± 35.786 msec).

Neural Responses to Modified Episodic Cueing

First, we tested whether structure and content modifications elicit common brain activation patterns compared with original episodes. To this end, we calculated the conjunction of the whole-brain contrasts str > ori and con > ori, which revealed a significant activation cluster in right posterior IPS (pIPS) extending into dorsal angular gyrus (AG; Table 1). As a more liberal approach to detect common activation, we contrasted modified episodes, aggregated over both modification types, with original episodes (mod > ori). Again, common activation was found in pIPS, extending into middle IPS (mIPS) in both hemispheres. Shared activation over both modification types was also detected in right precuneus, inferior frontal sulcus (IFS), middle frontal gyrus (MFG), OTC, and left cerebellum (Table 1, Figure 4).

To investigate which brain regions specifically respond to structure modifications in episodes, we inspected the contrast str > ori. Compared with episodes without modification, structurally modified episodes activated right SFS/PMd, MFG, IFS (Brodmann's area [BA] 44 and 45), supramarginal gyrus (SMG), posterior superior temporal sulcus (pSTS), IPS, and AG. In addition, we found activation in left and right precuneus (Table 2, Figure 5A).

To characterize brain responses to content modification during episodic cueing, we investigated the contrast con > ori. Compared with episodes without modification, content modifications bilaterally elicited higher activity in pIPS and OTC, including FG and parahippocampal gyrus (PHG). In the right hemisphere, there was a significant activation cluster in IFS, including BA 44 and 45. In addition, we found activation in the left cerebellum (Table 2, Figure 5B).

To further verify the specificity of brain responses to both modification types, we also investigated the direct

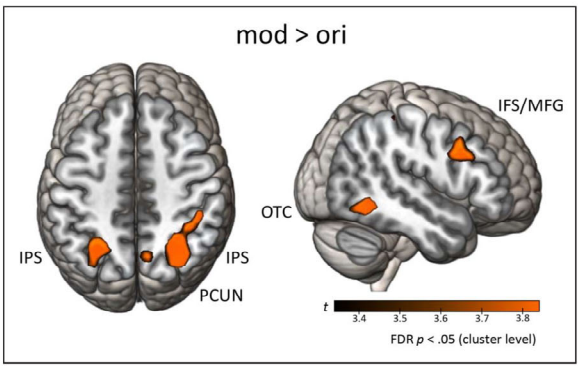


Figure 4. Whole-brain activation for episodic modifications, aggregated over both modification types. FDR-corrected ($p < .05$) *t*-map for mod > ori contrast. Ori = original; mod = modification (aggregated); PCUN = precuneus.

contrasts between them (str > con, con > str). In contrast to content modifications, structure modifications elicited higher activation in right precuneus, MFG, SFS/PMd, and lingual gyrus (LG). Bilaterally, we found significant activation in SMG and pSTS, extending into posterior middle temporal gyrus (pMTG) in the right hemisphere (Table 3, Figure 6A). Content compared with structure modifications triggered an elevated brain response in pIPS and OTC, including FG and PHG (Table 3, Figure 6B).

Neural Effects of Episodic Reactivation

In addition to our main research question, we investigated which brain regions were activated during episodic retrieval in general. To this end, we contrasted each type of episode (ori, str, con) with novel videos (nov). For all three episode types, we found significant activation in LG (only right for con > nov), cuneus and precuneus. Original episodes and those with structure modifications additionally activated posterior cingulate cortex (PCC) and ACC. For structurally modified episodes, ACC activation extended into medial frontal gyrus and we found another significant activation cluster in right AG (Table 4, Figure 7).

Table 2. Whole-Brain Activation for Different Episodic Modifications at FDR $p < .05$ (Cluster Level)

			MNI Coordinates			
Localization	*H*	*Cluster Extent*	*x*	*y*	*z*	*t Value*
		str > ori				
SFS/PMd	R	389	30	8	53	5.25
MFG, extending into IFS (BA 44/45)	R	l.m.	39	8	38	5.08
SMG	R	501	45	−40	47	6.15
Posterior superior temporal gyrus	R	l.m.	57	−49	23	5.47
IPS	R	l.m.	42	−46	41	5.35
AG	R	l.m.	39	−64	32	5.29
Precuneus	L	173	−3	−64	38	5.12
	R	l.m.	6	−64	62	4.02
		con > ori				
pIPS	L	361	−27	−61	50	7.22
	R	489	33	−67	35	7.08
IFS (BA 44)	R	127	42	8	32	5.18
BA 45	R	l.m.	42	23	26	3.84
OTC	L	282	−42	−58	−7	7.76
Fusiform gyrus, extending into PHG	L	l.m.	−33	−46	−16	6.13
OTC	R	299	51	−52	−10	9.07
Fusiform gyrus, extending into PHG	R	l.m.	30	−40	−19	5.70
Cerebellum	L	75	−6	−82	−34	5.71

H = Hemisphere; L = Left; R = Right; str = structure modification; con = content modification; ori = original; l.m. = local maximum.

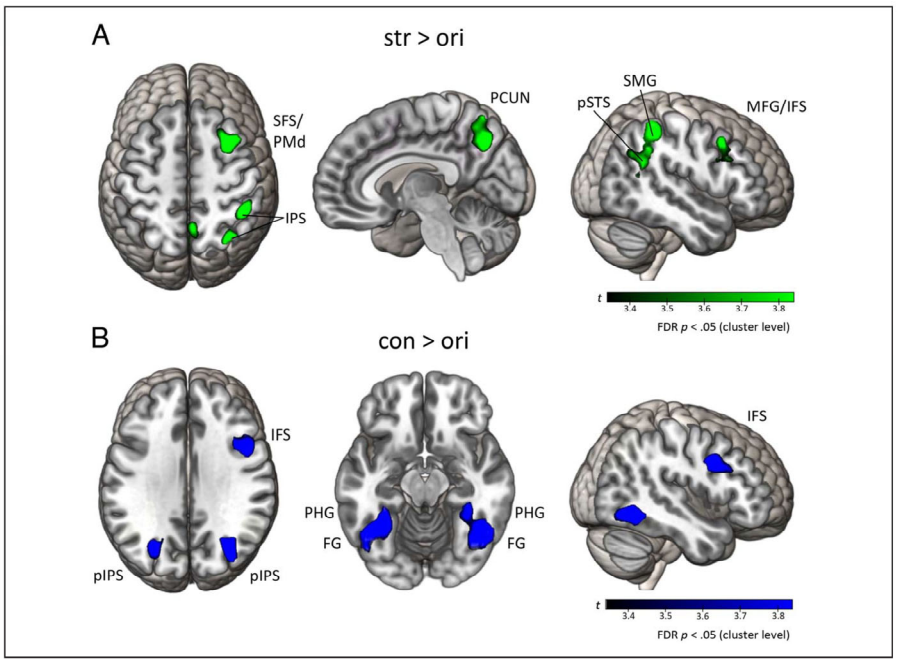


Figure 5. Whole-brain activation for different episodic modifications. (A) FDR-corrected ($p < .05$) t map for str > ori contrast. (B) FDR-corrected ($p < .05$) t-map for con > ori contrast. Ori = original; str = structure modification; con = content modification; (p)IPS = (posterior) intraparietal sulcus; PCUN = precuneus.

Table 3. Whole-Brain Activation for Direct Contrast of Different Episodic Modifications at FDR $p < .05$ (Cluster Level)

			MNI Coordinates			
Localization	*H*	*Cluster extent*	*x*	*y*	*z*	*t Value*
		str > con				
Precuneus	R	76	9	−49	62	5.45
SMG	R	602	51	−37	32	6.21
pSTS, extending into pMTG	R	l.m.	45	−34	2	4.80
MFG	R	320	42	35	32	5.65
SFS/PMd	R	l.m.	24	14	59	5.28
pSTS	L	295	−57	−64	17	5.01
SMG	L	l.m.	−57	−40	29	4.08
LG	R	71	12	−79	−4	5.09
		con > str				
pIPS	R	81	27	−55	47	5.00
	L	218	−24	−61	44	5.97
OTC	R	435	48	−61	−10	10.92
Fusiform gyrus	R	l.m.	30	−43	−19	7.14
PHG	R	l.m.	18	−31	−16	3.87
OTC	L	349	−42	−61	−7	8.26
Fusiform gyrus	L	l.m.	−33	−49	−16	5.87
PHG	L	l.m.	−33	−34	−16	5.09

H = Hemisphere; L = Left; R = Right; str = structure modification; con = content modification; ori = original; l.m. = local maximum.

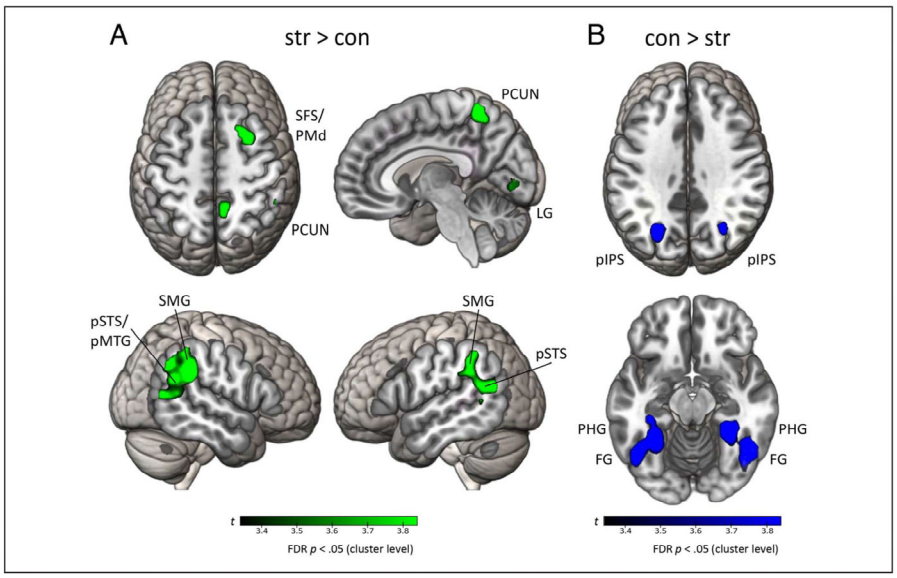


Figure 6. Whole-brain activation for direct contrasts between episodic modifications. (A) FDR-corrected ($p < .05$) *t*-map for str > con contrast. (B) FDR-corrected ($p < .05$) *t*-map for con > str contrast. Str = structure modification; con = content modification; PCUN = precuneus.

Post-fMRI Memory Test

Corrected Hit Rates

First, we investigated the general memory accuracy in the memory test, using the corrected hit rate. We confirmed that participants did not merely guess when rating videos as corrected hit rates for each factorial combination were significantly larger than zero (no-str: $z = -4.43, p < .001$; yes-str: $z = -4.17, p < .001$; no-con: $z = -5.44, p < .001$; yes-con: $z = -5.30, p < .001$). A nonparametric rmANOVA with the factors $\text{Modification}_{\text{FMRI}}$ (yes, no) and $\text{Version}_{\text{MT}}$ (str, con) revealed a significant main effect of $\text{Modification}_{\text{FMRI}}$, $F(1, 35) = 10.17$, $p = .003$, $\eta p^2 = .23$, which was driven by higher corrected hit rates for no ($M = .580 \pm .041$) compared with yes ($M = .460 \pm .035$), indicating a better memory performance when no modifications had been presented during the fMRI session. There was also a significant main effect of $\text{Version}_{\text{MT}}$, $F(1, 35) = 109.65$, $p < .001$, $\eta p^2 = .76$, which was explained by higher corrected hit rates values for con ($M = .757 \pm .038$) than for str ($M = .283 \pm .042$). There was no significant interaction of $\text{Modification}_{\text{FMRI}}$ and $\text{Version}_{\text{MT}}$, $F(1, 35) = 0.11$, $p = .75$, $\eta p^2 = .00$ (Figure 8A).

Hit Rates

Hit rates for $\text{original}_{\text{MT}}$ episodes were close to ceiling for all factorial combinations. A nonparametric rmANOVA with the factors $\text{Modification}_{\text{FMRI}}$ (yes, no) and $\text{Version}_{\text{MT}}$ (str, con) revealed a significant main effect of $\text{Modification}_{\text{FMRI}}$, $F(1, 35) = 12.43, p = .001$, $\eta p^2 = .26$, which was driven by higher hit rates for no ($M = .962 \pm .016$) than for yes ($M = .929 \pm .014$). Thus, participants were more prone to erroneously reject original episodes after the presentation of modified videos during the fMRI session. In addition, we found a significant main effect of $\text{Version}_{\text{MT}}$, $F(1, 35) = 5.36$, $p = .027$, $\eta p^2 = .13$, with higher hit rates for con ($M = .955 \pm .011$) than for str ($M = .936 \pm .015$). This indicates that participants were generally better at recognizing originally encoded episodes of which they also knew the content-modified version. The interaction of both factors was also significant, $F(1, 35) = 8.18$, $p = .007$, $\eta p^2 = .19$, and post hoc pairwise comparisons revealed that hit rates only decreased significantly after pre-experience with structure ($z = -2.70, p = .015$), but not content-modified episodes ($z = -0.47, p = 1$). Please note, however, that all differences in absolute values were quite small and thus should be interpreted with caution (Figure 8B).

False Alarm Rates

We computed a nonparametric rmANOVA with the factors $\text{Modification}_{\text{FMRI}}$ (yes, no) and $\text{Version}_{\text{MT}}$ (str, con) to analyze false alarm rates for $\text{modified}_{\text{MT}}$ episodes. There was a significant main effect of $\text{Modification}_{\text{FMRI}}$, $F(1, 35) = 4.93$, $p = .033$, $\eta p^2 = .12$, which was driven by higher false alarm rates for yes ($M = .469 \pm .039$) than for no ($M = .382 \pm .037$). Thus, participants were more prone to accept modified episode videos as originally encoded when a modified version had already been presented during the fMRI session. In addition, there was a significant main effect of $\text{Version}_{\text{MT}}$, $F(1, 35) = 113.51$, $p < .001$, $\eta p^2 = .76$, as false alarm rates were higher for str ($M = .653 \pm .046$) than for con ($M = .198 \pm .035$). This shows that participants generally accepted videos with modified structure much more readily than alternatives with modified content. We found a nonsignificant trend for an interaction of $\text{Modification}_{\text{FMRI}}$ and $\text{Version}_{\text{MT}}$, $F(1, 35) = 3.25$, $p = .080$, $\eta p^2 = .08$. Descriptively, false alarm rates for structure modified videos were increased less by the previous experience of

Table 4. Whole-Brain Activation for Episodic recall at FDR $p < .05$ (Cluster Level)

			MNI Coordinates			
Localization	*H*	*Cluster Extent*	*x*	*y*	*z*	*t Value*
		ori > nov				
PCC	R + L	168	0	−22	32	7.73
ACC	L	292	−6	26	23	5.21
	R	l.m.	6	26	20	5.18
LG	R	2508	6	−91	−4	13.11
Cuneus, extending into precuneus	R	l.m.	9	−88	38	11.13
	L	l.m.	−3	−85	14	9.25
		str > nov				
AG	R	121	48	−55	53	5.74
Medial frontal gyrus	R + L	242	0	32	35	5.63
ACC	L	l.m.	−3	32	26	4.99
	R	l.m.	9	38	11	4.42
PCC	R + L	144	3	−22	32	9.02
LG	R	2459	9	−91	−7	13.30
Cuneus, extending into precuneus	R	l.m.	18	−85	20	8.99
	L	l.m.	−3	−85	14	7.91
		con > nov				
LG	R	1382	9	−88	−7	11.66
Cuneus	R	l.m.	6	−88	35	5.93
	L	l.m.	0	−94	20	5.37
Precuneus	R	l.m.	15	−64	32	5.84
	L	l.m.	−9	−70	38	5.55

H = Hemisphere; L = Left; R = Right; str = structure modification; con = content modification; ori = original; nov = novel; l.m. = local maximum.

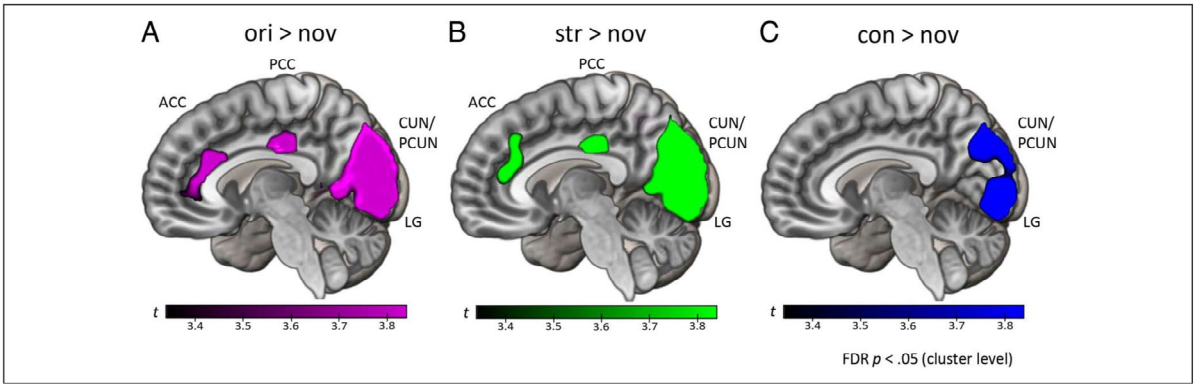


Figure 7. Whole-brain activation for videos showing original or slightly modified episodes, contrasted with novel videos. (A) FDR-corrected ($p < .05$) *t* map for ori > nov contrast. (B) FDR-corrected ($p < .05$) *t*-map for str > nov contrast. (C) FDR-corrected ($p < .05$) *t*-map for con > nov contrast. Ori = original; str = structure modification; con = content modification; nov = novel ; CUN = cuneus; PCUN = precuneus.

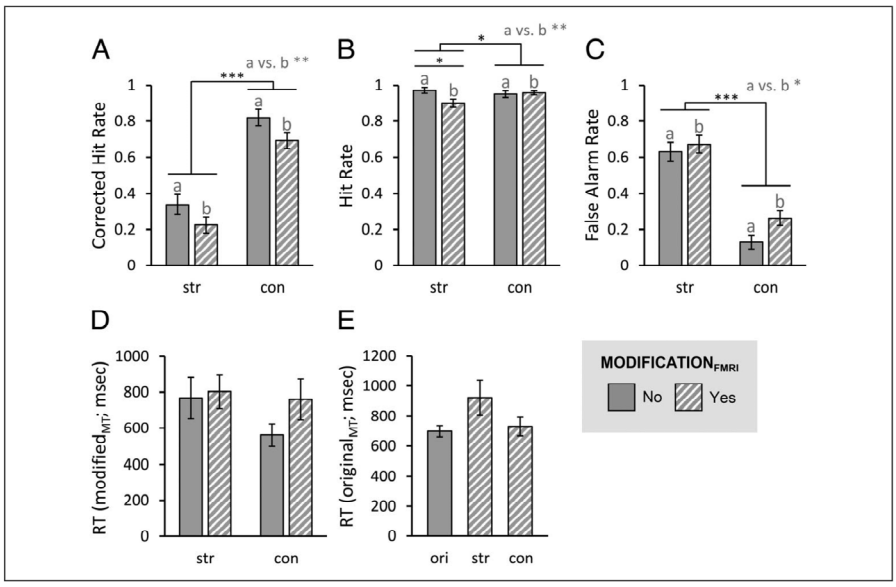


Figure 8. Behavioral results from post-fMRI memory test. For modified$_{MT}$ and original$_{MT}$ videos, participants rated whether they showed originally encoded episodes or not. Original$_{MT}$ videos were the targets whereas modified$_{MT}$ videos were distractors. (A) Corrected hit rate. Statistics: nonparametric rmANOVA with the factors MODIFICATION$_{FMRI}$ (yes, no) and VERSION$_{MT}$ (str, con), n = 36. (B) Hit rates for original$_{MT}$ videos. Statistics: nonparametric rmANOVA with the factors MODIFICATION$_{FMRI}$ (yes, no) and VERSION$_{MT}$ (str, con) and Wilcoxon signed-ranks tests, n = 36. (C) False alarm rates for modified$_{MT}$ videos. Statistics: nonparametric rmANOVA with the factors MODIFICATION$_{FMRI}$ (yes, no) and VERSION$_{MT}$ (str, con), n = 36. (D) RTs for modified$_{MT}$ videos. Statistics: rmANOVA with the factors MODIFICATION$_{FMRI}$ (yes, no) and VERSION$_{MT}$ (str, con), n = 23. (E) RTs for original$_{MT}$ videos. Statistics: rmANOVA with the factor VERSION$_{FMRT}$ (ori, str, con), n = 36. Bar plots show means and standard errors. *p < .05, **p < .01, ***p < .001. Ori = original; str = structure modification; con = content modification; a versus b indicates the main effect of MODIFICATION$_{FMRI}$.

episodic modifications than those for content-modified videos (Figure 8C). To control for a general acceptance bias, we compared false alarm rates for novel videos and modified$_{MT}$ videos using the Wilcoxon signed-ranks test (one-tailed). False alarm rates for novel videos were at a floor level (M = .007 ± .005) and significantly lower than those for modified$_{MT}$ videos (z = −5.35, p < .001; M = .425 ± .034; Figure 8C).

RTs

A rmANOVA on RTs for modified$_{MT}$ videos with the factors MODIFICATION$_{FMRI}$ (yes, no) and VERSION$_{MT}$ (str, con) revealed a near significant effect of MODIFICATION$_{FMRI}$, F(1, 22) = 3.86, p = .062, ηp^2 = .15, as participants tended to take longer to correctly reject modified$_{MT}$ episodes when the same had already been presented in the scanner (M = 782.520 msec ± 81.481 msec vs. M = 665.384 msec ± 79.401 msec). Descriptively, RTs were shorter for videos with content than with structure modification (M = 661.447 msec ± 76.010 msec vs. M = 786.458 msec ± 85.829 msec), but this difference did not reach significance either, F(1, 22) = 3.06, p = .094, ηp^2 = .12. We found no significant interaction effect, F(1, 22) = 0.58, p = .45, ηp^2 = .03 (Figure 8D). There was a near significant effect of VERSION$_{FMRI}$ (ori, str, con) on RTs for original$_{MT}$ videos, F(1.72, 60.09) = 3.21, p = .055, ηp^2 = .08. Descriptively, RTs were longest for str (M = 923.092 msec ± 117.161 msec) compared with ori (M = 697.608 msec ± 37.210 msec) and con (M = 729.762 msec ± 62.904 msec; Figure 8E).

DISCUSSION

In this study, we investigated brain and behavioral responses to violation of episodic expectancy induced by cues with modified details in structure or content. As hypothesized, brain responses differed for these two types of episodic prediction errors, reflecting the processing of divergent object and structure information. Modified episodes were mistaken for veridical originals more often in a post-fMRI memory test when already presented during the fMRI scan, whereas correct recognition of originally encoded episodes decreased. Together, findings provide evidence that different types of mnemonic prediction errors are processed differently by the brain and may contribute to memory changes.

Neural Responses to Episodes and Episodic Modifications

The presentation of original and slightly modified videos of encoded episodes recruited several brain regions associated with episodic retrieval compared with novel videos (Jeong, Chung, & Kim, 2015; Rugg & Vilberg, 2013; Wiggs, Weisberg, & Martin, 1999). The activation patterns for episodic retrieval closely resemble the one we detected in a previous study with a similar paradigm, comprising ACC, PCC, precuneus, cuneus, and LG (Jainta et al., 2022). Thus, in parallel to replicating our previous findings, we validated that participants had successfully encoded episodes and that the presentation of videos thereof cued episodic memories. Importantly, this was true for original and for modified episodes. This was to be expected because of

the subtle changes in modified episodes, which, overall, were still highly familiar to the participants.

New content and structure information of the episodic cue was expected to draw on distinct brain areas, but also to share some common activation in medial frontal cortex and the hippocampal formation. Although we could not confirm this hypothesis in this study, we found significant common activation in (right) pIPS, as revealed by the conjunction analysis and the aggregated modification contrast. Therefore, we suggest that superior parietal regions might be involved in processing of prediction errors in the context of episodic memory, potentially by guiding updating mechanisms. This interpretation fits the finding that dorsal parietal cortex plays an important role in the formation of episodic memories (Uncapher & Wagner, 2009). In addition, this area is well known to be involved in the reorientation of attention to salient and unexpected stimuli (Molenberghs, Mesulam, Peeters, & Vandenberghe, 2007; Corbetta & Shulman, 2002). It has been suggested that the superior parietal lobe, including pIPS, regulates top–down attention in memory. This is especially important when additional postretrieval processes are necessary to discriminate between what is true memorized content and what is not (Cabeza et al., 2011; Ciaramelli, Grady, & Moscovitch, 2008).

Furthermore, the more liberal approach of aggregating episodes over different types of modification revealed common activation in right IFS. This reflects our previous finding that activity in ventrolateral pFC increases for inconsistent or highly informative detail in observed actions (El-Sourani et al., 2019; Hrkać, Wurm, Kühn, & Schubotz, 2015; Wurm & Schubotz, 2012).

Structure and content modifications each recruited a set of brain regions unique to the modification type. Activation patterns in contrast to original episodes closely resembled those for the direct contrasts between structure and content modifications, indicating high specificity of brain responses for each modification type.

On the one hand, we had expected that structure modifications specifically lead to activation in brain regions involved in the temporal organization of episodes. We found that structure modifications co-activated right PMd/SFS, SMG, pSTS, and precuneus, suggesting this network contributes to the updating of predictive models because of unexpected new structure in episodes. This interpretation is consistent with previous reports about the functional characteristics of said areas. Accordingly, activity of a region comprising PMd/SFS is related to linking successive action steps (Pomp et al., 2021; Hrkać et al., 2014; Schubotz et al., 2012; Stadler et al., 2011) and could contribute to updating the current event or action model with respect to each next segment (Pomp et al., 2021; Schubotz et al., 2012; Tamber-Rosenau et al., 2011; Kurby & Zacks, 2008). SMG and precuneus have been demonstrated to be important for the sequential organization of memories (Foudil, Kwok, & Macaluso, 2020; Guidali, Pisoni, Bolognini, & Papagno, 2019) and involved in sequential learning (Burke, Bramley, Gonzalez, & McKeefry, 2013; Oishi et al., 2005). In line with this, we recently found that SMG is sensitive for perceived break points in actions (Pomp et al., 2021). Activity in (right) pSTS is characteristic for the processing of biological motion (Gilaie-Dotan, Kanai, Bahrami, Rees, & Saygin, 2013; Grossman, Battelli, & Pascual-Leone, 2005) and, in this context, action adaptation (Thurman, van Boxtel, Monti, Chiang, & Lu, 2016). pSTS activation for actions has been found to be goal-sensitive, responding more strongly when expected spatial transport targets are not met (Shultz, Lee, Pelphrey, & McCarthy, 2011).

On the other hand, we found content modifications to specifically recruit pIPS and OTC, including FG, which were hypothesized on the basis of their role for processing of object properties in the context of actions (El-Sourani et al., 2019; Lingnau & Downing, 2015; Wiggett & Downing, 2011; Reber et al., 2005; Grill-Spector, Kourtzi, & Kanwisher, 2001). More specifically, pIPS encodes basic visual features of graspable objects (Mruczek, von Loga, & Kastner, 2013; Creem-Regehr, 2009), reflecting the interaction with toy objects in our paradigm. In addition, content modifications elicited activity in the hippocampal formation (PHG), which likely represents ongoing learning because of the detected mismatch. In general, hippocampus and PHG are important in learning contexts (Davachi & Wagner, 2002; Köhler, Crane, & Milner, 2002; O'Reilly & Rudy, 2000; Aguirre, Detre, Alsop, & D'Esposito, 1996) and there is evidence that the PHG is involved in processing of competing memories (Kuhl, Bainbridge, & Chun, 2012). Moreover, the hippocampal formation is believed to generate mismatch signals when predictions do not fit perceptual inputs (Long et al., 2016; Duncan, Curtis, & Davachi, 2009; Kumaran & Maguire, 2007). Because our post-fMRI memory test data imply that content changes were more salient than structural changes, one could speculate that the discrepancy between what was predicted and what was perceived in case of content modifications was strong enough to be reflected in the activation of the hippocampal formation. Interestingly, the overall activation pattern we found for content modifications closely resembles the one Gläscher, Daw, Dayan, and O'Doherty (2010) identified for what they call *state prediction errors*, which is characterized by a mismatch of the expected and current state.

Taken together, we found that structure and content modifications activated distinct networks, each specifically representing the processing of the type of unexpected new information. We therefore achieved the main aim of this study, that is, characterize brain responses to different types of prediction errors.

Influence of Episodic Modifications on Post-fMRI Memory Performance

As expected, those original episodes that had been presented in a modified version during the fMRI were recalled

less reliably, as reflected by a lower corrected hit rate. More specifically, previously modified videos were later more often mistaken for original ones, as evidenced by significantly increased false alarm rates. Others have reported that mnemonic prediction errors lead to an intrusion of new information into an established memory repertoire (Sinclair & Barense, 2018; Long et al., 2016). Also, in our study, the recognition of original videos was impaired after encounters of alternative versions. Similarly, Kim et al. reported that prediction violations led to decreased recognition of original memory content (Kim et al., 2014). In particular, prior presentation of structurally modified videos led to a decrease in the hit rate in the post-fMRI memory test. This might be taken as a first hint that different types of episodic modifications could influence memory traces differently, which would fit the specific brain responses we detected for structure and content modifications. Interestingly, participants tended to take longer to correctly classify a video if they had already seen the same video in a modified version during fMRI. Longer response times in cued-memory paradigms are interpreted as indicative of increased difficulty of retrieval because of higher cognitive processing demands (Noppeney & Price, 2004; Larsen & Plunkett, 1987). Thus, it is likely that it became more difficult to differentiate between alternative competing versions of episodes when versions diverging from the original experience had been already encountered in the fMRI session. However, effects concerning RTs need to be interpreted with caution because they only approached significance. Taken together, the behavioral findings suggest that structure and content modifications during cueing of episodic retrieval influenced subsequent memory for these episodes. Our findings corroborate the observation that mnemonic prediction errors can trigger episodic memory modification (Sinclair & Barense, 2019).

What remains unclear is how exactly memory traces were influenced by our intervention. For example, it has been discussed that memory modification can result from an interference of old and new memory traces (Sinclair & Barense, 2018, 2019; Klingmüller, Caplan, & Sommer, 2017; Sederberg, Gershman, Polyn, & Norman, 2011) or from source confusion (Hekkanen & McEvoy, 2002), which could both explain our findings. Then again, participants remembered correctly that novel videos had not been part of the original episode repertoire although novels had been repeatedly presented during the fMRI experiment as well. This speaks against source confusion in its simplest form as an explanation of the results of our post-fMRI memory test.

Another interesting behavioral finding was that participants generally had a strong tendency to accept structurally modified versions as originals in the memory test. Although both types of modifications resulted from a single change in the story, it is likely that structural modifications were generally less salient than content modifications. This would be matched by the fact that after the fMRI session, nearly all participants (86%) reported noticing at least one object swap, whereas only half of them had noticed a change in the sequence of action steps (53%). Recently, it was reported that memory performance following prediction errors differed depending on whether changes were detected (and remembered) by participants or not (Wahlheim & Zacks, 2019). Whereas undetected changes led to reduced memory performance, detected changes had the opposite effect. Depending on contextual factors, prediction errors can even improve subsequent memory (Greve, Cooper, Kaula, Anderson, & Henson, 2017; Smith, Hasinski, & Sederberg, 2013). Although our behavioral results suggest that structural changes were rarely noticed, corresponding to a reduced memory performance, more frequent detection in the case of content modification did not lead to enhanced, but on the contrary, also to decreased memory performance. Thus, the final impact of a prediction violation on memory appears to be multifactorially determined.

Limitations and Implications for Future Research

One factor that may limit the generalizability of our findings is that, for practical reasons, only women participated in the study. However, because the processing of episodic memory in the brain seems to be broadly similar between women and men (Nyberg, Habib, & Herlitz, 2000), we are confident that our findings are applicable to a more general population. In the future, our paradigm could be adapted to circumvent such practical limitations, for example, by applying virtual reality techniques so that encoding could be detached from the true physical appearance of participants' hands.

Second, we used new content and structure information, which contrasted details of the encoded episodes to elicit prediction errors. Although these interventions can also be interpreted as contextual and associative novelty, respectively, the unexpected new input within the familiar context will give rise to mnemonic prediction errors according to the predictive coding framework (see the work of Reichardt et al., 2020, for a review). In addition, we could show that episodes with structure and content modifications, in contrast to completely novel videos, recruited different brain regions associated with episodic retrieval (Jeong et al., 2015; Rugg & Vilberg, 2013; Wiggs et al., 1999). Still, it would be interesting to find a way to keep the novelty constant and have participants make active predictions that are then either violated or not.

Our findings from the post-fMRI memory test revealed that structure modifications were likely less salient and harder to detect than content modifications. Thus, we cannot exclude that the neural differences we detected were confounded by differences in prediction error strength. However, brain responses to structure modifications were highly specific and located in hypothesized areas. Moreover, structure modifications elicited equally strong activation as content modifications. It is thus highly unlikely that those differences simply arose because of quantitative

difference between modification types. For these reasons, we are confident that our findings indeed represent differential neural processing because of different types of episodic information.

As a caveat, we wish to point out that, in normal life, episodic memories are not trained or repeated in the strict sense as they were in our paradigm. Although experimental procedures must be rigorous to be able to test hypotheses, the rigor takes away from the applicability of the research.

Last, our trial-wise analysis of later false alarms and correct rejections in the memory test did not reveal significant brain activation predicting false memories. Our aim for future investigations is to optimize our paradigm to further analyze how later true and false memories are encoded by the brain.

Conclusion

When recalling episodes, our memory can change, for instance because of prediction errors in the reactivation process. Our results suggest that structural and content prediction errors in episode retrieval differ in their neural processing. The tendency to misclassify modified episodes as originally experienced episodes increased after experiencing repeated structural and content prediction errors. Accordingly, different types of prediction errors can confuse episodic memory and possibly lead to the emergence of alternative versions of the same memory trace. Our results may provide a fruitful starting point for further research on the mutability of episodic memories.

Acknowledgments

The authors thank Monika Mertens, Lena Puder, Simon Wieczorek, Jamuna Halscheid, Leandra Feldhusen, and Anne Glombitza for their help during data collection. Furthermore, we thank Annika Garlichs, Helena Sydlik, and Yuyi Xu for their assistance during the creation of stimulus material and Christin Schwarzer for training new student assistants. Last, we thank Jennifer Pomp, Lena Schliephake, Falko Mecklenbrauck, and Nina Heins for advice regarding data analysis and the members of research unit FOR 2812 for valuable discussions.

Reprint requests should be sent to Sophie Siestrup, University of Münster, Fliednerstraße 21, 48149 Münster, or via e-mail: s.siestrup@uni-muenster.de.

Data Availability Statement

All data reported here is publicly available at https://osf.io/m7dcu/?view_only=575d6ed3fbf544ada3bcb0519c86f94b.

Author Contributions

Sophie Siestrup: Formal analysis; Investigation; Methodology; Visualization; Writing—Original draft; Writing—Review & editing. Benjamin Jainta: Investigation; Methodology; Writing—Review & editing. Nadiya El-Sourani: Methodology; Writing—Review & editing. Ima Trempler: Formal analysis; Methodology; Writing—Review & editing. Moritz F. Wurm: Writing—Review & editing. Oliver T. Wolf: Writing—Review & editing. Sen Cheng: Conceptualization; Writing—Review & editing. Ricarda I. Schubotz: Conceptualization; Funding acquisition; Methodology; Resources; Supervision; Writing—Original draft; Writing—Review & editing.

Funding Information

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) – project numbers 419037023, 419039274, 419037518. The funders had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Aguirre, G. K., Detre, J. A., Alsop, D. C., & D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cerebral Cortex*, *6*, 823–829. https://doi.org/10.1093/cercor/6.6.823, PubMed: 8922339

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology*, *192*, 101821. https://doi.org/10.1016/j.pneurobio.2020.101821, PubMed: 32446883

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology*, *4*, 907. https://doi.org/10.3389/fpsyg.2013.00907, PubMed: 24376428

Bein, O., Duncan, K., & Davachi, L. (2020). Mnemonic prediction errors bias hippocampal states. *Nature Communications*, *11*, 3451. https://doi.org/10.1038/s41467-020-17287-1, PubMed: 32651370

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, *8*, 3–62.

Burke, M. R., Bramley, P., Gonzalez, C. C., & McKeefry, D. J. (2013). The contribution of the right supra-marginal gyrus to sequence learning in eye movements. *Neuropsychologia*, *51*,

3048–3056. https://doi.org/10.1016/j.neuropsychologia.2013.10.007, PubMed: 24157539
Cabeza, R., Mazuz, Y. S., Stokes, J., Kragel, J. E., Woldorff, M. G., Ciaramelli, E., et al. (2011). Overlapping parietal activity in memory and perception: Evidence for the attention to memory model. *Journal of Cognitive Neuroscience*, *23*, 3209–3217. https://doi.org/10.1162/jocn_a_00065, PubMed: 21568633
Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top–down and bottom–up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46*, 1828–1851. https://doi.org/10.1016/j.neuropsychologia.2008.03.022, PubMed: 18471837
Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1
Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215. https://doi.org/10.1038/nrn755, PubMed: 11994752
Creem-Regehr, S. H. (2009). Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. *Neurobiology of Learning and Memory*, *91*, 166–171. https://doi.org/10.1016/j.nlm.2008.10.004, PubMed: 18996216
Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: Insights from relational and item-based learning. *Journal of Neurophysiology*, *88*, 982–990. https://doi.org/10.1152/jn.2002.88.2.982, PubMed: 12163547
Duncan, K., Curtis, C., & Davachi, L. (2009). Distinct memory signatures in the hippocampus: Intentional states distinguish match and mismatch enhancement signals. *Journal of Neuroscience*, *29*, 131–139. https://doi.org/10.1523/JNEUROSCI.2998-08.2009, PubMed: 19129391
El-Sourani, N., Trempler, I., Wurm, M. F., Fink, G. R., & Schubotz, R. I. (2019). Predictive impact of contextual objects during action observation: Evidence from functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *32*, 326–337. https://doi.org/10.1162/jocn_a_01480, PubMed: 31617822
Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, *76*, 1057–1070. https://doi.org/10.1016/j.neuron.2012.12.002, PubMed: 23259943
Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories—The role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011, PubMed: 25453746
Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of Prediction Error. *Neuroscience and Biobehavioral Reviews*, *68*, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004, PubMed: 27287939
Foudil, S. A., Kwok, S. C., & Macaluso, E. (2020). Context-dependent coding of temporal distance between cinematic events in the human precuneus. *Journal of Neuroscience*, *40*, 2129–2138. https://doi.org/10.1523/JNEUROSCI.2296-19.2020, PubMed: 31996453
Franz, E. A., Ford, S., & Werner, S. (2007). Brain and cognitive processes of imitation in bimanual situations: Making inferences about mirror neuron systems. *Brain Research*, *1145*, 138–149. https://doi.org/10.1016/j.brainres.2007.01.136, PubMed: 17349983
Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622, PubMed: 15937014
Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210. https://doi.org/10.1002/hbm.460020402
Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *364*, 1211–1221. https://doi.org/10.1098/rstb.2008.0300, PubMed: 19528002
Gilaie-Dotan, S., Kanai, R., Bahrami, B., Rees, G., & Saygin, A. P. (2013). Neuroanatomical correlates of biological motion detection. *Neuropsychologia*, *51*, 457–463. https://doi.org/10.1016/j.neuropsychologia.2012.11.027, PubMed: 23211992
Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*, 585–595. https://doi.org/10.1016/j.neuron.2010.04.016, PubMed: 20510862
Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001, PubMed: 28579691
Griffiths, D., Dickinson, A., & Clayton, N. (1999). Episodic memory: What can animals remember about their past? *Trends in Cognitive Sciences*, *3*, 74–80. https://doi.org/10.1016/S1364-6613(98)01272-8, PubMed: 10234230
Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*, 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6, PubMed: 11322983
Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, *45*, 2847–2853. https://doi.org/10.1016/j.visres.2005.05.027, PubMed: 16039692
Guidali, G., Pisoni, A., Bolognini, N., & Papagno, C. (2019). Keeping order in the brain: The supramarginal gyrus and serial order in short-term memory. *Cortex*, *119*, 89–99. https://doi.org/10.1016/j.cortex.2019.04.009, PubMed: 31091486
Hekkanen, S. T., & McEvoy, C. (2002). False memories and source-monitoring problems: Criterion differences. *Applied Cognitive Psychology*, *16*, 73–85. https://doi.org/10.1002/acp.753
Horner, A. J., & Doeller, C. F. (2017). Plasticity of hippocampal memories in humans. *Current Opinion in Neurobiology*, *43*, 102–109. https://doi.org/10.1016/j.conb.2017.02.004, PubMed: 28260633
Hrkać, M., Wurm, M. F., Kühn, A. B., & Schubotz, R. I. (2015). Objects mediate goal integration in ventrolateral prefrontal cortex during action observation. *PLoS One*, *10*, e0134316. https://doi.org/10.1371/journal.pone.0134316, PubMed: 26218102
Hrkać, M., Wurm, M. F., & Schubotz, R. I. (2014). Action observers implicitly expect actors to act goal-coherently, even if they do not: An fMRI study. *Human Brain Mapping*, *35*, 2178–2190. https://doi.org/10.1002/hbm.22319, PubMed: 23983202
Jainta, B., Siestrup, S., El-Sourani, N., Trempler, I., Wurm, M. F., Werning, M., et al. (2022). Seeing what i did (not): Cerebral and behavioral effects of agency and perspective on episodic memory re-activation. *Frontiers in Behavioral Neuroscience*, *15*, 793115. https://doi.org/10.3389/fnbeh.2021.793115, PubMed: 35069141
Jeong, W., Chung, C. K., & Kim, J. S. (2015). Episodic memory in aspects of large-scale brain networks. *Frontiers in Human Neuroscience*, *9*, 454. https://doi.org/10.3389/fnhum.2015.00454, PubMed: 26321939

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, 8997–9002. https://doi.org/10.1073/pnas.1319438111, PubMed: 24889631

Klingmüller, A., Caplan, J. B., & Sommer, T. (2017). Intrusions in episodic memory: Reconsolidation or interference? *Learning and Memory*, *24*, 216–224. https://doi.org/10.1101/lm.045047.117, PubMed: 28416633

Köhler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus*, *12*, 718–723. https://doi.org/10.1002/hipo.10077, PubMed: 12542224

Kuhl, B. A., Bainbridge, W. A., & Chun, M. M. (2012). Neural reactivation reveals mechanisms for updating memory. *Journal of Neuroscience*, *32*, 3453–3461. https://doi.org/10.1523/JNEUROSCI.5846-11.2012, PubMed: 22399768

Kumaran, D., & Maguire, E. A. (2007). Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus*, *17*, 735–748. https://doi.org/10.1002/hipo.20326, PubMed: 17598148

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*, 72–79. https://doi.org/10.1016/j.tics.2007.11.004, PubMed: 18178125

Larsen, S. F., & Plunkett, K. (1987). Remembering experienced and reported events. *Applied Cognitive Psychology*, *1*, 15–26. https://doi.org/10.1002/acp.2350010104

Lee, J. L. C., Nader, K., & Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in Cognitive Sciences*, *21*, 531–545. https://doi.org/10.1016/j.tics.2017.04.006, PubMed: 28495311

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, *19*, 268–277. https://doi.org/10.1016/j.tics.2015.03.006, PubMed: 25843544

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *Journal of Neuroscience*, *36*, 12677–12687. https://doi.org/10.1523/JNEUROSCI.1850-16.2016, PubMed: 27821577

Maguire, E. A., Intraub, H., & Mullally, S. L. (2016). Scenes, spaces, and memory traces: What does the hippocampus do? *Neuroscientist*, *22*, 432–439. https://doi.org/10.1177/1073858415600389, PubMed: 26276163

Molenberghs, P., Mesulam, M. M., Peeters, R., & Vandenberghe, R. R. C. (2007). Remapping attentional priorities: Differential contribution of superior parietal lobule and intraparietal sulcus. *Cerebral Cortex*, *17*, 2703–2712. https://doi.org/10.1093/cercor/bhl179, PubMed: 17264251

Mruczek, R. E. B., von Loga, I. S., & Kastner, S. (2013). The representation of tool and non-tool object information in the human intraparietal sulcus. *Journal of Neurophysiology*, *109*, 2883–2896. https://doi.org/10.1152/jn.00658.2012, PubMed: 23536716

Nader, K. (2015). Reconsolidation and the dynamic nature of memory. *Cold Spring Harbor Perspectives in Biology*, *7*, a021782. https://doi.org/10.1101/cshperspect.a021782, PubMed: 26354895

Nader, K., & Einarsson, E. Ö. (2010). Memory reconsolidation: An update. *Annals of the New York Academy of Sciences*, *1191*, 27–41. https://doi.org/10.1111/j.1749-6632.2010.05443.x, PubMed: 20392274

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*, 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005, PubMed: 15808966

Noppeney, U., & Price, C. J. (2004). Retrieval of abstract semantics. *Neuroimage*, *22*, 164–170. https://doi.org/10.1016/j.neuroimage.2003.12.010, PubMed: 15110006

Nyberg, L., Habib, R., & Herlitz, A. (2000). Brain activation during episodic memory retrieval: Sex differences. *Acta Psychologica*, *105*, 181–194. https://doi.org/10.1016/s0001-6918(00)00060-3, PubMed: 11194411

Oishi, K., Toma, K., Bagarinao, E. T., Matsuo, K., Nakai, T., Chihara, K., et al. (2005). Activation of the precuneus is related to reduced reaction time in serial reaction time tasks. *Neuroscience Research*, *52*, 37–45. https://doi.org/10.1016/j.neures.2005.01.008, PubMed: 15811551

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4, PubMed: 5146491

O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, *10*, 389–397. https://doi.org/10.1002/1098-1063(2000)10:4<389::AID-HIPO5>3.0.CO;2-P

Peters, G. J., David, C. N., Marcus, M. D., & Smith, D. M. (2013). The medial prefrontal cortex is critical for memory retrieval and resolving interference. *Learning and Memory*, *20*, 201–209. https://doi.org/10.1101/lm.029249.112, PubMed: 23512936

Pomp, J., Heins, N., Trempler, I., Kulvicius, T., Tamosiunaite, M., Mecklenbrauck, F., et al. (2021). Touching events predict human action segmentation in brain and behavior. *Neuroimage*, *243*, 118534. https://doi.org/10.1016/j.neuroimage.2021.118534, PubMed: 34469813

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2005). Priming effects in the fusiform gyrus: Changes in neural activity beyond the second presentation. *Cerebral Cortex*, *15*, 787–795. https://doi.org/10.1093/cercor/bhh179, PubMed: 15371295

Reichardt, R., Polner, B., & Simor, P. (2020). Novelty manipulations, memory performance, and predictive coding: The role of unexpectedness. *Frontiers in Human Neuroscience*, *14*, 152. https://doi.org/10.3389/fnhum.2020.00152, PubMed: 32410975

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27. https://doi.org/10.1016/j.tics.2010.09.003, PubMed: 20951630

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*, 1432–1463. https://doi.org/10.1037/a0037559, PubMed: 25150680

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*, 255–260. https://doi.org/10.1016/j.conb.2012.11.005, PubMed: 23206590

Schiffer, A. M., Ahlheim, C., Ulrichs, K., & Schubotz, R. I. (2013). Neural changes when actions change: Adaptation of strong and weak expectations. *Human Brain Mapping*, *34*, 1713–1727. https://doi.org/10.1002/hbm.22023, PubMed: 22422724

Schiffer, A. M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS One*, *7*, e36445. https://doi.org/10.1371/journal.pone.0036445, PubMed: 22570715

Schubotz, R. I. (2004). *Human premotor cortex: Beyond motor performance*. MPI Series in Human Cognitive and Brain Sciences (Vol. 50). Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences.

Schubotz, R. I. (2015). Prediction and expectation. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference* (Vol. 3, pp. 295–302). Academic Press, Elsevier. https://doi.org/10.1016/B978-0-12-397025-1.00205-0

Schubotz, R. I., Korb, F. M., Schiffer, A. M., Stadler, W., & von Cramon, D. Y. (2012). The fraction of an action is more than a movement: Neural signatures of event segmentation in fMRI. *Neuroimage*, *61*, 1195–1205. https://doi.org/10.1016/j.neuroimage.2012.04.008, PubMed: 22521252

Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology*, *5*, 636. https://doi.org/10.3389/fpsyg.2014.00636, PubMed: 25009519

Scully, I. D., Napper, L. E., & Hupbach, A. (2017). Does reactivation trigger episodic memory change? A meta-analysis. *Neurobiology of Learning and Memory*, *142*, 99–107. https://doi.org/10.1016/j.nlm.2016.12.012, PubMed: 28025069

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin and Review*, *18*, 455–468. https://doi.org/10.3758/s13423-011-0086-9, PubMed: 21512839

Shultz, S., Lee, S. M., Pelphrey, K., & McCarthy, G. (2011). The posterior superior temporal sulcus is sensitive to the outcome of human and non-human goal-directed actions. *Social Cognitive and Affective Neuroscience*, *6*, 602–611. https://doi.org/10.1093/scan/nsq087, PubMed: 21097958

Siestrup, S., Jainta, B., Trempler, I., Cheng, S., & Schubotz, R. (in preparation). *Solidity meets surprise: How memory consolidation affects cerebral and behavioral processing of episodic prediction errors*. Department of Psychology, University of Münster.

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning and Memory*, *25*, 369–381. https://doi.org/10.1101/lm.046912.117, PubMed: 30012882

Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neurosciences*, *42*, 727–739. https://doi.org/10.1016/j.tins.2019.08.007, PubMed: 31506189

Smith, T. A., Hasinski, A. E., & Sederberg, P. B. (2013). The context repetition effect: Predicted events are remembered better, even when they don't happen. *Journal of Experimental Psychology: General*, *142*, 1298–1308. https://doi.org/10.1037/a0034067, PubMed: 23957285

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50. https://doi.org/10.1037/0096-3445.117.1.34, PubMed: 2966230

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*, 1643–1653. https://doi.org/10.1038/nn.4650, PubMed: 28967910

Stadler, W., Schubotz, R. I., von Cramon, D. Y., Springer, A., Graf, M., & Prinz, W. (2011). Predicting and memorizing observed action: Differential premotor cortex involvement. *Human Brain Mapping*, *32*, 677–687. https://doi.org/10.1002/hbm.20949, PubMed: 20225220

Tamber-Rosenau, B. J., Esterman, M., Chiu, Y. C., & Yantis, S. (2011). Cortical mechanisms of cognitive control for shifting attention in vision and working memory. *Journal of Cognitive Neuroscience*, *23*, 2905–2919. https://doi.org/10.1162/jocn.2011.21608, PubMed: 21291314

Thurman, S. M., van Boxtel, J. J. A., Monti, M. M., Chiang, J. N., & Lu, H. (2016). Neural adaptation in pSTS correlates with perceptual aftereffects to biological motion and with autistic traits. *Neuroimage*, *136*, 149–161. https://doi.org/10.1016/j.neuroimage.2016.05.015, PubMed: 27164327

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114, PubMed: 11752477

Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: Insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiology of Learning and Memory*, *91*, 139–154. https://doi.org/10.1016/j.nlm.2008.10.011, PubMed: 19028591

Wahlheim, C. N., & Zacks, J. M. (2019). Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General*, *148*, 30–50. https://doi.org/10.1037/xge0000458, PubMed: 29985021

Wiggett, A. J., & Downing, P. E. (2011). Representation of action in occipito-temporal cortex. *Journal of Cognitive Neuroscience*, *23*, 1765–1780. https://doi.org/10.1162/jocn.2010.21552, PubMed: 20807060

Wiggs, C. L., Weisberg, J., & Martin, A. (1999). Neural correlates of semantic and episodic memory retrieval. *Neuropsychologia*, *37*, 103–118. https://doi.org/10.1016/s0028-3932(98)00044-x, PubMed: 9920476

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *CHI '11: CHI Conference on Human Factors in Computing Systems* (pp. 143–146). https://doi.org/10.1145/1978942.1978963

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—Again. *Neuroimage*, *2*, 173–181. https://doi.org/10.1006/nimg.1995.1023, PubMed: 9343600

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: Contextual information modulates action recognition. *Neuroimage*, *59*, 1551–1559. https://doi.org/10.1016/j.neuroimage.2011.08.038, PubMed: 21878395

3.2 Study 2: Solidity Meets Surprise: Cerebral and Behavioral Effects of Learning from Episodic Prediction Errors

Running title: Learning from Episodic Prediction Errors

Sophie Siestrup, Benjamin Jainta, Sen Cheng, & Ricarda I. Schubotz (2023)

Journal of Cognitive Neuroscience, *35*, 291–313

Solidity Meets Surprise: Cerebral and Behavioral Effects of Learning from Episodic Prediction Errors

Sophie Siestrup[1,2*], Benjamin Jainta[1*], Sen Cheng[3], and Ricarda I. Schubotz[1,2]

Abstract

■ How susceptible a memory is to later modification might depend on how stable the episode has been encoded. This stability was proposed to increase when retrieving information more (vs. less) often and in a spaced (vs. massed) practice. Using fMRI, we examined the effects of these different pre-fMRI retrieval protocols on the subsequent propensity to learn from episodic prediction errors. After encoding a set of different action stories, participants came back for two pre-fMRI retrieval sessions in which they encountered original episodes either two or eight times in either a spaced or a massed retrieval protocol. One week later, we cued episodic retrieval during the fMRI session by using original or modified videos of encoded action stories. Recurrent experience of modified episodes was associated with increasing activity in the episodic memory network including hippocampal and cortical areas, when leading to false memories in a post-fMRI memory test. While this observation clearly demonstrated learning from episodic prediction errors, we found no evidence for a modulatory effect of the different retrieval protocols. As expected, the benefit of retrieving an episode more often was reflected in better memory for originally encoded episodes. In addition, frontal activity increased for episodic prediction errors when episodes had been less frequently retrieved pre-fMRI. A history of spaced versus massed retrieval was associated with increased activation throughout the episodic memory network, with no significant effect on behavioral performance. Our findings show that episodic prediction errors led to false memories. The history of different retrieval protocols was reflected in memory performance and brain responses to episodic prediction errors, but did not interact with the brain's episodic learning response. ■

INTRODUCTION

Through episodic memories, we can mentally relive events from our personal past (Tulving, 2002), and yet, episodic memories are not always veridical reconstructions of our experiences (Scully & Hupbach, 2020; Lee, Nader, & Schiller, 2017; Nader, 2015; Nader & Einarsson, 2010). In everyday life, there is always a certain discrepancy between our expectations, which we derive from memories, and new experiences. According to the predictive coding framework, this discrepancy gives rise to prediction errors, which serve as bottom–up learning signals to the brain (Reichardt, Polner, & Simor, 2020; Barto, Mirolli, & Baldassarre, 2013). It has been suggested that memory modification is fueled by mnemonic prediction errors (Sinclair & Barense, 2018; Fernández, Boccia, & Pedreira, 2016), that is, subtle discrepancies between the remembered situation and the current situation that are encountered during retrieval. Memory modification could be functional in that it allows us to maintain valid predictions in a highly dynamic environment. Thus, new experiences can be integrated into existing memories or *internal models* (Barron, Auksztulewicz, & Friston, 2020) to successfully guide our behavior in the long run (Fernández et al., 2016; Exton-McGuinness, Lee, & Reichelt, 2015).

In the current fMRI study, we examined the influence of different episodic memory retrieval protocols on learning from mnemonic prediction errors. To this end, we used an episode-modification paradigm adapted from our previous study (Jainta et al., 2022). First, participants went through a training during which they encoded episodes. Then, they completed two active pre-fMRI retrieval sessions. Half of the episodes were retrieved in a spaced schedule, that is, during both sessions, whereas the other half was retrieved in a massed schedule, that is, only in the second session. Overall, half of the episodes were retrieved only two times, whereas the other half were retrieved eight times. During the subsequent fMRI session, participants were presented videos reminiscent of the encoded episodes. Whereas some episode videos were presented in their original form, others were manipulated to induce prediction errors. Finally, participants completed a post-fMRI memory test that queried their memory of episodes originally experienced.

A fundamental factor mediating memory solidity is consolidation, a process that stabilizes a memory trace after its initial encoding. It occurs not only during sleep because of off-line replay (Poe, Walsh, & Bjorness, 2010), but also

[1]University of Münster, Germany, [2]Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany, [3]Ruhr University Bochum, Germany
*These authors contributed equally and share first authorship.

Journal of Cognitive Neuroscience 35:2, pp. 291–313
https://doi.org/10.1162/jocn_a_01948

when a memory is actively retrieved (Antony, Ferreira, Norman, & Wimber, 2017). Retrieval practice does not only improve memory performance, establishing the so-called "testing effect" (Rowland, 2014), but can also protect memories from later modification (Elsey, Van Ast, & Kindt, 2018; Antony et al., 2017). On the one hand, memory performance improves when an item is practiced more often (Lyle, Bego, Hopkins, Hieb, & Ralston, 2020; Karpicke & Roediger, 2007; Underwood, 1970). On the other hand, spreading the information to be practiced over several repetitions, so-called spaced practice, is more efficient for long-term retention than practicing all items on a single occasion, called massed practice (Lyle et al., 2020; YeckehZaare, Resnick, & Ericson, 2019; Kang, 2016; Gerbier & Toppino, 2015; Dempster, 1989; Underwood, 1970). The positive effect of spaced practice has been explained by more effortful preretrieval processing (Feng et al., 2019; Dobson, Perez, & Linderholm, 2017) that determines later retrieval success (Rowland, 2014). Against this backdrop, we expected that spaced (vs. massed) retrieval of episodic memories as well as retrieving them more (vs. less) frequently would lead to particularly stable memory traces, which should render episodic memories less susceptible to later change.

First, we tested the behavioral hypothesis that more consolidated episodes would be less susceptible to change during retrieval. As previously reported (Jainta et al., 2022; Siestrup et al., 2022), we found that repeated prediction violation leads to increased acceptance of modified episode videos as originally encoded, and decreased acceptance for unmodified videos, corroborating the idea that mnemonic prediction errors can contribute to memory modification. The behavioral data suggested that this modification consisted of incorporating additional variants of the episode rather than overwriting the original episode. Building on this observation, we now tested whether this learning process depends on the solidity of memory, which we manipulated using different retrieval protocols. In particular, we expected that modified videos of more solid episodes, that is, those that were retrieved more frequently and/or following a spaced protocol, would not be as readily accepted as original as modified videos of less solid ones (Schiffer, Ahlheim, Ulrichs, & Schubotz, 2013).

Second, using fMRI, we aimed to elucidate the neural processes underlying learning from episodic prediction errors. Previous studies found elevated activation in the hippocampal formation and medial frontal cortex (FMC) during the recall of supposedly more consolidated memories (Bosshardt et al., 2005), either because of more frequent prior retrieval (Wiklund-Hörnqvist, Stillesjö, Andersson, Jonsson, & Nyberg, 2021; Schiffer et al., 2013) or a spaced retrieval schedule (Li & Yang, 2020; Ezzyat, Inhoff, & Davachi, 2018; Zhan, Guo, Chen, & Yang, 2018; Takashima et al., 2009). On the basis of own previous studies, we expected that the moment of modification in an episodic cue triggers increases activity in areas that process mnemonic prediction errors (Schiffer et al., 2013; Schiffer, Ahlheim, Wurm, & Schubotz, 2012). We especially addressed the role of two brain areas, the hippocampal formation and the FMC, in learning from prediction errors during retrieval of differently consolidated episodes, as previously suggested (Bein, Duncan, & Davachi, 2020; van Kesteren, Ruiter, Fernández, & Henson, 2012). The hippocampus is known to be relevant for both associative learning (Suzuki, 2007) and initial consolidation through its link to neocortical areas (Squire, Genzel, Wixted, & Morris, 2015). Presumably because of this dual function, the hippocampus mediates the comparison of new information with stored memories, which may lead to an updating of the internal model (Long, Lee, & Kuhl, 2016; Duncan, Curtis, & Davachi, 2009; Kumaran & Maguire, 2007). The FMC is thought to play a more general role in the retrieval of consolidated memories (Preston & Eichenbaum, 2013; Sterpenich et al., 2009; Takashima et al., 2009). Against this background, we expected that the hippocampal complex (HC) and FMC respond more strongly to mnemonic prediction errors when more solid memories are involved, as divergent information triggers stronger mismatch signals for more consolidated memory traces (Schiffer et al., 2012, 2013).

However, it is also possible that less stable memories might be more difficult to retrieve and thus lead to higher activation in areas that are typically involved in episodic retrieval (e.g., Nadel, Campbell, & Ryan, 2007), including the hippocampus and FMC. Accordingly, it could be that neural prediction error signals are stronger for weaker memories, resulting in higher learning rates. To account for both options, we examined all effects for both supposedly strongly versus weakly consolidated memories and vice versa. Together with the behavioral findings, we would then be able to make assumptions about how learning from episodic prediction errors is accomplished by the brain.

As an important final step, we aimed at combining our functional and behavioral measures to investigate which neuronal processes give rise to learning from prediction errors, and how they might interact with memory solidity. In principle, learning from prediction errors can include two different phenomena (Gershman, Monfils, Norman, & Niv, 2017). In that sense, one possibility is that prediction errors could lead to a modification of old memory traces. In the memory test, this would manifest through more false negatives (misses). Alternatively, prediction errors might drive the acquisition of alternative variants of earlier memories, which would result in more false positives (false alarms). Because we observed high hit rates (i.e., low miss rates) combined with high false alarms rates in our previous study (Jainta et al., 2022), we sought here to test which brain areas reflect a gradual establishment of false memories. There is evidence suggesting increased hippocampal activity during successful encoding (Davachi, Mitchell, & Wagner, 2003) and retrieval of episodic information while activation decreases with

familiarity (Yonelinas, Otten, Shaw, & Rugg, 2005). In addition, stronger hippocampus activity during encoding is associated with better episodic memory (Davachi, 2006). Previous studies have shown that episodic memory encoding is characterized by increasing neocortical activity and decreasing hippocampal activity with number of repetitions (Brodt et al., 2016, 2018) and also with ongoing consolidation (Takashima et al., 2006). However, no studies have yet examined the cerebral reflection of incremental learning because of repeated episodic prediction errors. Hence, we were specifically interested in the dynamic increase of brain responses with accumulating evidence for new episode "alternatives." We expected that brain activity increases in areas related to memory formation, including hippocampal and parahippocampal regions (Ritchey, Libby, & Ranganath, 2015).

METHODS

This article is based on experimental data that were previously published in a companion paper (Siestrup et al., 2022). Please note that the factors addressed in this article were statistically independent of those reported in the companion paper.

Participants

Forty-five participants took part in the study. Like in our previous study (Jainta et al., 2022), participants were all female to achieve a good match between the hands in the videos and the hands of the participants. This was important for the credibility of our cover story that participants would be presented videos of themselves during the fMRI session. Four participants started the experiment but did not finish, either because of technical problems during the second retrieval session (three participants) or personal reasons (one participant). Data from five additional participants were excluded from analyses because of the incorrect presentation of video stimuli during the fMRI session (one participant) and increased movement during the fMRI session (four participants, approx. 5-mm movement). Consequently, 36 participants were part of the final sample (M = 22 years, SD = 2.78 years, range = 18–30 years). This sample size yielded stable results in our previous work, where we used not only the same number of participants but also an equivalent experimental and statistical design with the same number of conditions and trials (e.g., Jainta et al., 2022).

Participants had (corrected-to-) normal vision, were native German speakers, were and right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). They reported no history of neurological or psychiatric disorders or substance abuse. Participants received course credits or money for their participation and gave written informed consent to participate in this study. The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the University of Münster.

Stimuli

We used the same set of videos as previously reported (Jainta et al., 2022; available upon request at https://www.uni-muenster.de/IVV5PSY/AvicomSrv/). These were 78 short films (duration = 8.80–17.88 sec, M = 12.71 sec; six to nine action steps, M = 7.4 steps) showing stories played with PLAYMOBIL toys from a first-person perspective. Videos depicted toys from above being manipulated in front of a matte white background and the hands and underarms of an actress wearing a black pullover and black gloves. The back of the right hand was additionally marked with a yellow dot to facilitate imitation from demo videos (Franz, Ford, & Werner, 2007). Videos were filmed with a digital single-lens reflex camera (Nikon D5300). The section captured by the camera (47.5 cm × 28 cm; in the following referred to as camera frame) was marked on the background with tape. For each video, objects that were needed for the story were placed next to the camera frame and moved into view when they first appeared in the story. For editing video material, we used Adobe Premiere Pro CC (Adobe Systems Software, Version 12.1.2). Videos were cut so that they started with seven frames of white background and ended after seven frames of the final toy constellation. The frame size of the videos was 1920 × 1080 pixels, and the frame rate was 25 frames per second. Videos were presented at a visual angle of approximately 7.3° × 13° with the stimulus presentation software Presentation (Version 20.3 02.25.19, Neurobehavioral Systems) throughout the study.

A subset of 24 stories existed in three different versions. First, there was an original version (*ori*), which was used for encoding and retrieval. Second, in a modified version of the story, two adjacent action steps were switched (structure modification) to elicit prediction errors based on episode structure (*str*). Third, one object was exchanged as compared with the original version of the story (content modification) to elicit prediction errors based on episode content (*con*). Modifications did not occur during the first two or last two action steps. Effects regarding the factor modification (*str*, *con*) were addressed in a companion paper (Siestrup et al., 2022). In this article, we aggregated the modified videos (*str*, *con*) and will refer to them as modified versions (*mod*) in the following sections. Four additional videos were first introduced in the fMRI session and will be referred to as novels in the following. Two more videos were used for practice trials for the different tasks and were not shown in the fMRI experiment. These six videos existed in only one version each.

Procedure

For an overview of the procedure, please see Figure 1.

Encoding

For encoding of episodes, participants went through two initial training sessions in a computer lab at the

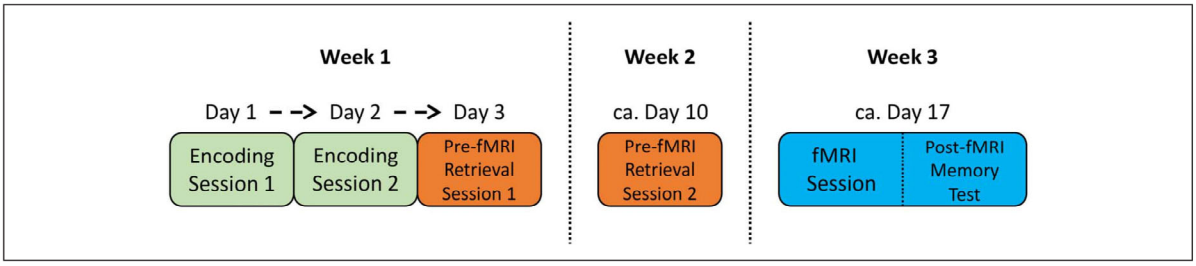


Figure 1. Schematic overview of the experimental procedure. During the first week, participants underwent two encoding sessions and one pre-fMRI retrieval session on three consecutive days. The first retrieval session only contained videos belonging to the spaced condition. Approximately 1 week later, participants returned to the behavioral laboratory for the second pre-fMRI retrieval session, during which they re-encountered videos belonging to the spaced as well as the massed condition. In Week 3, participants came back for the fMRI session, which was immediately followed by a post-fMRI memory test about which participants were not informed beforehand.

Department of Psychology at the University of Münster. The two sessions were conducted on two consecutive days and lasted about 2.25 and 1.75 hr, respectively, and during each session, participants encoded half of the 24 episodes.

Before the first encoding session started, participants were informed that their hands and forearms will be filmed to use these videos for the fMRI session (cover story). In contrast to our previous study, participants now imitated all of the 24 stories from demo videos (i.e., the original versions). For that, each video was presented three times in the first-person perspective (1 pp) and, afterward, had to be imitated correctly three additional times by the participant. For each participant, the order of videos was randomized. All stories were equally often encoded during the first and second sessions over the course of the study. Videos trained during Sessions 1 and 2 were balanced for the number of action steps.

During encoding sessions, participants wore the same black pullover and gloves as the actress when filming the demo videos and sat at the same filming setup. This way, the hands and arms of participants closely resembled those of the original actress in the videos, which was important for our cover story that participants would be presented videos of themselves during the fMRI session. The experimenter was present to monitor the participants' performance via a monitor that provided a live view of the participants' actions (Figure 2). Before a new video was shown, all toys included in the story were placed next to the camera frame in the same way as during the creation of the stimulus material. After three correct imitations of the story as judged by the experimenter, participants had to describe the story in detail to ensure that they had paid attention to all objects and had understood the story correctly. The experimenter immediately interrupted the participants as soon as they made a mistake during an imitation or description attempt to avoid encoding of wrong stories. In both cases, the experimenter then corrected the participant who had to start with a new imitation/description attempt. The experimenter monitored all details of the story as closely as possible, including the sequence of action steps and the correct positioning of hands and objects. Participants only performed one incorrect imitation attempt, on average (M = 1.037, SD = 0.469), so that they needed, on average, four attempts to complete the three correct imitations.

Pre-fMRI Retrieval Sessions

To further consolidate episodic memories, the second phase of the study was active retrieval of the before encoded stories. To this end, participants went through two pre-fMRI retrieval sessions. The first retrieval session took place 1 day after the second encoding session, the second retrieval session was conducted approximately 1 week later (range = 4–8 days; M = 6.36 days, SD = 0.93 days). Both sessions were conducted in a computer laboratory at the Department of Psychology at the University of Münster.

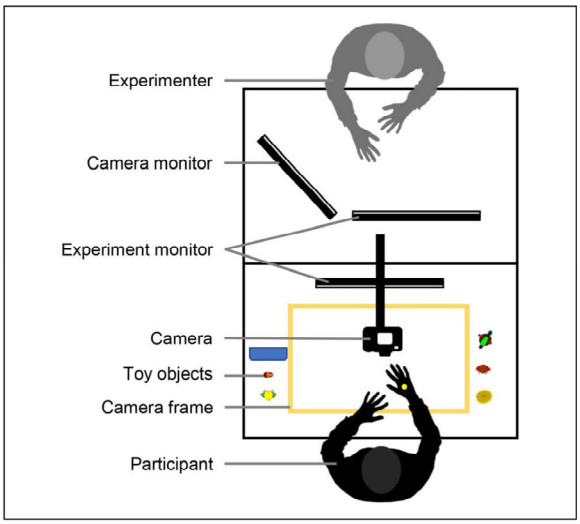


Figure 2. Encoding setup. During training, participants imitated PLAYMOBIL stories, while sitting at the filming setup. Their performance was monitored by the experimenter. Figure adapted from Siestrup et al. (2022).

During the retrieval task, participants always saw the first two steps of a demo video. Then, the video stopped and a question was displayed, which was either "Left?" or "Right?" Participants were instructed to visualize the rest of the story from memory and then answer how many steps of the entire story had been played with the left or right hand, respectively. Participants had to answer by pressing a number key (0–9) on their keyboard. We chose this approach because we wanted to encourage active retrieval of the encoded episodes without laying a special focus on aspects that would be modified during the fMRI session. Thus, we wanted to avoid asking for specific contents or steps involved in the actions. In addition, we chose this question because it could not be answered with mere gist-knowledge about the episode. Importantly, the number of steps that were conducted with the left or right hand did not change in videos containing either type of modification. Upon response delivery, the video played until the end. Participants were instructed to carefully watch the video to self-check their answers. Afterward, written feedback ("correct," "incorrect") was displayed on the screen for 1.5 sec. When feedback is included in retrieval tasks, consolidation has been shown to occur irrespective of the initial retrieval success (Rowland, 2014; Roediger & Butler, 2011). The task was self-paced, so participants could decide themselves when they wanted to proceed with the next video. As videos were not only presented once but several times (outlined below), participants were explicitly instructed to always visualize the story and not just remember previous responses they gave. Each type of question was presented equally often after each video and per session. For a schematic depiction of the retrieval task, see Figure 3.

Half of the stories were retrieved two times, the other half eight times (factor Times). Furthermore, half of the stories were retrieved during two separate sessions, that is, in a spaced manner, whereas the other half of the stories were retrieved only in Session 2, that is, in a massed manner (factor Schedule). All episodes that were retrieved in a spaced manner were trained in Session 1 either 1 or 4 times, depending on which factor level of Times they belonged to. Therefore, the retrieval task comprised 30 trials (plus two practice trials) in Session 1. The remaining repetitions, so either one our four additional trials per story, were conducted in Session 2 (30 trials). In addition, all episodes that were retrieved in a massed way were trained in Session 2 (60 trials). Thus, the retrieval task in Session 2 comprised 90 trials in total. Trials of different levels of the factor Schedule (*spaced, massed*) were not trained in an interleaved manner to keep the second retrieval of spaced items comparable to the first. Instead, trials were blocked according to the factor level of Schedule. Order of spaced and massed blocks was counterbalanced between participants. Within blocks, videos were trained in a pseudorandom order. Repetitions of the same video were also blocked together to avoid additional spacing effects on a trial basis.

To avoid additional practicing between pre-fMRI retrieval sessions, participants were informed that they would perform a similar, but different task in Session 2 and explicitly asked not to think about the episodes between experimental sessions. To control for additional practice, participants were asked at the end of each session whether they had tried to actively remember the videos before the session. None of the participants reported to have visualized the episodes outside the laboratory. At the end of Session 2, participants went through a short practice (four video trials, four question trials, one null event) of the task they would conduct during the fMRI scan. Please note that participants were informed during

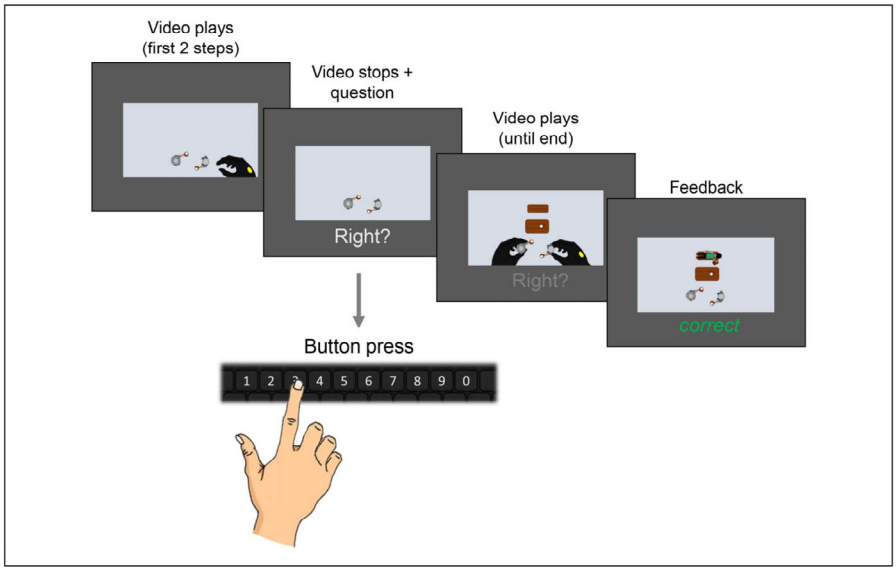


Figure 3. Schematic depiction of retrieval task. Participants were presented the first two steps of a demo video. Then, the video stopped and a question ("Left?" or "Right?") was displayed. Participants had to visualize the story from memory and answer how many steps were conducted with the left or right hand, respectively, using the number keys on their keyboard. Upon response delivery, the video proceeded to play until the end, and written feedback was provided ("correct" or "incorrect"). The task was self-paced. Because of copyright restrictions, we show schematic illustrations of the stimulus material.

the first encoding session that they will be presented with videos of their own during the fMRI session. Therefore, participants were aware from the beginning on that they would re-encounter the encoded action stories again.

fMRI Session

The fMRI session was conducted approximately 1 week after the second pre-fMRI retrieval session (range = 6–13 days; M = 7.69 days, SD = 1.31 days). Participants were presented with original and modified videos similar to the previously encoded and retrieved episodes. As in our previous study (Jainta et al., 2022), participants were told that videos of themselves playing the stories would be presented in the fMRI session. However, this was only a cover story to elevate personal identification with the videos to benefit episode reactivation. In fact, participants were never presented videos of themselves. They were fully debriefed after completing the experiment.

Each story was only shown in the original (eight videos) or one divergent form (structure or content modification; eight videos each). Before the fMRI session, stories had either been retrieved two or eight times and the total number of retrieval times had either been spaced out over the two pre-fMRI retrieval sessions or massed together in Session 2. Videos were allocated to conditions pseudorandomly so that individual videos belonged to each condition minimally two and maximally six times (3 times on average) over the course of the experiment. In addition, four novel stories were included in the fMRI session.

The fMRI experiment was divided into six blocks, each containing the 24 videos that had been previously encoded during the training sessions. Each of these 24 episodes was either presented in an original or a slightly modified version once per block and, thus, was repeated six times in total over the course of the fMRI session. Over the entire experiment, the trial order was pseudorandomized and transition probabilities were checked to ensure a balanced number of transitions between conditions. Therefore, we ensured each condition followed every other condition, including the same condition, equally often. Each block in addition contained three null events (fixation cross for 7 to 10 s) and four completely new videos (i.e., novels), leading to 18 null events and 24 novel video trials in the whole session.

During the fMRI session, the participants' task was to attentively watch the videos and answer a question after some video trials. The task was included to focus the participants' attention on the video stimuli, as applied in previous studies (Jainta et al., 2022; El-Sourani, Trempler, Wurm, Fink, & Schubotz, 2019). Questions were short descriptions (e.g., "Rescuing princess?") of stories and the participants had to indicate whether this description matched or did not match the previously displayed video. To do so, they had to press one of two buttons on a response box with the right index finger (yes) or middle finger (no), respectively (Figure 4). Question trials were pseudorandomly interspersed with video trials. Over the course of the experiment, each story (including novels) was once followed by a matching, once by a non-matching description. Thus, there were 56 question trials in total.

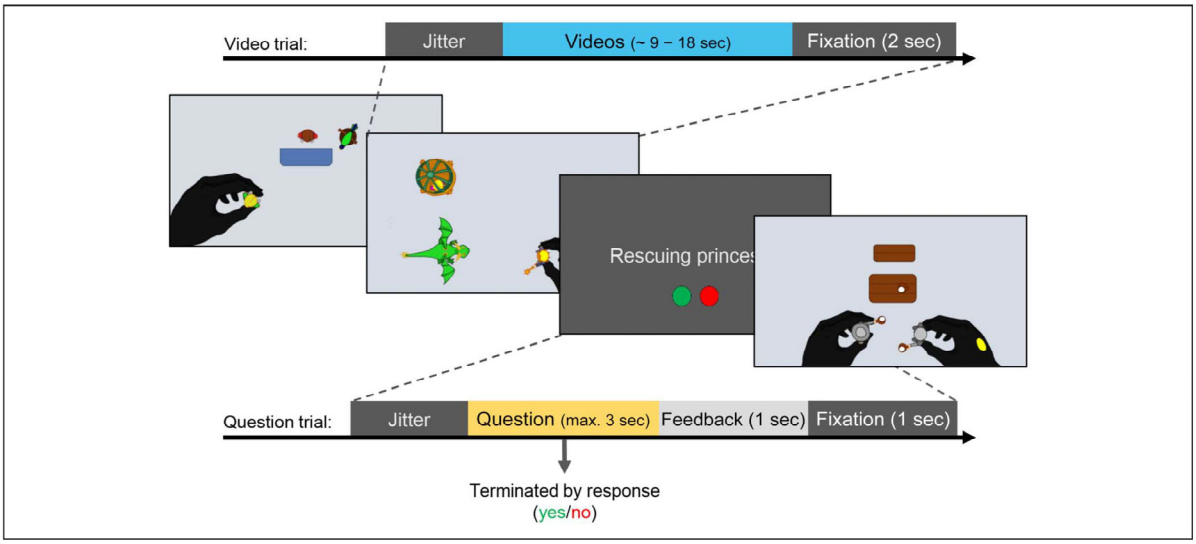


Figure 4. Schematic depiction of task during fMRI session. Video trials included a variable jitter (0, 0.5, 1, or 1.5 sec of fixation), a video showing a PLAYMOBIL story (ca. 9–18 sec), and a 2-sec ISI (fixation). Question trials consisted of a variable jitter, a question regarding the story shown in the preceding video (terminated by response or maximally 3 sec long), and a 2-sec ISI. The ISI after question trials was divided into a 1-sec feedback ("correct," "incorrect," "too late") and a 1-sec fixation. Under the question text, it was shown which button should be pressed to accept (left, green) or reject (right, red) the description.

Each block contained 9 to 10 question trials to achieve an approximately even distribution of question trials over the experiment. Maximally two successive videos were followed by questions. Like that, we aimed to ensure that question trials were neither highly irregular nor highly predictable so that participants would have to stay attentive at all times. Questions were presented for maximally three seconds or until participants responded. Participants then received one second of written feedback whether they had answered correctly, incorrectly or too late, in case no response was given.

During ISIs, a fixation cross was presented (duration: 2 sec, 1 sec after question trials). In addition, a variable jitter of 0, 0.5, 1 or 1.5 sec of fixation was added between trials for enhancement of the temporal resolution of the BOLD response (Figure 4). The fMRI task had a total duration of approximately 48 minutes.

Post-fMRI Memory Test

The memory test was conducted as previously described (Jainta et al., 2022). Participants were not informed that their memory for episodes would be tested at the end of the experiment.

Immediately after the fMRI session, participants were seated in a separate room in front of a laptop and instructed to remember their encoding sessions. The stories that had been part of the fMRI were now presented in two different versions. When during the fMRI experiment, an original version of a story had been presented, this was now presented as well. Half of these stories were then additionally presented in a structure-modified version, the other half in the content-modified version. When a modified version had been presented in the scanner already, stories were presented in the same modified version as well, and additionally in the original version. Novel videos which had been encountered during the fMRI session for the first time were now presented twice in the same version to keep the number of presentations per action story equal for all conditions. Importantly, participants were not explicitly informed that modified and novel videos would be presented. The video presentation order was pseudorandomized, so that half of the stories (of each experimental condition) were first presented in a modified version followed by an original version and vice versa. Importantly, videos depicting alternative versions of the same story were not shown in direct succession.

The participants' task was to rate after each video whether they knew this exact episode from the encoding sessions, using a Likert scale including 1 (yes), 2 (rather yes), 3 (rather no) and 4 (no). Answers were collected via four marked keys on the laptop's keyboard. Please note that for the analysis of ratings, we reversed the coding of responses, so that higher ratings indicate higher acceptance. Response time was unrestricted, but participants were instructed to react quickly and intuitively. In total, the memory test comprised 56 video trials and the completion of the task took approximately 15 min.

MRI Data Acquisition and Preprocessing

MRI scans were acquired with a 3-Tesla Siemens Magnetom Prisma MR tomograph and a 20-channel head coil. Participants lay on the scanner bed in a supine position, their index and middle finger positioned on the two buttons on the response box. Movements of the head and arms were minimized by fixation with form-fitting cushions. During the scan, participants wore earplugs and headphone to attenuate scanner noise. Stimuli were projected on a screen behind the fMRI machine which participants saw through an individually adjusted mirror on the head coil.

Before functional imaging, high-resolution anatomical images (T1 weighted) were created with a 3-D multiplanar rapidly acquired gradient-echo sequence (192 slices, voxel size = 1 mm^3, repetition time = 2130 msec, echo time = 2.28 msec, flip angle = 8°, field of view = 256 × 256 mm^2). Functional images were acquired in interleaved order along the AC–PC plane using a gradient-echo EPI sequence to measure BOLD contrast (33 slices, voxel size = 3 mm^3, repetition time = 2000 msec, echo time = 30 msec, flip angle = 90°, field of view = 192 × 192 mm^2).

Imaging data were processed with SPM12 (Wellcome Trust) implemented in MATLAB (Version R2020b, The MathWorks Inc.). We applied slice time correction to the middle slice, movement correction and realignment to the mean image, co-registration of functional to structural scans, normalization of functional and structural images into standard Montreal Neurological Institute (MNI) space, and spatial smoothing using a Gaussian kernel of FWHM of 8 mm. A 128-sec high-pass temporal filter was applied.

Statistical Data Analysis

Behavioral Data Analysis

The behavioral data analysis was conducted with RStudio (R Core Team, 2020; Version 1.3.1073).

To analyze the participants' performance during the pre-fMRI retrieval sessions, we applied a 2 × 2 within-subject factorial design. Factors were retrieval Times (*8*, *2*) and Schedule (*spaced*, *massed*), and we calculated the correct answer rate for all factorial combinations collapsed over both retrieval sessions.

For analyzing data from the fMRI session and post-fMRI memory test, we applied a 2 × 2 × 2 within-subject factorial design. Factors were retrieval Times (*8*, *2*), Schedule (*spaced*, *massed*), and $\text{Modification}_{\text{FMRI}}$ (*no*, *yes*). For the analysis of behavioral performance during the fMRI session, we first excluded those 0.2% of question trials during which no answer was given. We then calculated the

error rate and mean RT (including only correct responses) for each factorial combination.

For the analysis of data collected during the post-fMRI memory test, we analyzed responses to modified (in memory test) and unmodified (in memory test) videos separately, in the following referred to as $\text{modified}_{\text{MT}}$ and $\text{unmodified}_{\text{MT}}$. For each factorial combination, we calculated the mean rating score. Please note that for this analysis, we reversed the initial coding of responses as it is more intuitive that higher scores indicate higher acceptance. In addition, we calculated the mean RTs (correct responses only) for each factorial combination. RTs can be used as indicators of how long it takes to retrieve information (correctly) from memory (Collins & Quillian, 1969). Longer RTs can be interpreted as increased difficulty of retrieval because of elevated cognitive processing demands (Noppeney & Price, 2004; Larsen & Plunkett, 1987), which may also occur when competing versions of an episode are processed. As some participants did not give any correct answers for some factorial combinations, the sample size for the analysis of RTs was reduced to 35 ($\text{unmodified}_{\text{MT}}$) and 23 ($\text{modified}_{\text{MT}}$) participants.

For all behavioral analyses, we applied a significance level of $\alpha = .05$. Data were inspected for normal distribution with the Shapiro Wilk Test and checked for outliers as defined as values higher than the 75% quartile $+3 \times$ interquartile range or lower than the 25% quartile $-3 \times$ interquartile range. When data were normally distributed or could be transformed to fit normal distribution (RTs; logarithmic transformation) and showed no extreme outliers, we employed a three-way repeated-measures analyses of variance (rmANOVA). When the prerequisites for parametric analysis were not met, we used a nonparametric three-way rmANOVA based on aligned rank data (package *ARTool*; Wobbrock, Findlater, Gergle, & Higgins, 2011) and computed post hoc pairwise comparisons using the Wilcoxon signed-ranks test (one-tailed with respect to our hypotheses). p Values were adjusted according to the Bonferroni correction for multiple comparisons (Bonferroni, 1936). As descriptive statistics, we report mean values and standard errors of the means.

fMRI Design Specifications

For the analysis of fMRI data with SPM12, we used general linear models (GLM) for serially autocorrelated observations (Worsley & Friston, 1995; Friston et al., 1994). We set up four different GLMs, in each of which the six subject-specific rigid-body transformations obtained from realignment were included as regressors of no interest. All regressors were convolved with a canonical hemodynamic response function.

We applied gray matter masking on the first level of all analyses, by using the smoothed individual normalized gray matter image (8-mm FWHM), thresholded at .2 using ImCalc in SPM12, as a binary mask (https://jpeelle.net/mri/misc/creating_explicit_mask.html). Second-level group analyses were performed with one-sample t tests across participants. To control for false positive results, we applied false discovery rate (FDR) correction and a threshold of $p < .05$ or higher (voxel level) to resulting t-maps. For completeness of our analysis and with regard to our hypotheses, when no significant activation could be detected using this threshold, we applied a threshold of $p < .001$ (uncorrected) and included the results in the Appendix.

General Times *and* Schedule *effects.* The first GLM (GLM1) aimed to examine the general effects of Times and Schedule on episodic retrieval for both original and modified episodes. We included nine regressors for video trials, one per factorial combination of $\text{version}_{\text{FMRI}}$ (*ori*, *mod*), Times (*2*, *8*) and Schedule (*spaced*, *massed*), and one for novel videos. Please note, each factor level of the factors Times and Schedule contained two original and four modified videos. All video trials were modeled as epochs with onsets time-locked to the beginning of the videos and containing the full video duration. In addition, we included two regressors for the 18 null events and the 56 question trials. Null events were modeled as epochs, questions were modeled as events. For GLM1, we calculated the first-level-t-contrasts $8 > 2$ and $2 > 8$ to investigate the effect of Times. For the analysis of Schedule effects, we built the contrasts *spaced* > *massed* and *massed* > *spaced*.

Phasic Times *and* Schedule *effects at timepoint of modification.* With the second GLM (GLM2) we investigated the phasic effect of the expectancy violation at the precise moment it occurred. Regressors were the same as in GLM1, but video trials were modeled as events and onsets were time-locked to the point in the video at which the modification occurred. For the original videos, the onset used in each case was the time that corresponded, on average, to the onset of the structural and content change in the modified videos. For novel videos, the onset was set to the middle of the video. For GLM2, we calculated the contrasts $8 > 2$, $2 > 8$, *spaced* > *massed*, and *massed* > *spaced*. In addition, we calculated these contrasts separately for modified and unmodified episodes, that is, $mod_2 > mod_8$ and $mod_8 > mod_2$ as well as $mod_{spaced} > mod_{massed}$ and $mod_{massed} > mod_{spaced}$, and accordingly for originals. To investigate how Schedule and Times influence BOLD responses to mnemonic prediction errors, we built the interaction contrasts $(mod_2 > ori_2) > (mod_8 > ori_8)$, $(mod_8 > ori_8) > (mod_2 > ori_2)$, $(mod_{spaced} > ori_{spaced}) > (mod_{massed} > ori_{massed})$ and $(mod_{massed} > ori_{massed}) > (mod_{spaced} > ori_{spaced})$. As a control, we calculated another model in which the onsets for original videos corresponded to the times at which a modification would occur. The whole brain as well as ROI analyses yielded the same results as GLM2.

Neuronal effects of later false alarm and correct rejections. We used a third GLM (GLM3) to conduct an analysis that modeled BOLD responses according to the behavioral performance from the post-fMRI memory test to investigate which brain activity predicted later false memories, that is, false alarms in the memory test. GLM3 included regressors for null events (epochs), questions, original videos, and novel videos (events). Modified video trials were split into two separate regressors: those that were later (in the memory test) erroneously accepted as originals (false alarms [*fa*], originally Ratings 1 and 2) and those that were later correctly rejected (correct rejections [*cr*], originally Ratings 3 and 4). Please note that *fa* and *cr* for those episodes that had been presented in the modified version only in the post-fMRI memory test but not during scanning were not considered. Moreover, we added three further parametric modulators to model the repeated presentation of each video for original videos, modified videos resulting in *fa*, and modified videos resulting in *cr*. Six participants who did not have at least three false alarms (*fa*) or correct rejections (*cr*) in the memory test were excluded from this particular analysis. On average, there were 7.833 *fa* trials (SD = 2.730 trials) and 8.167 *cr* trials (SD = 2.730 trials) per participant.

Effect of Times on neuronal response to later false alarms. Finally, to investigate whether learning from prediction errors modeled by GLM3 depended on the employed retrieval protocol, we set up a fourth design, GLM4. As only Times yielded effects on memory performance, we focused on this factor and split the *fa* regressor into videos, which had been seen *2* times and videos that had been seen *8* times before participants entered the scanner (fa_2, fa_8). Following the same criterion as described above (minimally three later *fa* and three *cr* per regressor), this analysis included 22 participants. We calculated the contrasts for the parametric modulators from GLM3 and GLM4 to identify those brain regions in which the BOLD response increased with the number of presentations of videos that resulted in false memories.

Regions of Interest Analyses

With regard to our hypotheses on specific brain regions involved in memory formation as well as the detection of mismatching information, we performed ROI analyses for each GLM as described below.

To further explore the effect of retrieval Schedule and Times in episodic memory, we performed an ROI analysis based on GLM1. Anatomical ROIs were the bilateral HC and the bilateral FMC, according to our hypotheses. Based on the finding by Jainta et al. (2022), we restricted the FMC analysis to ACC, which was found to be activated during episodic recall in the current paradigm. Following our whole-brain analysis, we extracted mean contrast estimates (y) against the implicit baseline for the regressors *2*, *8*, *spaced*, and *massed*, and calculated paired-samples *t* tests (two-tailed) in each region. Next, we investigated whether the different retrieval protocols influenced neuronal processing of mnemonic prediction errors with a ROI analysis based on GLM2. ROIs were the bilateral HC (anatomical ROI) and regions we previously found to respond to mnemonic prediction errors, bilateral superior parietal lobe (SPL) and bilateral inferior frontal sulcus (IFS) (functional ROIs; data from Jainta et al., 2022). Contrast estimates for separate *mod* > *ori* contrasts, allocated to the factors Schedule (*spaced*, *massed*) and Times (*2*, *8*), were extracted and rmANOVAs with the same factors were computed (nonparametric for hippocampal ROI as prerequisites for parametric analysis were not met). In addition, we conducted a ROI analysis for GLM3 to investigate the increasing parametric response to later false memories (i.e., gradual memory formation). Anatomical ROIs were the HC and the parahippocampal gyrus (PHG), both highly involved in learning (Davachi & Wagner, 2002; Köhler, Crane, & Milner, 2002; O'Reilly & Rudy, 2000; Aguirre, Detre, Alsop, & D'Esposito, 1996). Contrast estimates extracted from the three parametric modulators for *fa*, *cr*, and *ori* were first subjected to one-sample *t* tests (PHG) or one-sample Wilcoxon tests (HC) to analyze whether there was significant activity increase/decrease in these areas (one-tailed for *cr* and *fa*, two-sided for *ori*). Then, contrast estimates were compared with a rmANOVA (nonparametric for hippocampal ROI) and post hoc pairwise comparisons were conducted using paired *t* tests (one-tailed with respect to our hypotheses). We used the same ROIs to extract contrast estimates from GLM4. Here, we compared contrast estimates between the parametric modulators of fa_2 and fa_8 (paired *t* test, two-tailed). We report means and standard errors.

All anatomical ROIs were created using the automated anatomical labeling atlas (Tzourio-Mazoyer et al., 2002) from the Wake Forest University Pickatlas toolbox (Maldjian, Laurienti, Kraft, & Burdette, 2003) in SPM12. To create functional ROIs, we used the peak voxel coordinates from our previous study (contrast *mod* > *ori*; IFS: $x = 45, y = 26, z = 20$; $x = -42, y = 17, z = 23$; SPL: $x = 33, y = -61, z = 44$; $x = -30, y = -64, z = 41$) as the central points for spheres with a diameter of 6 mm. Mean contrast estimates were extracted using the MarsBar Toolbox (Brett, Anton, Valabregue, & Poline, 2002), aggregated over the left and right hemispheres.

RESULTS

Behavioral Results from Pre-fMRI Retrieval Sessions

To analyze the participants' performance during the pre-fMRI retrieval sessions, we calculated the correct answer rate for each combination of the two factors Schedule and Times and performed a nonparametric rmANOVA. We found a significant main effect of Times, $F(1, 35)$ =

190.39, $p < .001$, $\eta_p^2 = .84$, as participants gave more correct answers when they retrieved the episode 8 instead of 2 times ($M_8 = .897 \pm .008, M_2 = .667 \pm .022$). There was a trend toward higher correct answer rates in the massed condition ($M_{massed} = .805 \pm .013$, $M_{spaced} = .758 \pm .020$), but the difference was not significant, $F(1, 35) = 3.13$, $p = .09$, $\eta_p^2 = .08$. There was no significant interaction, $F(1, 35) = 1.19$, $p = .28$, $\eta_p^2 = .03$.

Behavioral Results from the Cover Task during fMRI

To ensure that participants paid attention during the scanning phase, they occasionally (after 33.33% of the videos) had to answer questions about the content of the video. A nonparametric rmANOVA on error rates during the fMRI experiment with the factor stimulus retrieval Times (*8*, *2*), Schedule (*spaced*, *massed*), and fMRI modification mode $\text{Modification}_{\text{FMRI}}$ (*yes*, *no*) revealed a trend for an interaction of $\text{Modification}_{\text{FMRI}}$ and Schedule, $F(1, 35) = 2.95$, $p = .095$, $\eta_p^2 = .078$. Descriptively, error rates were higher for original episodes, which had been retrieved in a spaced compared with a massed fashion. This difference could not be observed for modified videos ($M_{no\text{-}spaced} = .045 \pm .012$; $M_{no\text{-}massed} = .010 \pm .006$; $M_{yes\text{-}spaced} = .031 \pm .006$; $M_{yes\text{-}massed} = .033 \pm .008$). Participants generally made only few mistakes, with error rates of .028 ± .007 (mean ± standard error of mean) for unmodified videos and .032 ± .005 for modified videos. The error rate for the control condition (novel videos) was .073 ± .014.

Regarding RTs on correct trials, we found a significant main effect of $\text{Modification}_{\text{FMRI}}$, $F(1, 35) = 5.70$, $p = .022$, $\eta_p^2 = .140$, indicating that participants correctly recognized the story content of a video faster when presented with an unmodified version ($M_{no} = 951.653 \pm 26.821$ msec) compared with videos containing a modification ($M_{yes} = 973.455 \pm 27.912$ msec), although modification was task-irrelevant. Furthermore, we found a trend that RTs were longer for actions trained in a spaced ($M_{spaced} = 970.808 \pm 27.584$ msec) compared with a massed fashion ($M_{massed} = 954.300 \pm 26.939$ msec), $F(1, 35) = 3.49$, $p = .070$, $\eta_p^2 = .091$. There were no significant interactions, but we found a trend for an interaction of Schedule and Times, $F(1, 35) = 2.94$, $p = .095$, $\eta_p^2 = .078$.

Behavioral Results from the Post-fMRI Memory Test

To investigate the effects of different retrieval protocols on learning from prediction errors, we averaged rating scores and RTs to $\text{modified}_{\text{MT}}$ and $\text{unmodified}_{\text{MT}}$ videos separately for each factorial combination. Memory performance for $\text{modified}_{\text{MT}}$ videos reflects how successfully participants rejected modified videos as not matching the originally experienced episodes. In contrast, memory performance for $\text{unmodified}_{\text{MT}}$ videos reflects how successfully participants accepted unmodified videos as truly matching the originally experienced episodes. For both, successful rejection and successful detection, we examined how they were modulated by the factors Times and Schedule during retrieval.

Rejection and RTs for Modified Videos

We found a significant main effect of $\text{Modification}_{\text{FMRI}}$ on rating scores, $F(1, 35) = 12.45$, $p = .001$, $\eta_p^2 = .262$, showing reduced rejection for modified videos already presented in a modified version during fMRI ($M_{yes} = 2.429 \pm 0.103$ vs. $M_{no} = 2.125 \pm 0.100$). Moreover, there was a significant main effect of Times, $F(1, 35) = 10.37$, $p = .003$, $\eta_p^2 = .229$, suggesting that successful rejection of a modified version was reduced when the original version had been presented less often during the pre-fMRI retrieval sessions ($M_8 = 2.177 \pm 0.102$; $M_2 = 2.377 \pm 0.091$). There were no significant interaction effects (Figure 5A).

Thus, episodic prediction errors during fMRI may have resulted in additional encoding of these particular stories (i.e., learning of modified episodes), but to a greater extent for less frequently retrieved memories, whereas spaced or massed retrieval protocols had no effect. Therefore, participants' judgments on previously encoded memories were influenced by these additional memory traces and correct responses on old memories were more difficult to make.

This was also reflected in RTs. Here, participants took significantly longer to rate modified videos already presented in a modified version during fMRI ($M_{no} = 588.038 \pm 61.449$ msec; $M_{yes} = 695.545 \pm 96.154$ msec; $F(1, 22) = 4.63$, $p = .043$, $\eta_p^2 = .174$). There were neither significant interaction effects nor further main effects (Figure 5B).

Acceptance and RTs for Unmodified Videos

We found a significant main effect of $\text{Modification}_{\text{FMRI}}$, $F(1, 35) = 40.03$, $p < .001$, $\eta_p^2 = .534$, indicating that participants were more likely to accept unmodified versions as originally encoded when already presented in their original form during fMRI ($M_{no} = 3.858 \pm 0.043$; $M_{yes} = 3.753 \pm 0.039$). We further found a significant main effect of Times, $F(1, 35) = 5.49$, $p = .025$, $\eta_p^2 = .136$, reflecting higher acceptance for more frequently retrieved episodes ($M_8 = 3.825 \pm 0.029$ vs. $M_2 = 3.786 \pm 0.048$). The interaction between Schedule and Times was also significant, $F(1, 35) = 10.12$, $p = .003$, $\eta_p^2 = .224$). Post hoc analyses with Wilcoxon pairwise tests (one-tailed) did not reveal significant differences between the levels of the factors Schedule and Times (*massed-2* vs. *massed-8*: $Z = -0.87$, $p = 1$; *massed-8* vs. *spaced-8*: $Z = -1.48$, $p = .552$; *spaced-2* vs. *spaced-8*: $Z = -1.06$, $p = 1$; *massed-2* vs. *spaced-2*: $Z = -1.26$, $p = .824$). However, when

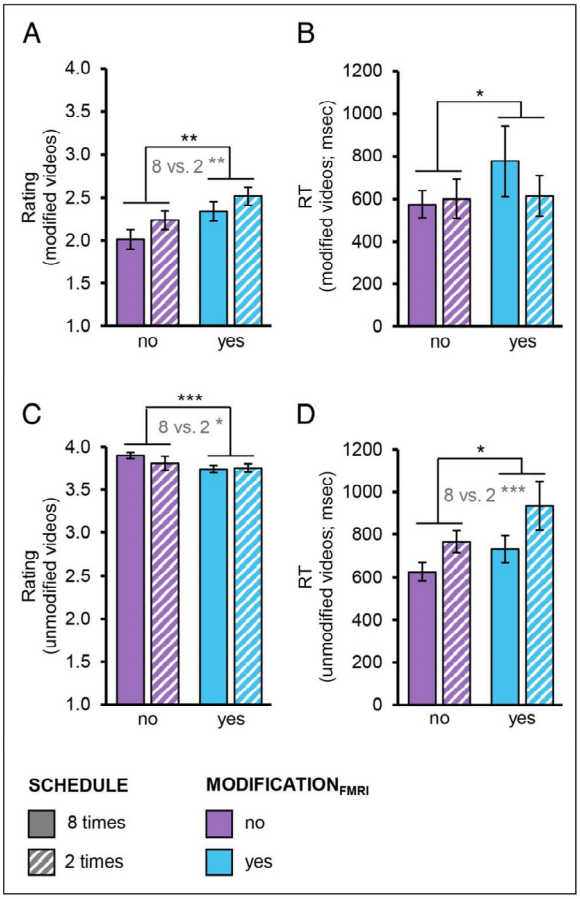


Figure 5. Behavioral results from post-fMRI memory test. Episodes had either been retrieved 8 or 2 times (factor Times) following a spaced or massed schedule (factor Schedule) before entering the fMRI session during which episodes were shown in an original or modified version (factor $\text{Modification}_{\text{FMRI}}$). Higher ratings reflect higher acceptance. (A) Ratings for modified videos. Statistics: nonparametric rmANOVA based on aligned rank data. (B) RTs for modified videos. Statistics: rmANOVA. (C) Ratings for unmodified videos. Statistics: nonparametric rmANOVA based on aligned rank data. (D) RTs for unmodified videos. Statistics: rmANOVA. *$p < .05$, **$p < .01$, ***$p < .001$. For more clarity, the factor Schedule is not depicted here. For a graph covering all factors, please see Appendix A (Figure A1).

retrieval had been scheduled in a massed fashion, participants were descriptively more likely to correctly recognize an unmodified episode after retrieving it 8 times ($M_{massed\text{-}8} = 3.795 \pm 0.042$) versus twice ($M_{massed\text{-}2} = 3.764 \pm 0.062$), and this pattern was even more pronounced in absolute values for the spaced condition ($M_{spaced\text{-}8} = 3.854 \pm 0.031$; $M_{spaced\text{-}2} = 3.809 \pm 0.054$). That is, the factor Times descriptively had an overadditive effect on the spaced Schedule. Finally, there was a significant three-way interaction of $\text{Modification}_{\text{FMRI}}$, Schedule, and Times, $F(1, 35) = 7.37, p = .01, \eta_p^2 = .174$ (Figure 5C). Accordingly, repeated triggering of the original episode led to better recall performance, especially if it had been retrieved more frequently before. Please note, however, that acceptance ratings in all conditions were at ceiling level (Figure 5C).

In line with the ratings, RTs to unmodified videos were faster when presented in their original version during fMRI ($M_{no} = 696.243 \pm 38.372$ msec; $M_{yes} = 832.948 \pm 81.726$ msec), $F(1, 34) = 7.14, p = .012, \eta_p^2 = .173$. In addition, a significant main effect of Times, $F(1, 34) = 13.63, p < .001, \eta_p^2 = .286$, indicated faster responses for more frequently retrieved episodes ($M_8 = 679.031 \pm 44.506$ msec, $M_2 = 850.160 \pm 74.799$ msec; Figure 5D).

Ratings for Novel Videos

Participants were clearly aware that novel videos had not been part of the originally encoded episodes, as indicated by floor-level rating scores (i.e., successful rejections; rating: $M = 1.021 \pm 0.015$) and very fast RTs ($M = 624.316 \pm 55.878$ msec).

fMRI Results

Neural Effects of Different Retrieval Protocols

To analyze the influence of the factor Times on brain activation during re-exposure to episodes (GLM1), we calculated whole-brain contrasts *8 > 2* and *2 > 8*. For both contrasts, we did not detect significant increases in brain activity.

Following our hypotheses, we performed ROI analyses to investigate how retrieval Times influence brain activation. Based on a previous study (Jainta et al., 2022), we used anatomical ROIs of the bilateral ACC and the bilateral HC and calculated paired-samples *t* tests (two-tailed) for each brain region to compare conditions. Contrast estimates were extracted from conditions for 2 and 8 (contrasted against the implicit baseline) using GLM1. We did not find a significant effect of retrieval Times. Descriptively, contrast estimates in ACC were lower for 8 versus 2 times, $t(35) = 0.93, p = .36, d = .155; M_2 = -0.811 \pm 0.128, M_8 = -0.872 \pm 0.121$. This descriptive pattern was also present in the HC, $t(35) = 1.18, p = .25, d = .196; M_2 = -0.116 \pm 0.044, M_8 = -0.137 \pm 0.044$.

Based on our hypothesis that spacing practice of retrieval between encoding and fMRI sessions will lead to more stable memory trace, we investigated how spaced versus massed Schedule during retrieval influence neural activation for episodes in general. Here, we calculated the whole-brain contrasts *spaced > massed* and *massed > spaced*. For the contrast *spaced > massed*, we found increased activity in the bilateral posterior cingulate cortex (pCC), the bilateral SPL/ intraparietal sulcus (IPS), and the right posterior precuneus (PCUN; Table 1, Figure 6) whereas we did not find significantly greater activity for the contrast *massed > spaced*.

Following our hypotheses that retrieval Schedule influences brain activation, we again performed ROI analyses and extracted contrast estimates from spaced and massed

Table 1. Peak Activations from Second-level Whole-brain Analyses of Retrieval Schedule

Localization	H	Cluster Extent	MNI Coordinates			t Value
			x	y	z	
spaced > massed (FDR-corrected at p < .05)						
Superior parietal lobe/intraparietal sulcus	R	25	27	−67	59	5.27
	L	66	−30	−64	56	4.95
Posterior PCUN	R	44	12	−61	32	4.24
Posterior cingulate cortex	R	60	6	−37	29	5.59
	L	l.m.	−6	−34	32	5.15

H = Hemisphere; MNI = Montreal Neurological Institute; L = Left; R = Right; l.m. = local maximum.

conditions (GLM1). We found a near-significant trend of retrieval Schedule on brain activity for episodes in HC, as contrast estimates were higher for spaced compared with massed retrieval, $t(35) = 1.90$, $p = .066$, $d = .316$; $M_{spaced} = -0.111 \pm 0.042$, $M_{massed} = -0.142 \pm 0.046$. This was not the case in ACC, $t(35) = 0.02$, $p = .99$, $d = .003$; $M_{spaced} = -0.84 \pm 0.128$, $M_{massed} = -0.842 \pm 0.120$.

Effects of Retrieval Protocols on Neural Responses to Modified Cueing

In our previous analysis of this data set, we demonstrated that several brain regions respond to mnemonic prediction errors (*mod* > *ori*; Siestrup et al., 2022). To now understand the influence of previous retrieval Times on brain activation for modified episodes, we calculated the contrasts $mod_2 > mod_8$ and $mod_8 > mod_2$ (GLM2). Regarding the $mod_2 > mod_8$ contrast, we found increased activity in ACC (Brodmann's area [BA] 24 and BA 32; Table 2, Figure 6B). The contrasts *2* > *8* as well as $ori_2 > ori_8$ did not yield significant results, indicating that this effect might be specific for modified videos. The reverse contrast, $mod_8 > mod_2$, did not result in significant effects. For further (subthreshold) results from GLM2, please see Appendix B. Furthermore, we analyzed interaction effects of retrieval protocols, which are reported in Appendix C.

To further elucidate the impact of different retrieval protocol on neural responses for mnemonic prediction errors, we conducted ROI analyses by extracting contrast estimates from GLM2 for separate *mod* > *ori* contrasts, allocated to the factors Schedule (*spaced*, *massed*) and Times (*2*, *8*). These yielded no significant results in IFS or SPL. In the hippocampal ROI, we found a significant main effect of Schedule, $F(1, 35) = 4.86$, $p = .034$, $\eta_p^2 = .122$, driven by higher contrast estimates for massed than for spaced ($M_{massed} = 0.035 \pm 0.051$, $M_{spaced} = -0.151 \pm 0.050$). There was no main effect of Times, $F(1, 35) = 0.05$, $p = .829$, $\eta_p^2 = .001$, and no significant interaction, $F(1, 35) = 0.04$, $p = .85$, $\eta_p^2 = .001$.

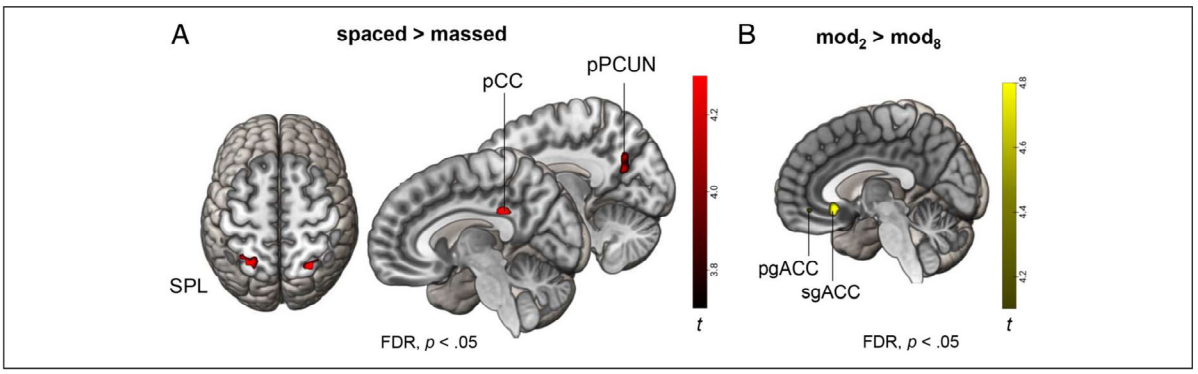


Figure 6. Whole-brain activation for retrieval Schedule and Times effects. (A) FDR-corrected *t*-map ($p < .05$) for the *spaced* > *massed* contrast in episodes (based on GLM 1). (B) FDR-corrected *t*-map ($p < .05$) for the $mod_2 > mod_8$ contrast (based on GLM 2). pCC = posterior cingulate cortex; SPL = superior parietal lobe; pPCUN = posterior precuneus; pgACC = pregenual anterior cingulate cortex; sgACC = subgenual anterior cingulate cortex.

Table 2. Peak Activations from Second-level Whole-brain Analyses of Retrieval Times for Modified Episodes

			MNI Coordinates			
Localization	*H*	*Cluster Extent*	*x*	*y*	*z*	*t Value*
mod_2 > mod_8 (FDR-corrected at p < .05)						
Subgenual ACC (BA 24)	R	13	3	23	−13	6.04
Precgenual ACC (BA 32)	L + R	7	0	44	−13	5.09

H = Hemisphere; MNI = Montreal Neurological Institute; L = Left; R = Right; BA = Brodmann's area.

Parametric Effects of Episode Repetition for Later False Alarms

We were interested in how behavioral performance was reflected in brain activity during repeated reactivation of episodes. To investigate the neural effects of memory modification, we analyzed the parametric increase in brain activation to modified episodes that later elicited false memories (GLM3). The BOLD response increased with repeated presentation of a modified episode, when it was later misclassified as original, in several regions: superior frontal gyrus (SFG) extending into ACC, as well as in inferior frontal gyrus, IPS, midcingulate cortex, pCC, and middle temporal cortex (Table 3, Figure 7A).

Table 3. Peak Activations from Second-level Whole-brain Analyses of Parametric Effect (Increase) for Later False Memories

			MNI Coordinates			
Localization	*H*	*Cluster Extent*	*x*	*y*	*z*	*t Value*
Parametric modulator (fa; FDR-corrected at p < .01)						
Superior frontal gyrus e.i. ACC (lateral BA 9 e.i. BA 10)	L	550	−9	38	50	5.48
Middle frontal gyrus	L	l.m.	−33	23	47	5.12
Superior frontal gyrus	R	6	6	38	56	4.39
Middle frontal gyrus	R	7	33	35	44	4.15
Angular gyrus/inferior parietal sulcus	R	37	57	−46	44	4.35
Angular gyrus	R	l.m.	48	−55	38	3.94
PCUN	R	7	−12	−43	41	4.01
Angular gyrus	L	124	−51	−61	38	5.30
Supramarginal gyrus	L	l.m.	−51	−52	29	4.89
Midcingulate cortex	L	57	−6	−16	32	4.89
	R	l.m.	3	−19	35	4.41
Posterior cingulate cortex	L	21	−3	−43	26	4.58
Superior temporal gyrus	R	9	−63	−22	14	4.10
Inferior frontal gyrus (pars triangularis)	L	44	−54	29	11	4.75
Inferior frontal gyrus (pars orbitalis)	L	l.m.	−51	35	−10	4.20
Subgenual ACC	R	12	3	20	−4	4.65
Middle temporal gyrus	L	407	−51	−31	−7	6.77
	R	322	57	−22	−10	7.33
Inferior temporal gyrus	R	l.m.	51	−16	−22	4.70
Middle temporal pole	R	16	48	14	−28	5.14
Cerebellum	L	52	−24	−91	−34	4.69
	R	233	30	−85	−40	5.54

H = Hemisphere; MNI = Montreal Neurological Institute; L = Left; R = Right; e.i. = extending into; l.m. = local maximum.

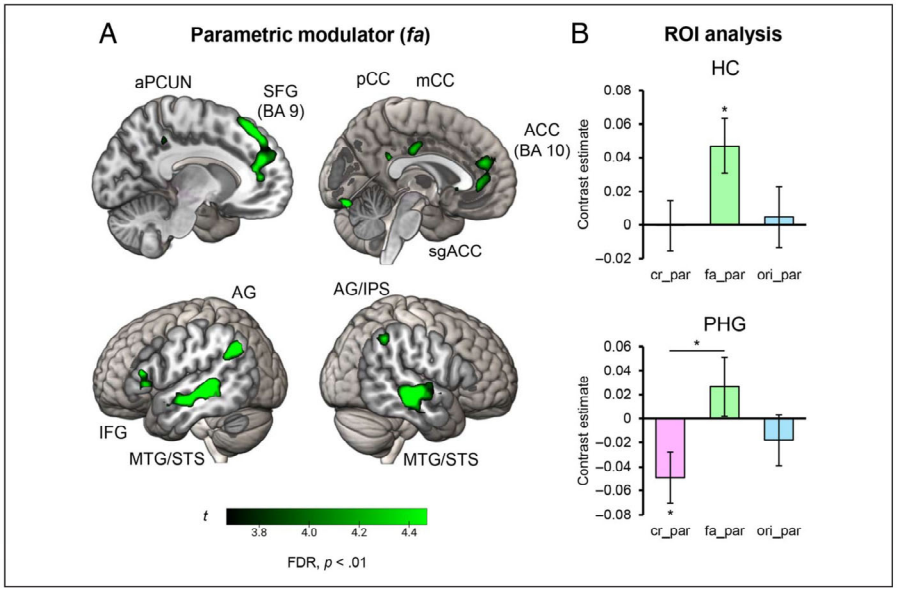


Figure 7. Results from the parametric analysis. (A) FDR-corrected *t*-map ($p < .01$) for the whole-brain contrast of the parametric modulator modeling the repeated presentation of modified episodes, which lead to false alarm responses in the post-fMRI memory test. Activation clusters indicate an increase of brain activity with repeated exposure. aPCUN = anterior precuneus; SFG = superior frontal gyrus; BA = Brodmann's area; pCC = posterior cingulate cortex; mCC = midcingulate cortex; sgACC = subgenual anterior cingulate cortex; AG = angular gyrus; IFG = inferior frontal gyrus; MTG = middle temporal gyrus; STS = superior temporal sulcus; IPS = intraparietal sulcus. (B) ROI analysis for parametric modulators modeling the repeated presentation of modified episodes, which were later false alarms (fa_par) or correct rejections (cr_par), as well as parametric response to original episodes (ori_par). Contrast estimates were extracted from hippocampal complex (HC) and parahippocampal gyrus (PHG). Statistics: one-sample *t*/Wilcoxon tests; rmANOVA (nonparametric for HC ROI) and paired *t* tests (one-tailed). *$p < .05$.

Following this whole-brain analysis, we performed a ROI analysis to specifically investigate brain responses in areas that are highly involved in memory formation, HC and PHG, as hypothesized. In HC, we found a significant increase of activation for *fa* ($Z = -2.92, p = .01$), whereas the change of activation was not significant for *cr* ($Z = -0.56, p = 1$), and *ori* ($Z = -0.18, p = 1$). In PHG, there was a significant decrease in activation for *cr*, $t(29) = -2.28$, $p = .04$, and no significant change of activation for *fa*, $t(29) = 1.07, p = .44$, and *ori*, $t(29) = -0.86, p = 1$. Using rmANOVA, we compared contrast estimates from the three parametric modulators for *fa*, *cr*, and *ori*. In HC, we found a trend for an effect of condition, $F(2, 58) = 3.0, p = .058$, $\eta_p^2 = .094$, as descriptively seen from an increase for false alarms versus a decrease for correct rejections (Figure 7B). In PHG, the same effect was significant, $F(2, 58) = 3.26, p = .045$, $\eta_p^2 = .101$. Paired *t* tests showed that contrast estimates in PHG were significantly higher for the parametric *fa* than for the parametric *cr* response, $t(29) = -2.40$, $p = .034$, $d = -.439$, as expected (Figure 7B). The remaining pairwise comparisons did not yield significant results (cr_par vs. ori_par: $t(29) = -1.05, p = .45, d = -.193$; fa_par vs. ori_par: $t(29) = 1.58, p = .19, d = .289$).

Regarding effects of how the factor Times influenced learning from prediction errors (GLM 4), we performed ROI analysis on HC and PHG. As we only found sub-threshold effects, we report the results in Appendix D.

DISCUSSION

When we encounter a previously experienced episode, the brain reactivates a memory trace, corroborating its stability while also allowing flexible adaptations to cope with upcoming changes in the world (Lee et al., 2017). In this study, we took the view that retrieval, and modification of episodic memories upon retrieval, should be influenced by how strongly memories were consolidated. Specifically, we investigated the effects of different retrieval protocols, namely, the temporal schedule of retrieval practice (in the following referred to as retrieval schedule) and the amount of practice. To separately test stability induced by these two factors, we violated episodic expectations using subtly modified retrieval cues and assessed the effects of this prediction error on brain activity and subsequent memory performance.

Evidence for Learning through Prediction Errors

Concerning the post-fMRI memory test, we replicated our previous finding that repeatedly experiencing modified episodes during the fMRI session leads to a higher misattribution of the same modified versions as part of the original episode repertoire (Jainta et al., 2022). Again, we saw that, following the presentation of the modified version in the scanner, original videos received less acceptive ratings. Still, there was a ceiling effect concerning ratings for original videos; as in all conditions, acceptance was extremely high. Our findings suggest that episodic prediction errors lead to the encoding of an alternative representation of the same episode, rather than replacing the original representation based on new information. What remains unclear is how exactly original memories were influenced by the acquisition of alternative (nonveridical) episodes. For example, memories might be modified through

remodeling of the original memory trace, or through source confusion or interference effects. A detailed discussion of this aspect is provided in Siestrup et al. (2022).

A set of neocortical areas and the hippocampal formation showed a BOLD response that increased with repeated presentation of the same modified episode when this modified episode was later mistaken as originally experienced. This cerebral reflection of incremental learning because of repeated prediction errors was found in superior and inferior frontal areas, the cingulate cortex, as well as in the middle temporal and superior parietal cortex. Hypothesis-driven ROI analyses in the hippocampal formation revealed that, whereas activation to later correct rejections decreased with repeated encounters, activation to later false alarms increased, as expected. To our knowledge, this is the first demonstration of a specific and dynamic learning effect induced by episodic prediction errors.

Hippocampus and neocortical areas are suggested to be concurrently active during learning of new events, and encoded memories become more and more independent of hippocampal activity, which initially aids encoding by detailed but short-lived storage run (Frankland & Bontempi, 2005). Thus, the hippocampus was found to support the development of a neocortical memory representation during first stimulus encounters, but then decreased in activity during further encounters with the learned stimulus (Brodt et al., 2016). Notably, in Brodt and colleagues' work, learning was based on recurrent experience of unmodified object–location information without a particular necessity to update memory. Against this backdrop, our findings speak in favor of sustained learning in both neocortical and hippocampal areas when participants repeatedly encountered slightly modified episodic cues. In our paradigm, sustained hippocampal engagement during ongoing learning could be because of the fact that the repetitions of a modified episode did not occur consecutively, but in randomized sequences intermixed with other modified and original episodes. Another unique feature was that a preexisting memory trace was activated and involved in a learning process. This mixture of familiar and novel parts in the same stimulus placed particularly high demands on associative learning, resulting in participants accepting both the original and the modified episode as belonging to the originally learned set. Note that during the fMRI session, each participant was exposed only to either the modified or the original version of an episode.

For the interpretation of our findings, it was particularly informative that successful learning of the modified episode depended on the presence of sustained hippocampal activity. This could be concluded from the fact that modified episodes that were later correctly rejected as nonoriginal showed no such increase in hippocampal areas.

Our results suggest the formation of additional memory traces, specifically "alternative versions" of previously encoded episodes, which are later erroneously taken as veridically experienced, through engagement of the episodic memory network, including the HC, medial frontal cortex, posterior cingulate, lateral temporal areas, and temporo-parietal junction (Jeong, Chung, & Kim, 2015). In addition to these well-known components for episodic encoding and retrieval, there were two brain sites that do not typically show up for episodic processes: BA 9m and mid-cingulate cortex. We refrain from broadly discussing these findings post hoc but would like to suggest how they may contribute to modification processes of episodic information, awaiting to be tested in future studies. On the one hand, the mid-cingulate cortex has been suggested to contribute to the monitoring of the other's decisions (Apps, Lockwood, & Balsters, 2013). This functional description dovetails with the fact that our participants experienced a subtle breach of episodic expectation regarding a change in the videotaped toy story. Although they later judged this modified episode to be veridically self-experienced, the brain seemed to detect this specific prediction error as if witnessing an unexpected decision. On the other hand, BA 9 has been reported to be increasingly activated for the emergence of coherence between contextual relations (Ferstl, Neumann, Bogler, & von Cramon, 2008), for example, in the presentation of syntactically independent but episodically related sentences (Ferstl & von Cramon, 2001, 2002). Although these studies were very different from ours, the concept of coherence could be an interesting starting point for understanding the role of BA 9 in episodic modification: This area could detect and code coherence between similar episodic events, which in our study led to modified episodes being accepted as originals during post-fMRI memory testing.

Effects of Pre-fMRI Retrieval Frequency

In line with our hypotheses, memory for originally encoded episodes was better when participants had retrieved episodes 8 times compared with 2 times before the fMRI session. This was evidenced by more rejective ratings for modified videos, slightly (but significantly) more acceptive ratings for unmodified videos and shorter RTs for the latter. This pattern already emerged during the pre-fMRI retrieval sessions, as episodes retrieved 8 times had an overall higher percentage of correct responses. Our findings contribute to the large body of literature reporting the same effect (Lyle et al., 2020; Karpicke & Roediger, 2007; Underwood, 1970). The benefit of eight previous repetitions was also present when modified episodes had been presented during the fMRI session, which is in line with reports that retrieval practice can protect memories from modification (Scully & Hupbach, 2020; Pastötter, Eberle, Aue, & Bäuml, 2017; Bäuml, Holterman, & Abel, 2014; Rowland, 2014; Potts & Shanks, 2012; Halamish & Bjork, 2011).

Interestingly, we did not find increased neural activation after more frequent retrievals (*8* vs. *2*) when investigating

the general effects of retrieval protocols, as would be expected in areas where memory representation is established. On the contrary, when focusing on modifications in episodes, we found evidence for decreasing activation as one would associate it with decreasing effort of retrieval ($mod_8 < mod_2$); this effect was found for ACC, an area that was reported to decrease in activity with progressing consolidation (Long et al., 2016). In addition, we found that the brain response to mnemonic prediction errors was modulated by previous retrieval times in a similar fashion, as the interaction contrast ($mod_2 > ori_2$) > ($mod_8 > ori_8$) revealed subthreshold activation in ACC and FMC (BA 9 and 10). Hence, it seems that less stable memories might be more effortful to retrieve, which yields elevated brain activation in areas associated with episodic recall (Nadel et al., 2007). Schiffer and colleagues (2013) detected activity in ACC and medial frontopolar cortex (BA 10) when comparing the influence of prediction errors on biased versus balanced internal models, suggesting that the adaptation of internal models because of prediction errors occurred more slowly for strong and faster for weak internal models (Schiffer et al., 2013). As mentioned, the FMC is activated through the experience of prediction errors (Malekshahi et al., 2016; Schiffer et al., 2013). Regarding our findings, the FMC may indicate the level of coherence between previously encoded and currently perceived modified episodes, which further corroborates our interpretation of these areas being involved in model updating. We interpret activity in these areas to indicate the detection of mismatching information as well as the increasing familiarity with perceived modifications over time while comparing them to mnemonic representations. Accordingly, in the case of weaker mnemonic representations (hence, weaker predictive models), prediction errors potentially serve as a more potent signal for model updating (i.e., learning). These findings extend the general functional view of ACC as a region involved in conflict processing (Vassena, Holroyd, & Alexander, 2017; Botvinick, Cohen, & Carter, 2004) and learning from (prediction) errors to adapt behavior (Vassena et al., 2017; Rushworth, Noonan, Boorman, Walton, & Behrens, 2011).

It has previously been discussed that there is likely no linear relationship between the number of retrievals and later retention (Rowland, 2014; Roediger & Butler, 2011). Potentially, most mnemonic benefit is gained from increasing the retrieval frequency from one to two retrievals (Lyle et al., 2020; Rawson & Dunlosky, 2011). Therefore, it could be that several meaningful changes in neuronal activation arise at this threshold. In this study, the lowest retrieval frequency already exceeded this potential level of highest consolidation benefit, which might explain our partially subtle findings.

Effects of Spaced versus Massed Retrieval Protocols

Our data did not confirm previous findings that spaced retrievals result in better memory for originally encoded episodes compared with massed retrievals (Latimier, Peyre, & Ramus, 2021; Lyle et al., 2020; YeckehZaare et al., 2019; Kang, 2016; Gerbier & Toppino, 2015; Dempster, 1989; Underwood, 1970). There is still an ongoing debate on how to design spaced practice schedules to maximize the positive influence on retention. While some suggest that spacing out practice to two sessions is already highly effective (Gerbier & Toppino, 2015), others report that the spacing effect is more powerful for a higher number of distributed sessions (Dobson et al., 2017). In addition, some researchers suggest that the first of multiple retrieval sessions should occur shortly after initial encoding, like in our paradigm, as this maximizes the likelihood of successful retrieval. However, others argue that this limits the benefit for long-term retention as the first retrieval might not be effortful enough (Kang, 2016; Roediger & Butler, 2011). Our paradigm followed the naturalistic timeline of first encoding an episode into memory and then retrieving it at different timepoints, which contributes to memory consolidation. Similar designs have been used previously in memory research when investigating the influence of massed versus spaced retrieval on retention (e.g., Lyle et al., 2020; Karpicke & Bauernschmidt, 2011; Landauer & Björk, 1978). However, practice might only be truly "massed" if carried out during a single encounter (e.g., Kang, 2016). Encountering episodes in the encoding as well as in the retrieval session might have decreased the differences between spaced and massed retrieval in the current work.

The fMRI analysis of general retrieval protocol effects revealed significant activation for episodes that had been retrieved following a spaced schedule in SPL/IPS, PCUN, and pCC, which are all involved in episodic remembering. Hippocampal involvement was slightly higher (trend) in the spaced condition, as revealed by the ROI analysis (Jainta et al., 2022; Sestieri, Shulman, & Corbetta, 2017; Rugg & Vilberg, 2013; Trimble & Cavanna, 2008; Cavanna & Trimble, 2006; Iidaka, Matsumoto, Nogawa, Yamamoto, & Sadato, 2006; Wagner, Shannon, Kahn, & Buckner, 2005). Interestingly, we found that for the processing of prediction errors, massed retrieval generally led to increased brain activation in comparison to spaced retrieval. The interaction analysis concerning the influence of retrieval schedule on the processing of mnemonic prediction errors ($mod_{massed} > ori_{massed}$) > ($mod_{spaced} > ori_{spaced}$) yielded subthreshold activation in pre- and postcentral gyrus as well as in the superior temporal gyrus. The ROI analysis revealed higher hippocampal engagement in the processing of prediction errors in the massed condition. However, because of the lack of a behavioral effect, it is difficult to interpret these findings in terms of memory stability. Descriptively increased error rates and RTs during the fMRI task for episodes from the spaced condition might indicate that they were more difficult to retrieve, that is, less consolidated. Notably, this interpretation is speculative, because the fMRI task was designed to ensure constant attention and not to probe memory. However,

together with the lack of behavioral effects, it opens the possibility that our spaced versus massed retrieval schedule was not yet quantitatively sufficient to benefit from spaced retrieval and to produce behavioral effects. As outlined above, more research is needed to identify an optimal spacing strategy. We believe that repeating our study with an updated spacing schedule will benefit the understanding of its influence on memory consolidation on a neuronal level.

Limitations

In this study, we found effects of retrieval schedule and retrieval times for modified, but not original, videos. When interpreting the results of this study, we must acknowledge that the statistical power of modified videos compared with originals may be stronger because of an unbalanced number of trials within the two conditions. Participants were presented with twice as many modified videos (16) compared with originals (eight) during fMRI. We cannot rule out the possibility that null-results in original videos may result from a lower number of videos per condition. However, we showed in a previous analysis that subdividing modified videos into two different types of expectation violation including eight videos per condition still led to substantial brain responses (Siestrup et al., 2022). Compared with original videos, both violation types shared activity patterns in frontal and parietal areas. Although these previous results increase the reliability of our present results, further research is still needed.

Furthermore, alternative explanations for our behavioral findings can be considered. On the one hand, it is possible that modified episodes presented during the fMRI session were later endorsed as originals because of the additional recent encoding opportunity. In this respect, it is worth noting that novel videos, which were also repeatedly encountered during the fMRI session, were not mistaken for original episodes by the participants. Therefore, it is unlikely that recency drove the observed effects. On the other hand, it is possible that participants mistook modified episodes as original ones after several repetitions during the fMRI session because of increased familiarity with these episodes. However, familiarity alone would not account for the clear distinction of neuronal responses to later false alarms and correct rejections, as all modified videos were equally familiar. Nevertheless, further research is needed to understand the direct contribution of prediction errors to memory modification.

Conclusion

Episodic memories can change upon retrieval, and episodic prediction errors may trigger this process. In this study, we found that episodic prediction errors led to a dynamic learning process in the episodic memory network, resulting in the acceptance of false memories as veridical. Moreover, different retrieval protocols modulated the brain responses to episodic prediction errors and changed the subsequent propensity to learn from these events. These observations may be a fruitful starting point for further investigation of episodic prediction errors and their role in episodic memory changes.

APPENDIX A

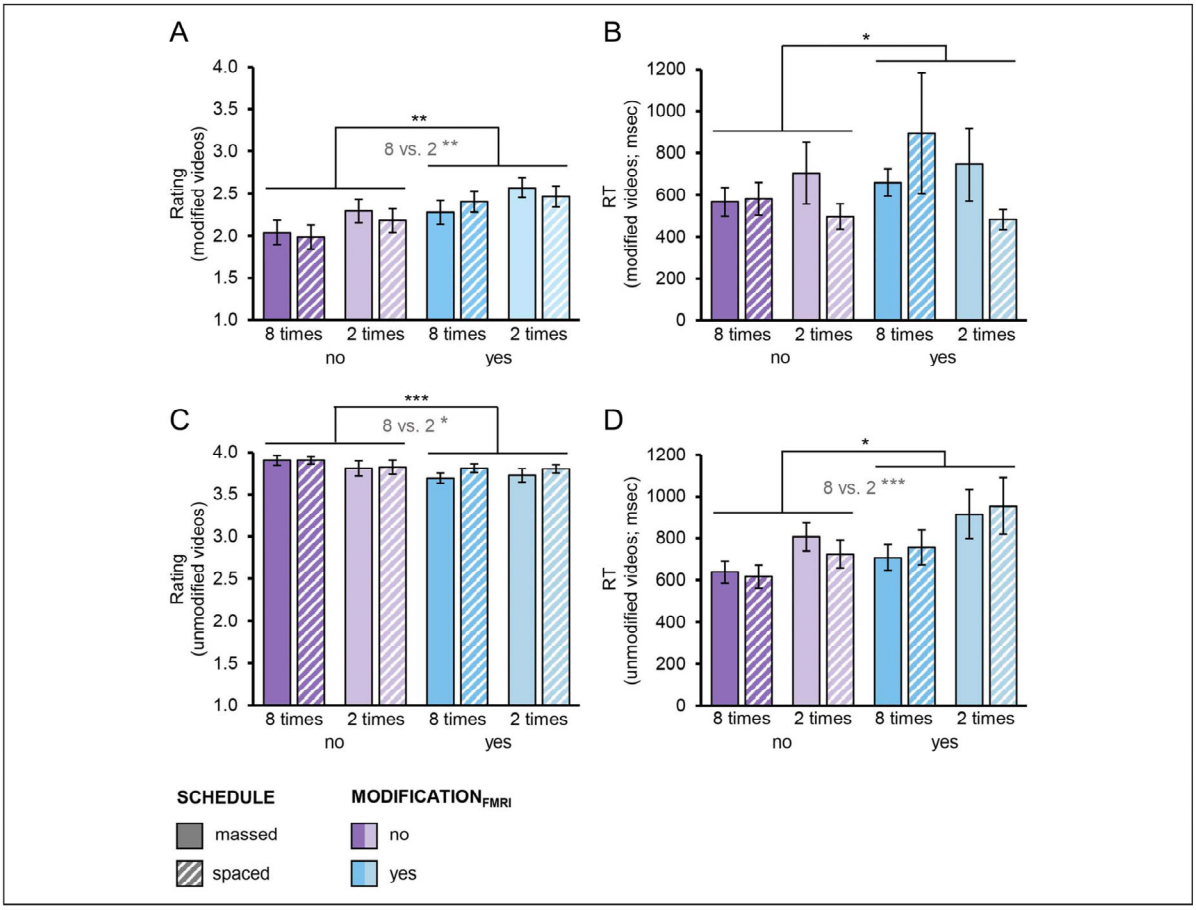


Figure A1. Behavioral results from post-fMRI memory test for all factors. Episodes had either been retrieved 8 or 2 times (factor TIMES) following a spaced or massed schedule (factor SCHEDULE) before entering the fMRI session during which episodes were shown in an original or modified version (factor $\text{MODIFICATION}_{\text{FMRI}}$). Higher ratings reflect higher acceptance. (A) Ratings for modified videos. Statistics: nonparametric rmANOVA based on aligned rank data. (B) RTs for modified videos. Statistics: rmANOVA. (C) Ratings for unmodified videos. Statistics: nonparametric rmANOVA based on aligned rank data. (D) RTs for unmodified videos. Statistics: rmANOVA. *$p < .05$, **$p < .01$, ***$p < .001$.

APPENDIX B

To investigate the effects of previous retrieval SCHEDULE on brain activation for modified episodes, we calculated the contrasts $mod_{spaced} > mod_{massed}$ and $mod_{massed} > mod_{spaced}$ (GLM2). Both contrasts did not reveal significant activation with FDR correction. However, for $mod_{massed} > mod_{spaced}$, we found subthreshold activation in cerebellum (right hemisphere: 23 voxels, $x = 42, y = -46, z = -31$; left hemisphere: 24 voxels, $x = -18, y = -61, z = -43$). No contrast for original episodes yielded significant results. Subthreshold activation for $ori_8 > ori_2$ was found in the right superior frontal sulcus and insula, as well as in the left middle frontal gyrus. For $ori_{spaced} > ori_{massed}$, we detected subthreshold activation in the left central sulcus and in the superior temporal sulcus (Table B1).

Table B1. Peak Activations from Second-level Whole-brain Analyses of Retrieval TIMES and SCHEDULE for Original Episodes

			MNI Coordinates			
Localization	*H*	*Cluster Extent*	*x*	*y*	*z*	*t Value*
$ori_8 > ori_2$ *(uncorrected at p < .001)*						
Superior frontal gyrus	R	19	21	59	23	3.83
Middle frontal gyrus	L	10	−39	53	17	3.77
Insula	R	11	39	14	−16	4.43
$ori_{spaced} > ori_{massed}$ *(uncorrected at p < .001)*						
Sulcus centralis	L	76	−57	−16	47	4.70
Temporal pole (superior temporal gyrus)	L	40	−42	5	−22	5.15

H = Hemisphere; MNI = Montreal Neurological Institute; L = Left; R = Right.

APPENDIX C

To further understand the influence of previous retrieval TIMES and SCHEDULE while controlling for potential baseline effects (GLM2), we analyzed interaction effects of retrieval protocols for mnemonic prediction errors. First, we investigated the influence of retrieval times on modified episodes in comparison to original episodes. The $(mod_2 > ori_2) > (mod_8 > ori_8)$ contrast revealed subthreshold activity in the left middle frontal gyrus (MFG; BA 9 and BA 10), bilateral SFG, right angular gyrus (AG), and left pregenual ACC (Table C1). Contrasting $(mod_{massed} > ori_{massed}) > (mod_{spaced} > ori_{spaced})$, we found a subthreshold activity in the bilateral postcentral gyrus, left superior temporal gyrus, and left amygdala (Table C1).

Table C1. Peak Activations from Second-level Whole-brain Analyses of Interaction Effects

			MNI Coordinates			
Localization	*H*	*Cluster Extent*	*x*	*y*	*z*	*t Value*
$(mod_2 > ori_2) > (mod_8 > ori_8)$ *(uncorrected at p < .001)*						
Angular gyrus	R	27	60	−55	32	4.49
Superior frontal gyrus	L	23	−18	53	29	4.01
	R	40	18	59	20	4.66
Medial superior frontal gyrus/mesial frontal cortex (BA 9)	L	78	−6	59	20	4.44
Middle frontal gyrus	L	8	−36	50	17	3.65
Caudate nucleus	R	4	6	14	8	4.21
ACC	R	8	6	26	−7	3.55
Pregenual ACC/straight gyrus	L	53	−3	47	−19	4.45
Right insula	R	7	36	14	−19	3.80
$(mod_{massed} > ori_{massed}) > (mod_{spaced} > ori_{spaced})$ *(uncorrected at p < .001)*						
Postcentral gyrus	L	165	−48	−22	56	4.69
	R	22	48	−13	32	4.16
Superior temporal gyrus	R	9	66	−10	5	3.87
Hippocampus	R	8	21	−7	−16	4.35
Superior temporal gyrus	L	40	−39	8	−22	4.54
Amygdala	L	l.m.	−30	2	−16	4.00

H = Hemisphere; MNI = Montreal Neurological Institute; L = Left; R = Right; BA = Brodmann's area; l.m. = local maximum.

APPENDIX D

With regard to GLM4, we were interested whether learning from prediction errors during scanning predicted later false memories during the memory test. To investigate whether BOLD responses were influenced by the former retrieval protocol, we reanalyzed the behavioral performance during the post-fMRI memory test. As TIMES, but not SCHEDULE, affected the stability of episodic memories as indicated by better recall, we tested whether different retrieval TIMES were reflected in brain areas involved during memory formation. Based on model GLM4, we extracted contrast estimates for parametric *fa* responses separately for episodes that had been retrieved 2 and 8 times. As a result, we found that the increasing BOLD effect did not depend on previous retrieval times in HC, $t(21) = -1.22$, $p = .23$, $d = -.261$; $M_{fa2\_par} = 0.014 \pm 0.020$, $M_{fa8\_par} = 0.072 \pm 0.039$, or PHG, $t(21) = -0.82$, $p = .42$, $d = -.176$; $M_{fa2\_par} = 0.014 \pm 0.029$, $M_{fa8\_par} = 0.056 \pm 0.048$.

Acknowledgments

The authors thank Monika Mertens, Lena Puder, Simon Wieczorek, Jamuna Halscheid, Leandra Feldhusen, and Anne Glombitza for their help during data collection and Annika Garlichs, Helena Sydlik, Christin Schwarzer and Yuyi Xu for further assistance. We thank Falko Mecklenbrauck for helpful comments on an earlier draft of this manuscript.

Reprint requests should be sent to Sophie Siestrup, Department of Psychology, University of Münster, Fliednerstraße 21, 48149 Münster, Germany, or via e-mail: s.siestrup@uni-muenster.de.

Data Availability Statement

The data sets generated for this study are available on request to the corresponding author.

Author Contributions

Sophie Siestrup: Formal analysis; Investigation; Methodology; Visualization; Writing—Original draft; Writing—Review & editing. Benjamin Jainta: Formal analysis; Investigation; Methodology; Visualization; Writing—Original draft; Writing—Review & editing. Sen Cheng: Conceptualization; Writing—Review & editing. Ricarda I. Schubotz: Conceptualization; Funding acquisition; Methodology; Resources; Supervision; Writing—Original draft; Writing—Review & editing.

Funding Information

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) – project numbers 419037023 and 419037518. The funders had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

Ethics Statements

The study was conducted in accordance with the Declaration of Helsinki and approved by the Local Ethics Committee of the University of Münster. Participants signed an informed consent before participation.

REFERENCES

Aguirre, G. K., Detre, J. A., Alsop, D. C., & D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cerebral Cortex*, *6*, 823–829. https://doi.org/10.1093/cercor/6.6.823, PubMed: 8922339

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences*, *21*, 573–576. https://doi.org/10.1016/j.tics.2017.05.001, PubMed: 28583416

Apps, M. A. J., Lockwood, P. L., & Balsters, J. H. (2013). The role of the midcingulate cortex in monitoring others' decisions. *Frontiers in Neuroscience*, *7*, 251. https://doi.org/10.3389/fnins.2013.00251, PubMed: 24391534

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology*, *192*, 101821. https://doi.org/10.1016/j.pneurobio.2020.101821, PubMed: 32446883

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology*, *4*, 907. https://doi.org/10.3389/fpsyg.2013.00907, PubMed: 24376428

Bäuml, K.-H. T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect: It enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1568–1581. https://doi.org/10.1037/xlm0000025, PubMed: 24933697

Bein, O., Duncan, K., & Davachi, L. (2020). Mnemonic prediction errors bias hippocampal states. *Nature Communications*, *11*, 3451. https://doi.org/10.1038/s41467-020-17287-1, PubMed: 32651370

Bonferroni, C. E. (1936). *Teoria statistica delle classi e calcolo delle probabilità, Vol. 8. Pubblicazioni del R. Istituto superiore di scienze economiche e commerciali di Firenze* (pp. 3–62).

Bosshardt, S., Schmidt, C. F., Jaermann, T., Degonda, N., Boesiger, P., Nitsch, R. M., et al. (2005). Effects of memory consolidation on human hippocampal activity during retrieval. *Cortex*, *41*, 486–498. https://doi.org/10.1016/S0010-9452(08)70189-8, PubMed: 16042025

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539–546. https://doi.org/10.1016/j.tics.2004.10.003, PubMed: 15556023

Brett, M., Anton, J. L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, Sendai, Japan. Available on CD-ROM in *NeuroImage*, Vol. 16, No. 2, Abstract 497. https://doi.org/10.1201/b14650-28

Brodt, S., Gais, S., Beck, J., Erb, M., Scheffler, K., & Schönauer, M. (2018). Fast track to the neocortex: A memory engram in the posterior parietal cortex. *Science*, *362*, 1045–1048. https://doi.org/10.1126/science.aau2528, PubMed: 30498125

Brodt, S., Pöhlchen, D., Flanagin, V. L., Glasauer, S., Gais, S., & Schönauer, M. (2016). Rapid and independent memory formation in the parietal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, 13251–13256. https://doi.org/10.1073/pnas.1605719113, PubMed: 27803331

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*, 564–583. https://doi.org/10.1093/brain/awl004, PubMed: 16399806

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*, 693–700. https://doi.org/10.1016/j.conb.2006.10.012, PubMed: 17097284

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences, U.S.A.*, *100*, 2157–2162. https://doi.org/10.1073/pnas.0337195100, PubMed: 12578977

Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: Insights from relational and item-based learning. *Journal of Neurophysiology*, *88*, 982–990. https://doi.org/10.1152/jn.2002.88.2.982, PubMed: 12163547

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*, 309–330. https://doi.org/10.1007/BF01320097

Dobson, J. L., Perez, J., & Linderholm, T. (2017). Distributed retrieval practice promotes superior recall of anatomy information. *Anatomical Sciences Education*, *10*, 339–347. https://doi.org/10.1002/ase.1668, PubMed: 27860396

Duncan, K., Curtis, C., & Davachi, L. (2009). Distinct memory signatures in the hippocampus: Intentional states distinguish match and mismatch enhancement signals. *Journal of Neuroscience*, *29*, 131–139. https://doi.org/10.1523/JNEUROSCI.2998-08.2009, PubMed: 19129391

Elsey, J. W. B., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin*, *144*, 797–848. https://doi.org/10.1037/bul0000152, PubMed: 29792441

El-Sourani, N., Trempler, I., Wurm, M. F., Fink, G. R., & Schubotz, R. I. (2019). Predictive impact of contextual objects during action observation: Evidence from functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *32*, 326–337. https://doi.org/10.1162/jocn_a_01480, PubMed: 31617822

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories-the role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011, PubMed: 25453746

Ezzyat, Y., Inhoff, M. C., & Davachi, L. (2018). Differentiation of human medial prefrontal cortex activity underlies Long-term resistance to forgetting in memory. *Journal of Neuroscience*, *38*, 10244–10254. https://doi.org/10.1523/JNEUROSCI.2290-17.2018, PubMed: 30012697

Feng, K., Zhao, X., Liu, J., Cai, Y., Ye, Z., Chen, C., et al. (2019). Spaced learning enhances episodic memory by increasing neural pattern similarity across repetitions. *Journal of Neuroscience*, *39*, 5351–5360. https://doi.org/10.1523/JNEUROSCI.2741-18.2019, PubMed: 31036763

Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of prediction error. *Neuroscience & Biobehavioral Reviews*, *68*, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004, PubMed: 27287939

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*, 581–593. https://doi.org/10.1002/hbm.20422, PubMed: 17557297

Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research*, *11*, 325–340. https://doi.org/10.1016/S0926-6410(01)00007-6, PubMed: 11339984

Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *Neuroimage*, *17*, 1599–1612. https://doi.org/10.1006/nimg.2002.1247, PubMed: 12414298

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, *6*, 119–130. https://doi.org/10.1038/nrn1607, PubMed: 15685217

Franz, E. A., Ford, S., & Werner, S. (2007). Brain and cognitive processes of imitation in bimanual situations: Making inferences about mirror neuron systems. *Brain Research*, *1145*, 138–149. https://doi.org/10.1016/j.brainres.2007.01.136, PubMed: 17349983

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210. https://doi.org/10.1002/hbm.460020402

Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, *4*, 49–59. https://doi.org/10.1016/j.tine.2015.01.001

Gershman, S. J., Monfils, M.-H., Norman, K. A., & Niv, Y. (2017). The computational nature of memory modification. *eLife*, *6*, e23763. https://doi.org/10.7554/eLife.23763, PubMed: 28294944

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 801–812. https://doi.org/10.1037/a0023219, PubMed: 21480751

Iidaka, T., Matsumoto, A., Nogawa, J., Yamamoto, Y., & Sadato, N. (2006). Frontoparietal network involved in successful retrieval from episodic memory. Spatial and temporal analyses using fMRI and ERP. *Cerebral Cortex*, *16*, 1349–1360. https://doi.org/10.1093/cercor/bhl040, PubMed: 16861334

Jainta, B., Siestrup, S., El-Sourani, N., Trempler, I., Wurm, M. F., Werning, M., et al. (2022). Seeing what I did (not): Cerebral and behavioral effects of agency and perspective on episodic memory re-activation. *Frontiers in Behavioral Neuroscience*, *15*, 793115. https://doi.org/10.3389/fnbeh.2021.793115, PubMed: 35069141

Jeong, W., Chung, C. K., & Kim, J. S. (2015). Episodic memory in aspects of large-scale brain networks. *Frontiers in Human Neuroscience*, *9*, 454. https://doi.org/10.3389/fnhum.2015.00454, PubMed: 26321939

Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy*

Insights From the Behavioral and Brain Sciences, *3*, 12–19. https://doi.org/10.1177/2372732215624708
Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1250–1257. https://doi.org/10.1037/a0023436, PubMed: 21574747
Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances Long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 704–719. https://doi.org/10.1037/0278-7393.33.4.704, PubMed: 17576148
Köhler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus*, *12*, 718–723. https://doi.org/10.1002/hipo.10077, PubMed: 12542224
Kumaran, D., & Maguire, E. A. (2007). Match–mismatch processes underlie human hippocampal responses to associative novelty. *Journal of Neuroscience*, *27*, 8517–8524. https://doi.org/10.1523/JNEUROSCI.1677-07.2007, PubMed: 17687029
Landauer, T. K., & Björk, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.
Larsen, S. F., & Plunkett, K. (1987). Remembering experienced and reported events. *Applied Cognitive Psychology*, *1*, 15–26. https://doi.org/10.1002/acp.2350010104
Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review*, *33*, 959–987. https://doi.org/10.1007/s10648-020-09572-8
Lee, J. L. C., Nader, K., & Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in Cognitive Sciences*, *21*, 531–545. https://doi.org/10.1016/j.tics.2017.04.006, PubMed: 28495311
Li, C., & Yang, J. (2020). Role of the hippocampus in the spacing effect during memory retrieval. *Hippocampus*, *30*, 703–714. https://doi.org/10.1002/hipo.23193, PubMed: 32022387
Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *Journal of Neuroscience*, *36*, 12677–12687. https://doi.org/10.1523/JNEUROSCI.1850-16.2016, PubMed: 27821577
Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. S. (2020). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, *32*, 277–295. https://doi.org/10.1007/s10648-019-09489-x
Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*, *19*, 1233–1239. https://doi.org/10.1016/S1053-8119(03)00169-1, PubMed: 12880848
Malekshahi, R., Seth, A., Papanikolaou, A., Mathews, Z., Birbaumer, N., Verschure, P. F. M. J., et al. (2016). Differential neural mechanisms for early and late prediction error detection. *Scientific Reports*, *6*, 24350. https://doi.org/10.1038/srep24350, PubMed: 27079423
Nadel, L., Campbell, J., & Ryan, L. (2007). Autobiographical memory retrieval and hippocampal activation as a function of repetition and the passage of time. *Neural Plasticity*, *2007*, 90472. https://doi.org/10.1155/2007/90472, PubMed: 18274617
Nader, K. (2015). Reconsolidation and the dynamic nature of memory. *Cold Spring Harbor Perspectives in Biology*, *7*, a021782. https://doi.org/10.1101/cshperspect.a021782, PubMed: 26354895
Nader, K., & Einarsson, E. Ö. (2010). Memory reconsolidation: An update. *Annals of the New York Academy of Sciences*, *1191*, 27–41. https://doi.org/10.1111/j.1749-6632.2010.05443.x, PubMed: 20392274
Noppeney, U., & Price, C. J. (2004). Retrieval of abstract semantics. *Neuroimage*, *22*, 164–170. https://doi.org/10.1016/j.neuroimage.2003.12.010, PubMed: 15110006
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4, PubMed: 5146491
O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, *10*, 389–397. https://doi.org/10.1002/1098-1063(2000)10:4<389::AID-HIPO5>3.0.CO;2-P, PubMed: 10985278
Pastötter, B., Eberle, H., Aue, I., & Bäuml, K.-H. T. (2017). Retrieval practice fails to insulate episodic memories against interference after stroke. *Frontiers in Psychology*, *8*, 1074. https://doi.org/10.3389/fpsyg.2017.01074, PubMed: 28701985
Poe, G. R., Walsh, C. M., & Bjorness, T. E. (2010). Both duration and timing of sleep are important to memory consolidation. *Sleep*, *33*, 1277–1278. https://doi.org/10.1093/sleep/33.10.1277, PubMed: 21061847
Potts, R., & Shanks, D. R. (2012). Can testing immunize memories against interference? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1780–1785. https://doi.org/10.1037/a0028218, PubMed: 22686838
Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current Biology*, *23*, R764–R773. https://doi.org/10.1016/j.cub.2013.05.041, PubMed: 24028960
Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*, 283–302. https://doi.org/10.1037/a0023956, PubMed: 21707204
R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.r-project.org
Reichardt, R., Polner, B., & Simor, P. (2020). Novelty manipulations, memory performance, and predictive coding: The role of unexpectedness. *Frontiers in Human Neuroscience*, *14*, 152. https://doi.org/10.3389/fnhum.2020.00152, PubMed: 32410975
Ritchey, M., Libby, L. A., & Ranganath, C. (2015). Cortico-hippocampal systems involved in memory and cognition: The PMAT framework. *Progress in Brain Research*, *219*, 45–64. https://doi.org/10.1016/bs.pbr.2015.04.001, PubMed: 26072233
Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27. https://doi.org/10.1016/j.tics.2010.09.003, PubMed: 20951630
Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*, 1432–1463. https://doi.org/10.1037/a0037559, PubMed: 25150680
Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*, 255–260. https://doi.org/10.1016/j.conb.2012.11.005, PubMed: 23206590
Rushworth, M. F. S., Noonan, M. A. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, *70*, 1054–1069. https://doi.org/10.1016/j.neuron.2011.05.014, PubMed: 21689594
Schiffer, A. M., Ahlheim, C., Ulrichs, K., & Schubotz, R. I. (2013). Neural changes when actions change: Adaptation of strong and weak expectations. *Human Brain Mapping*, *34*, 1713–1727. https://doi.org/10.1002/hbm.22023, PubMed: 22422724

Schiffer, A. M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS One*, *7*, e36445. https://doi.org/10.1371/journal.pone.0036445, PubMed: 22570715

Scully, I. D., & Hupbach, A. (2020). Different reactivation procedures enable or prevent episodic memory updating. *Hippocampus*, *30*, 806–814. https://doi.org/10.1002/hipo.23159, PubMed: 31520566

Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal cortex to episodic memory. *Nature Reviews Neuroscience*, *18*, 183–192. https://doi.org/10.1038/nrn.2017.6, PubMed: 28209980

Siestrup, S., Jainta, B., El-Sourani, N., Trempler, I., Wurm, M. F., Wolf, O. T., et al. (2022). What happened when? Cerebral processing of modified structure and content in episodic cueing. *Journal of Cognitive Neuroscience*, *34*, 1287–1305. https://doi.org/10.1162/jocn_a_01862, PubMed: 35552744

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning & Memory*, *25*, 369–381. https://doi.org/10.1101/lm.046912.117, PubMed: 30012882

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*, a021766. https://doi.org/10.1101/cshperspect.a021766, PubMed: 26238360

Sterpenich, V., Albouy, G., Darsaud, A., Schmidt, C., Vandewalle, G., Dang Vu, T. T., et al. (2009). Sleep promotes the neural reorganization of remote emotional memory. *Journal of Neuroscience*, *29*, 5143–5152. https://doi.org/10.1523/JNEUROSCI.0561-09.2009, PubMed: 19386910

Suzuki, W. A. (2007). Making new memories. The role of the hippocampus in new associative learning. *Annals of the New York Academy of Sciences*, *1097*, 1–11. https://doi.org/10.1196/annals.1379.007, PubMed: 17413005

Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *Journal of Neuroscience*, *29*, 10087–10093. https://doi.org/10.1523/JNEUROSCI.0799-09.2009, PubMed: 19675242

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., et al. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 756–761. https://doi.org/10.1073/pnas.0507774103, PubMed: 16407110

Trimble, M. R., & Cavanna, A. E. (2008). The role of the precuneus in episodic memory. *Handbook of Behavioral Neuroscience*, *18*, 363–377. https://doi.org/10.1016/S1569-7339(08)00220-8

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114, PubMed: 11752477

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289. https://doi.org/10.1006/nimg.2001.0978, PubMed: 11771995

Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *9*, 573–580. https://doi.org/10.1016/S0022-5371(70)80104-9

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*, 211–219. https://doi.org/10.1016/j.tins.2012.02.001, PubMed: 22398180

Vassena, E., Holroyd, C. B., & Alexander, W. H. (2017). Computational models of anterior cingulate cortex: At the crossroads between prediction and effort. *Frontiers in Neuroscience*, *11*, 316. https://doi.org/10.3389/fnins.2017.00316, PubMed: 28634438

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*, 445–453. https://doi.org/10.1016/j.tics.2005.07.001, PubMed: 16054861

Wiklund-Hörnqvist, C., Stillesjö, S., Andersson, M., Jonsson, B., & Nyberg, L. (2021). Retrieval practice facilitates learning by strengthening processing in both the anterior and posterior hippocampus. *Brain and Behavior*, *11*, e01909. https://doi.org/10.1002/brb3.1909, PubMed: 33094555

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 143–146). New York: Association for Computing Machinery. https://doi.org/10.1145/1978942.1978963

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—Again. *Neuroimage*, *2*, 173–181. https://doi.org/10.1006/nimg.1995.1023, PubMed: 9343600

YeckehZaare, I., Resnick, P., & Ericson, B. (2019). A spaced, interleaved retrieval practice tool that is motivating and effective. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (pp. 71–79). New York: Association for Computing Machinery. https://doi.org/10.1145/3291279.3339411

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*, 3002–3008. https://doi.org/10.1523/JNEUROSCI.5295-04.2005, PubMed: 15772360

Zhan, L., Guo, D., Chen, G., & Yang, J. (2018). Effects of repetition learning on associative recognition over time: Role of the hippocampus and prefrontal cortex. *Frontiers in Human Neuroscience*, *12*, 277. https://doi.org/10.3389/fnhum.2018.00277, PubMed: 30050418

3.3 Study 3: Minor Changes Change Memories: FMRI and Behavioral Reflections of Episodic Prediction Errors

Running title: Minor Changes Change Memories

Sophie Siestrup & Ricarda I. Schubotz (2023)

Manuscript submitted for publication/Manuscript under review

Minor Changes Change Memories: FMRI and Behavioral Reflections of Episodic Prediction Errors

Sophie Siestrup[1,2] and Ricarda I. Schubotz[1,2]

[1]Department of Psychology, University of Münster,

[2]Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster

Author Note

We have no conflict of interest to disclosure

Correspondence concerning this article should be addressed to Sophie Siestrup,

University of Münster, Fliednerstraße 21, 48149 Münster. Email: s.siestrup@uni-muenster.de

Abstract

Episodic memories can be modified, a process which is potentially driven by mnemonic prediction errors. In the present study, we used modified cues to induce prediction errors of different episodic relevance. Participants encoded episodes in the form of short toy stories and then returned for an fMRI session on the subsequent day. Here, participants were presented either original episodes or slightly modified versions thereof. Modifications consisted of replacing a single object within the episode and either challenged the gist of an episode (gist modifications) or left it intact (surface modifications). On the next day, participants completed a post-fMRI memory test that probed memories for originally encoded episodes. Both types of modifications triggered brain activation in regions we previously found to be involved in the processing of content-based mnemonic prediction errors (i.e., the exchange of an object). Specifically, these were ventrolateral prefrontal cortex, the intraparietal cortex and lateral occipitotemporal cortex. In addition, gist modifications triggered pronounced brain-responses, while those for surface modification were only significant in right inferior frontal sulcus. Processing of gist modifications also involved the posterior temporal cortex and the precuneus. Interestingly, our findings confirmed the posterior hippocampal role of detail-processing in episodic memory, as evidenced by increased posterior hippocampal activity for surface modifications compared to gist modifications. In the post-fMRI memory test, previous experience with surface modified, but not gist modified episodes, increased erroneous acceptance of the same modified versions as originally encoded. While surface level prediction errors might increase uncertainty and facilitate confusion of alternative episode representations, gist level prediction errors seem to trigger the clear distinction of independent episodes.

Keywords: fMRI, episodic memory, prediction error, memory modification

1 Introduction

Episodic memories allow us to vividly remember events from our personal past (Tulving, 2002). These memories are not always exact, but can be modified on the basis of new relevant information (Nader & Einarsson, 2010; Nader, 2015; Lee, Nader, & Schiller, 2017; Scully, Napper, & Hupbach, 2017). Such modifications are potentially fueled by mnemonic prediction errors (PEs) which occur when there is mismatch between what was expected based on memories and the true situational evidence (e.g., Sinclair & Barense, 2019; Siestrup et al., 2022). In that sense, PEs serve as learning signals to update internal predictive models, so that we can maintain valid predictions in the long run (Friston, 2005; Friston & Kiebel, 2009; Schubotz, 2015).

Memories include different kinds of details about experienced episodes which potentially differ in their relevance for our predictions. While some perceptual and contextual details are not relevant to the overall storyline of an episode, there are central details that represent the episode's gist (Robin & Moscovitch, 2017). It has been suggested that both types of episode information, gist relevant and surface level details, are encoded and support memory retrieval (e.g., Brainerd & Reyna, 2002). While some argue that the gist and details of episodes are processed in two different traces (fuzzy trace theory; Brainerd & Reyna, 2002), a different view is that episodic memories can generally be assessed from different levels of specificity which exist on a continuum (Greene & Naveh-Benjamin, 2021). The gist of an episode can be defined as "[…] the central features of a particular episode (story line)" (Robin & Moscovitch, 2017). Non-central details of an episode are much more likely to be forgotten (Sekeres et al., 2016) or influenced by misinformation, specifically if such misinformation is in line with the gist of the episode (Reyna & Lloyd, 1997; Reyna, Corbin, Weldon, & Brainerd, 2016). Furthermore, gist information is highly relevant for guiding judgment and decision making (Corbin, Reyna, Weldon, & Brainerd, 2015). It has even been proposed that the episodic

memory trace only represents the gist of an episode, while non-central details are supplemented from semantic memory during retrieval (Cheng, Werning, & Suddendorf, 2016). Taken together, gist relevant and surface level details likely differ in their predictive relevance in the context of episodic memory. Mnemonic PEs based on the two types of information should therefore elicit different neural and behavioral responses.

The aim of the present study was to investigate how unexpected episodic modifications that either do or do not change the gist of an episode influence brain activity and memory. We previously demonstrated that subtle changes during episodic cueing lead to specific brain responses for qualitatively different types of modifications, either affecting episode content or structure (Siestrup et al., 2022). Furthermore, encounters with both types of modified episodes induced false memories in a post-fMRI memory test (Siestrup et al., 2022; Jainta et al., 2022; Siestrup, Jainta, Cheng, & Schubotz, 2023). In the present work, we focused on content modifications and specifically manipulated the impact of those on the episodes' gist.

In a first experimental session, participants encoded different episodes consisting of short toy stories from video material. On the next day, participants returned for an fMRI session during which they were presented originally encoded episodes or slightly modified versions thereof. Episode videos were modified by the exchange of a single object (Siestrup et al., 2022). While some of these modifications were irrelevant for the episode's storyline (in the following termed *surface modifications*), others intentionally changed it (in the following termed *gist modifications*). This intended impact of the content modifications was validated in a behavioral pilot study. On a third day, participants completed a post-fMRI memory test to probe memory for originally encoded episodes. To this end, participants re-watched original as well as modified episode videos and had to decide if these had been encoded during the first session. Lastly, participants went through a rating task to assess their subjective evaluations of episode modifications.

4

We hypothesized that both types of modifications would trigger activation in specific brain regions we previously identified to respond to content modifications in our paradigm (Siestrup et al., 2022), including occipitotemporal cortex (OTC), fusiform gyrus (FG), (posterior) intraparietal sulcus (IPS) and ventrolateral prefrontal cortex, specifically inferior frontal gyrus (IFG) and inferior frontal sulcus (IFS) in the right hemisphere. These brain regions are known to be relevant for the processing of (new) object information in actions (El-Sourani, Trempler, Wurm, Fink, & Schubotz, 2019; Lingnau & Downing, 2015; Wiggett & Downing, 2011; Reber, Gitelman, Parrish, & Mesulam, 2005; Grill-Spector, Kourtzi, & Kanwisher, 2001). Modifications that change the gist of the story were expected to elicit more pronounced activation in said areas due to their overall higher relevance for episode content and therefore enhanced processing at both encoding and retrieval.

Additionally, we expected that there would be differences in brain activation between the two types of content modifications, for example in the hippocampal complex. When a modification leads to inconsistency with the known episode storyline, there could be increased activation in areas relevant to novelty processing. In a previous study, we found that the hippocampus (HPC) responded more strongly to novel compared to familiar episodes in our paradigm (Jainta et al., 2022). Therefore, it is possible that gist modifications elicit stronger hippocampal responses than surface modifications. Accordingly, parahippocampal areas were found to be included in processing of scene gist (Schubotz & Von Cramon, 2009; Epstein, 2005; Oliva & Torralba, 2006). However, it may be that surface modifications trigger stronger activation in HPC than do gist modifications, as it has been suggested that the HPC is more relevant to details in episodes and that gist representations become independent of the HPC, but rather after longer delays between encoding and retrieval (e.g., Winocur & Moscovitch, 2011). It has also been proposed that episode details are predominantly represented in posterior HPC, while the anterior HPC processes episode gist (Robin & Moscovitch, 2017). Therefore, it is

5

likely that both types of modifications trigger hippocampal activation, but the localization of activity within HPC might differ between the two modification types. Aside from that, there is evidence that episode gist is mediated by anterior cingulate cortex (ACC; Sekeres, Winocur, & Moscovitch, 2018) and orbitofrontal cortex (OFC; Schubotz and von Cramon 2009), so activation in these brain areas might be higher for gist than for surface modifications.

Concerning memory performance in the post-fMRI memory test, we expected to replicate our general finding that mnemonic PEs during the fMRI session trigger false memories afterwards, putatively due to internal model updating (Schiffer, Ahlheim, Wurm, & Schubotz, 2012; Schiffer, Ahlheim, Ulrichs, & Schubotz, 2013). Therefore, one hypothesis was that prediction violations lead to increased acceptance of modified episode videos as originally encoded, potentially accompanied by a decreased acceptance for unmodified videos (Siestrup et al., 2022; Jainta et al., 2022; Siestrup et al., 2023). As mentioned above, memories for non-central details of episodes are usually less accurate. Therefore, we reasoned that memory accuracy for originally encoded episodes, as assessed by false alarm and hit rates and, more generally, the area under the curve (AUC; Brady, Robinson, Williams, & Wixted, 2022), might be generally lower for surface than for gist modified episodes. In addition, prediction violation through surface and gist modifications might influence subsequent memory differently, in line with previous reports that surface level detail is more likely to be influenced by misinformation (Reyna & Lloyd, 1997; Reyna et al., 2016). Thus, false memories might arise as a consequence of prediction violation due to surface, but not, or to a lesser extent, due to gist modifications.

6

2 Materials and Methods

2.1 Participants

Forty-two volunteers (28 women, 14 men) participated in the study. All were native German speakers and right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). Participants had (corrected-to-) normal vision, intact color perception, and reported no history of neurological or psychiatric disorders. Data from three participants were excluded from analyses due to increased movement during the fMRI session (two participants; ca. 5 mm and 7 mm movement) and a benign anatomical abnormality of the brain that impeded normalization (one participant). Two participants terminated the study preliminarily for personal reasons. The final sample consisted of 37 participants (25 women, 12 men) between the age of 18 and 29 years (M_{age} = 23 years, SD = 3.17 years). Similar sample sizes have previously yielded stable results (Jainta et al., 2022; Siestrup et al., 2022). All participants gave written informed consent to participate in this study and were reimbursed with course credits or money. The study was conducted in accordance with the Declaration of Helsinki and approved by the Local Ethics Committee of the University of Münster.

2.2 Stimuli

The video stimuli were very similar to those used in our previous studies (Jainta et al., 2022; Siestrup et al., 2022). We worked with 80 short videos (duration = 8.8 – 14.76 s, M = 12.31 s) which showed an actress manipulating PLAYMOBIL® toys. Only the toy objects and the hands and arms of the actress were visible and filmed from above with a digital single-lens reflex camera (Nikon D5300). For filming, the actress wore a black pullover and black rubber gloves. Matt white paper served as a base. Objects that were needed for a particular story were positioned on a table and were only moved into view in the moment at which they appeared in

the story. Stories included six to nine action steps ($M = 7.60$ steps) and four to eleven separable objects ($M = 6.23$ objects). The exact same object appeared in only one of the stories.

Video material was edited with Adobe Premiere Pro CC (Adobe Systems Software, Dublin, Ireland, Version 12.1.2). Final video stimuli had a frame of size 1920 x 1080 pixels and a frame rate of 25 frames per second. They always started with seven frames showing only background (white base) and ended after seven frames showing the final object constellation. Throughout the study, videos were presented using the stimulus presentation software Presentation (version 20.3 02.25.19, NeuroBehavioral Systems, Berkeley, CA, USA) at a visual angle of approximately 7.3° x 13°.

There were 24 stories that existed in three different versions each. The first version was the original version which was used for encoding. Additionally, we created two modified version. In those modified versions, one single object was exchanged compared to the original. The new object was never just the same object in a different color; the new object diverged from the old one in color and shape. In one modified version, the object exchange did not affect the storyline (i.e., gist) of the episode. In the following, this type of modification will be referred to as surface modification (sm). In the other modified version, the new object changed the storyline, i.e., affected the gist of the episode (gist modification; gm). We validated our *a priori* classification of sm and gm in a behavioral pilot study and confirmed that all modifications could be identified by participants. Modifications were not introduced in the first two action steps so that the beginning of a video served as a cue for prediction. No modifications were introduced in the last two action steps either. The exact time point of the modification in each video was determined by identifying the video frame which diverged from the original version. For an example of an episode and its modified versions, see Figure 1.

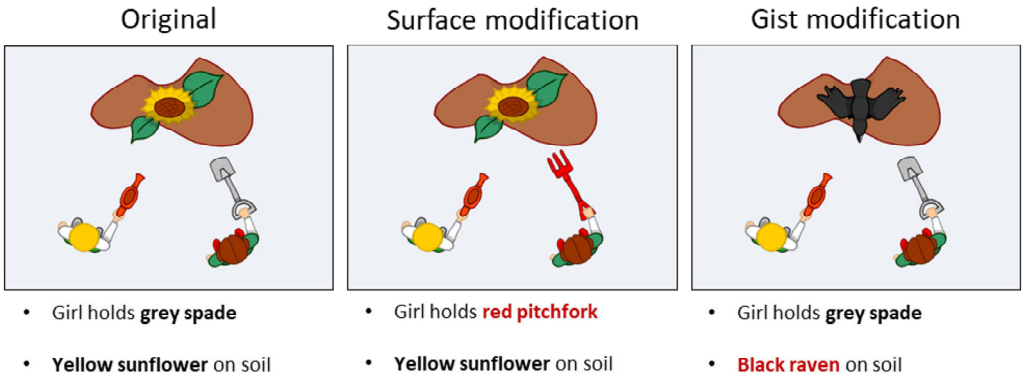


Figure 1. Example of episode modifications. Original (left) episodes were encoded by the participants. In one modified version, the object exchange did not affect the storyline of the episode (surface modification; here: grey spade is exchanged for red pitchfork). In the other modified version, the new object changed the storyline (gist modification; here: yellow sunflower is exchanged for black raven). Each participant only encountered one of the two modified version of an episode. We do not reproduce photos of our stimulus material because it is copyrighted material (PLAYMOBIL® figures); instead, we provide schematic images.

Two additional videos were used for practice trials for the encoding, fMRI and rating tasks. One of those existed in only one version and one additionally had two modified alternatives. Four more stories were first introduced in the fMRI experiment. They existed in one version only and will be referred to as novels in the following.

2.3 Procedure

2.3.1 Encoding

The encoding session was conducted in a computer laboratory at the University of Münster and lasted on average 2 hours and 9 min. For encoding, participants sat at the setup where the stimulus material had been filmed and wore black gloves like those seen in the video. Episodes were encoded by watching each video five times. Afterwards, participants had to give a detailed description to the experimenter, which had to include all objects (including their color) and the correct order of actions in the story. If the participant made a mistake, the

experimenter interrupted and corrected them, and the participant had to start the description anew, until one completely correct description was given. On average, participants needed 1.54 description attempts per story (*SD* = 0.22). The context and interpretation of the story were discussed and clarified, to ensure that all participants encoded the story the same way. Lastly, the participant was asked to summarize the story in a short sentence. Each participant encoded all 24 episodes in a randomized order. The session started with a short practice phase (2 videos) to familiarize the participant with the task. At the end of the encoding session, participants briefly practiced the tasks they would do during the fMRI session.

2.3.2 MRI session

The fMRI session was conducted on the day after the encoding. Participants went through two experimental parts, which are described in the following.

Incomplete reminder

During the anatomical measurements, participants went through a so-called incomplete reminder to initiate memory destabilization, presumably facilitating later modification (Sinclair & Barense 2018, 2019). Each video that had previously been encoded was presented, but interrupted during the step preceding the modification that would be presented during the fMRI experiment (for example while retracting the hand). Videos that would be shown in the original version during the fMRI experiment were shown in interrupted fashion as well. The time point of the interruption then matched the modified version that was later presented as an alternative in the memory test. It has been shown that allowing for complete retrieval following an incomplete cue can enhance memory (Antony, Ferreira, Norman, & Wimber, 2017). Therefore, participants had to solve a dummy task between incomplete video trials to prevent complete retrieval. This task was introduced to the participants as ‘warm-up-task’. After each video, a

number between 1 and 4 was presented for 300 ms and participants had to indicate via a button press whether the number was odd (index finger) or even (middle finger).

fMRI experiment

The fMRI experiment closely followed our previously published protocol (Jainta et al., 2022; Siestrup et al., 2022). During the fMRI session, participants were presented with videos reminiscent of the previously encoded episodes. Videos were either displayed in the original version as during encoding (8 videos), or in either the sm or gm version (8 videos each). Which stories belonged to which conditions was balanced over participants. As reported previously, all episodes were presented repeatedly to simulate natural circumstances that potentially benefit memory modification (Siestrup et al., 2022; Schiffer et al., 2012, Schiffer, Ahlheim, et al., 2013). Additionally, four novel stories were shown in the fMRI session for the first time.

The experiment consisted of six blocks. Each video was presented once per block. Additionally, each block contained three null events during which only a fixation cross was presented (duration: 7 – 10 s).

Participants were instructed to attentively watch the presented videos. They were told that after some videos they were required to solve a quick task about the previously presented video. Even though the participants were not aware of this, the task was merely included to ensure the participants' constant attention (El-Sourani et al., 2019; Jainta et al., 2022; Siestrup et al., 2022) and was not of interest for our actual research question. These task trials were characterized by the presentation of two verbs left and right of the fixation cross. One verb was closely associated with the video shown before, one was not at all associated with the video (as validated by three experimenters). Importantly, verbs did not describe any action actually depicted in the video and all associations remained intact in the case of modifications. The participants' task was to press the left button (index finger) if the left verb matched the video, and the right button (middle finger) if the right verb matched the video. Whether the associated

verb was presented on the left or right was pseudo-random and balanced between conditions for each participant. The verbs were presented for a maximum of 3 s or until participants responded. Upon response delivery, participants received 1 s written feedback whether they answered correctly, incorrectly, or too late, in case no response was given. Each video was followed by a task trial twice during the experiment, resulting in a total number of 56 task trials in the experiment.

A fixation cross was presented between trials for 2 s (1 s after task trials) to serve as an interstimulus interval. Additionally, a variable jitter of 0, .5, 1 or 1.5 s of fixation was added before each trial to enhance the temporal resolution of the BOLD response (Figure 2). In total, the fMRI task had a duration of approximately 48 min. Transition probabilities were balanced between different trial types over the whole experiment.

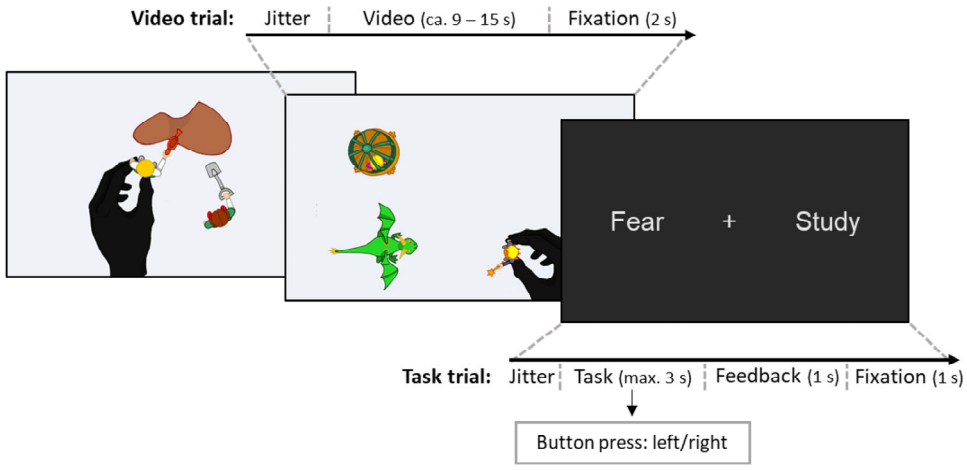


Figure 2. Schematic depiction of task during fMRI session. Video trials were comprised of a variable jitter (0, 0.5, 1, or 1.5 s of fixation), a video showing a toy story (ca. 9– 15 s) and a 2-s interstimulus interval (fixation). Task trials included a variable jitter and the presentation of two verbs (maximally 3 s long or terminated by response). The participants' task was to press the left or right button depending on which verb was more closely associated with the previously seen story. The 2-s interstimulus interval after the task was divided into a 1-sec feedback ("correct", "incorrect", "too late") and a 1-sec fixation.

2.3.3 Post-fMRI memory test

The post-fMRI memory test was conducted on the day after the MRI session and followed a similar protocol as previously reported (Jainta et al., 2022; Siestrup et al., 2022). Participants were not informed that their memory for encoded episodes would be tested at any time during the experiment.

Participants were seated in front of a computer and instructed to remember the encoding session. Then, they were presented all stories in two different versions. When modified videos had been presented during the fMRI experiment, these versions were presented again in the memory test and additionally, each original counterpart was presented. When original episodes had been presented during the fMRI experiment, they were presented again and additionally, each story was displayed either with a surface or a gist modification (in half of the cases each). Videos were presented in a pseudo-randomized order, so that half of the stories (of each experimental condition) were first presented in their original version followed by a modified version and vice versa. Novel videos were displayed twice in the same version. Versions of the same episode were not displayed in direct succession but with minimally two other videos in between. In total, the memory test comprised 56 video trials.

After each video, participants had to rate how confident they were that they had encoded exactly this episode, using a Likert scale from 1 to 6 (with 1 = 100 % no, 2 = 80 % no, 3 = 60 % no, 4 = 60 % yes, 5 = 80 % yes, 6 = 100 % yes). Participants were instructed to respond quickly and intuitively. The duration of the task was approximately 15 min.

2.3.4 Rating task

The rating task was conducted after the post-fMRI memory test on the third day of the study. Each original video was presented, followed by the same modified version that had been presented in the fMRI and/or memory test. Participants had to press a button as soon as they noticed the change in the modified video. The video was played completely and not interrupted

by the response. Participants were instructed to carefully watch the full video. They then were explicitly asked if they noticed a change or not and had to respond via a button press. This question was included to allow participants who missed to indicate the modification during the video to still rate the modification if they had seen it. If participants answered 'no', no further questions were displayed. If they had noticed the change, they were asked how unexpected they perceived the content modification on a Likert scale from 1 to 6 (with 1 = 0 % unexpected, 2 = 20 % unexpected, 3 = 40 % unexpected, 4 = 60 % unexpected, 5 = 80 % unexpected, 6 = 100 % unexpected; PE rating). Importantly, they were instructed to rate the unexpectedness in reference to the original story and to not consider that they might have noticed the modification previously during the experiment. Next, they were asked how much the change affected the story of the episode on a Likert scale from 1 to 6 (with 1 = 0 % different, 2 = 20 % different, 3 = 40 % different, 4 = 60 % different, 5 = 80 % different, 6 = 100 % different; story-change rating). Lastly, the participant had to describe the modification orally to the experimenter. The task was self-paced so that participants always started a new video themselves via a button press. The duration of this task was approximately 20 min.

2.4 MRI data acquisition and preprocessing

MR imaging was conducted with a 3-Tesla Siemens Magnetom Prisma MR tomograph (Siemens, Erlangen, Germany) using a 20-channel head coil. Participants lay supine on the scanner bed. Movements of head, arms and hands were minimized by fixation with form-fitting cushions. The participants' right index and middle finger were positioned on the two appropriate buttons on a response box. To attenuate scanner noise, participants were provided with earplugs and headphones. Visual stimuli were presented via a screen that participants saw through a mirror mounted on the head coil.

Prior to functional imaging, high resolution T1 weighted anatomical images were obtained with a 3D-multiplanar rapidly acquired gradient-echo (MPRAGE) sequence (scanning

paramters: 192 slices, slice thickness = 1 mm, TR = 2130 ms, TE = 2.28 ms, flip angle = 8°, FoV = 256 x 256 mm²). Functional images of the whole brain were acquired in interleaved order along the AC–PC plane using a gradient-echo echoplanar imaging (EPI) sequence to measure blood-oxygen-level-dependent (BOLD) contrast (scanning paramters: 33 slices, slice thickness = 3 mm, TR = 2000 ms, TE = 30 ms, flip angle = 90°, FoV = 192 x 192 mm²).

Imaging data was preprocessed with SPM12 (Wellcome Trust, London, England) implemented in MATLAB (version R2020b, The MathWorks Inc., Natick, MA, USA). Preprocessing included slice time correction to the middle slice, movement correction and realignment to the mean image, co-registration of the individual structural scans to the mean functional image, normalization of functional and structural images into the standard MNI space (Montreal Neurological Institute, Montreal, QC, Canada) on the basis of segmentation parameters and spatial smoothing using a Gaussian kernel of full-width at half maximum (FWHM) of 8 mm. A 128 s high-pass temporal filter was applied.

2.5 Statistical data analysis

2.5.1 fMRI data analysis

Design specifications

FMRI data were analyzed using general linear models (GLM) for serially autocorrelated observations (Friston et al., 1994; Worsley & Friston, 1995) implemented in SPM12. Regressors were convolved with the canonical hemodynamic response function. We used the smoothed individual normalized grey matter image (8 mm FWHM) which was thresholded at .2 using ImCalc in SPM12 to create a binary grey matter mask which was applied at the first level of analyses.

The first model (GLM1) included 15 regressors. These were original videos (ori), videos containing a surface modification (sm) and videos containing a gist modification (gm), each

comprising 48 trials. For modified videos, onsets of events were time-locked to the point in the video at which the modification occurred (time of modification). For ori trials, a hypothetical time of modification was calculated (mean of times that corresponded to points of sm and gm in the non-modified video) to serve as a comparable onset. As reported previously (Siestrup et al., 2022), these conditions were modeled as events to investigate phasic effect of the prediction violation at the precise moment of its occurrence. A parametric modulator was added to each of those regressors to model the repeated presentation of each video (descending coding). The 24 novel videos were modeled as events and onsets were timed to the middle of each video. Additionally, regressors for the 18 null events and 56 task trials were included, with onsets time-locked to their respective onsets. Null events were modeled as epochs, including the full presentation time (7 – 10 s). Task trials were modeled as events. The six subject-specific rigid-body transformations obtained from realignment were included as regressors of no interest. In sum, GLM1 included 15 regressors.

To control for participants' individual variation in the perception of sm and gm, as assessed by the rating task, we computed a second GLM2. Trials for ori and novel videos, null events and tasks were modeled as described for GLM1. Modified videos were not separated into sm and gm, but included in the model as one regressor (mod) to which we added a parametric modulator comprised of the participants' story-change rating for each video obtained from the rating task. No further parametric modulators were added. Including movement parameters, GLM2 comprised 12 regressors. The parametric analysis of differences between sm and gm gave rise to highly similar results as those we report for GLM1 with our *a priori* classification of modification types.

Whole brain analysis

We calculated first-level-*t*-contrasts for sm > ori and gm > ori as well as the direct contrasts sm > gm and gm > sm to analyze brain activity in response to the different

modification types. Additionally, we calculated the reverse contrasts, ori > sm and ori > gm to further validate the specificity of modification responses. Second-level group analyses were conducted with one-sample *t*-tests across participants. Furthermore, we computed the first-level-*t*-contrasts for each type of video vs. novel (nov) videos (ori > nov, sm > nov, gm > nov) and created the conjunction of these on the second level (Nichols et al., 2005), to replicate our previous findings of episodic retrieval in general (Jainta et al., 2022; Siestrup et al., 2022). We applied false discovery rate (FDR) correction with a threshold of $p < .05$ or higher (voxel level) to resulting *t*-maps. For completeness of our analysis and with regard to our hypotheses, when no significant activation could be detected using this threshold, we applied a threshold of $p < .001$ (uncorrected). Generally, we only report clusters with a minimum number of 10 voxels. Brain activation was visualized with MRIcroGL (version 1.2.20200331, McCausland Center for Brain Imaging, University of South Carolina, USA).

Region of interest analysis

With regard to our hypotheses, we performed region of interest (ROI) analyses to more specifically investigate which brain areas are in involved in the processing of sm and gm. We hypothesized that both, sm and gm, activate brain regions we identified to be involved in the processing of content modifications in general in our paradigm. We used functional ROIs of these regions, namely right IFS, bilateral OTC and IPS (Siestrup et al., 2022), and extracted mean contrast estimate values from the sm > ori and gm > ori contrasts using the MarsBar Toolbox (Brett, Anton, Valabregue, & Poline, 2002). Functional ROIs were created from the peak voxel coordinates we identified in our previous study (Siestrup et al., 2022) (rIFS: $x = 42$, $y = 8$, $z = 32$; lOTC: $x = -42$, $y = -58$, $z = -7$; rOTC: $x = 51$, $y = -52$, $z = -10$; lIPS: $x = -27$, $y = -61$, $z = 50$; rIPS: $x = 33$, $y = -67$, $z = 35$), which served as central points for spheres with a diameter of 6 mm.

Furthermore, we investigated the influence of sm and gm on activation in anterior HPC and posterior HPC. For creating hippocampal ROIs, we used the automated anatomical labeling atlas (Tzourio-Mazoyer et al., 2002) from the Wake Forest University (WFU) Pickatlas toolbox (Maldjian, Laurienti, Kraft, & Burdette, 2003) in SPM12 to extract maps of the left and right HPC. To divide the HPC into anterior and posterior sections, we used an anterior-posterior border at y = -21 (Poppenk, Evensmoen, Moscovitch, & Nadel, 2013). To avoid contamination between the two sections, a 2 mm coronal gap was introduced from this border in anterior and posterior directions (Li, Li, Wang, Li, & Li, 2018; Guo et al., 2020). Contrast estimates were extracted for sm > ori and gm > ori from anterior HPC and posterior HPC, aggregated over both hemispheres.

Contrast estimates for sm > ori and gm > ori were subjected to one-tailed one-sample *t*-tests to check for significant activation within ROIs. Further, we used one-tailed paired *t*-test to compare contrast estimates from sm > ori and gm > ori, according to our hypotheses for each ROI. *P*-values obtained from one-sample *t*-tests were Bonferroni-corrected for multiple comparison within each ROI.

2.5.2 Behavioral data analysis

The analysis of behavioral data was conducted with RStudio (R Core Team, 2020; version 1.3.1073).

To test the hypothesis that repeated presentations of modified videos during the fMRI decrease in memory accuracy in the post-fMRI memory test, we computed the area under the curve (AUC; Brady et al., 2022), as well as false alarm and hit rates (ratings 1, 2 and 3 aggregated as “rejection” and 4, 5 and 6 as “acceptance”). Additionally, we considered reaction times from the post-fMRI memory test, as longer reaction times in memory tasks are believed to be indicative of increased retrieval difficulty due to higher cognitive processing demands (Larsen & Plunkett, 1987; Noppeney & Price, 2004). For the analysis of AUC, hit rates, false

alarm rates and reaction times for modified videos in the post-fMRI memory test ($modified_{MT}$), a 2 x 2 within-subject factorial design with the factors $MODIFICATION_{FMRI}$ (yes, no) and $VERSION_{MT}$ (sm, gm) was applied. For analyzing reaction times for original videos in the post-fMRI memory test ($original_{MT}$), we applied a within-subject design with the factor $VERSION_{FMRI}$ (ori, sm, gm).

To again confirm our classification of sm and gm, we analyzed participants' story-change ratings from the rating task. Further, we considered the reaction time to notice changes in modified videos. The same 2 x 2 within-subject factorial design as described above was applied. In individual cases, participants' story-change ratings were inconsistent with our a priori classification. On average, these were 1.70 out of 24 videos per participant. To address this individual variation, we repeated all behavioral analyses with individually rearranged sm and gm conditions, which consistently gave rise to the same results we report on the basis of our *a priori* classification.

Furthermore, we analyzed the behavioral data from the cover task in the fMRI session by calculating the error rate and mean reaction time according to the within-subject factor $VERSION_{FMRI}$ (ori, sm, gm, nov) per participant. In only 0.77 % of all question trials, no response was given so that these trials were not further considered in the analysis.

For all reaction time analyses, only trials with correct answers were considered. Further, extreme outliers in reaction time (as defined as values above quartile 3 + 3 * interquartile range (IQR) or lower than quartile 1 – 3 * IQR) were removed from each participant's data. Reaction times were averaged over all trials of the same factorial combination for each participant. For choosing statistical tests, data were inspected for normal distribution using the Shapiro Wilk Test and checked for extreme outliers. When data were normally distributed or could be transformed to fit normal distribution (reaction times from memory test and rating task; logarithmic transformation) and included no extreme outliers, we used conventional repeated

measures analyses of variance (rmANOVA). When the prerequisites for parametric analysis were not met, we applied non-parametric rmANOVA based on aligned rank-transformed data (package ARTool; Wobbrock et al., 2011). Post-hoc pair-wise comparisons were conducted with paired *t*-tests or Wilcoxon signed rank tests (one-tailed with respect to the hypotheses, always two-tailed for explorative analysis of fMRI task).

We report mean values and standard errors of the mean. A significance level of $\alpha = .05$ was applied, and we report Bonferroni-corrected *p*-values for pair-wise comparisons (Bonferroni, 1936). If the assumption of sphericity was violated (Mauchly's test of sphericity), we report Greenhouse-Geisser-corrected degrees of freedom and *p*-values.

3 Results

3.1 Behavioral pilot study

We conducted a behavioral pilot study in an independent sample with 18 participants (M_{age} = 23.839 years, SD = 4.730, age range = 18 – 32 years, 15 women, 3 men) to validate the suitability of our stimulus material and procedure. Participants encoded episodes the same way as participants of the fMRI study did. On the next day, participants returned to complete a memory test and a rating task. Both were conducted as described for the main study.

In total, memorability of all videos was comparable, as we did not detect any extreme outliers in AUC per video (M = .950 ± .010). As expected, participants showed a reduced memory performance in the sm condition compared to the gm condition (M_{sm}= .907 ± .013; M_{gm} = .995 ± .003; Z = -3.88, p < .001) in the memory test. While we expected that all modifications, no matter if sm or gm, could be visually identified by the participants, it might take longer to become subjectively aware of sm due to the smaller impact on the storyline. Indeed, participants did not miss significantly more sm than gm (M_{sm} = 0.278 ± 0.135, M_{gm} = 0.0 ± 0.0; Z = -1.70, p = .089). Over the whole pilot study, five out of 24 sm were missed only

once each; gm were never missed. When analyzing the time it took participants to notice the modifications in the rating task, we did not detect a significant difference between conditions (M_{sm} = 907.682 ± 75.694 ms, M_{gm} = 822.913 ± 64.316 ms; $t_{(17)}$ = -1.44 , p = .084). The rating for the change of storyline was significantly higher for gm than for sm (M_{sm} = 1.304 ± 0.059, M_{gm} =4.894 ± 0.110; $t_{(17)}$ = 38.88, p < .001), as expected. Therefore, we validated that sm and gm were perceived according to our *a priori* categorization.

3.2 FMRI Results

3.2.1 Behavioral performance during fMRI session

During the fMRI session, participants solved a cover task during which two verbs were presented after some videos and participants had to indicate which verb was more closely associated with that video. A non-parametric rmANOVA on error rates with the factor $\text{VERSION}_{\text{FMRI}}$ (ori, sm, gm, nov) did not reveal significant differences ($F_{(3,108)}$ = 0.22, p = .883, ηp^2 = 0.01). For all factor levels, error rates were generally low (M_{ori} = .061 ± .009; M_{sm} = .056 ± .011; M_{gm} = .058 ± .009; M_{nov} = .100 ± .024). $\text{VERSION}_{\text{FMRI}}$ had a significant effect on reaction times in the fMRI task ($F_{(3,108)}$ = 7.89, p < .001, ηp^2 = 0.18). Post-hoc paired t-tests revealed that participants took significantly longer to respond in the task after novel videos than after any other version (ori vs. nov: $t_{(36)}$ = -4.08, p = .001; sm vs. nov: $t_{(36)}$ = -3.35, p = .012; gm vs. nov: $t_{(36)}$ = -3.48, p = .008; M_{ori} = 1267.910 ± 35.721 ms; M_{sm} = 1275.592 ± 32.560 ms; M_{gm} = 1277.184 ± 33.558 ms; M_{nov} = 1328.929 ± 35.486 ms), while other pair-wise comparisons did not reach significance ($t_{(36)}$ > -0.80, p = 1).

3.2.2 Whole brain analysis

Neural responses to modified episodic cueing

To investigate which brain regions respond to surface or gist modification in episodic cueing, we calculated the contrasts gm > ori and sm > ori. To better understand differences between different types of episodes, we also computed the reverse contrasts, ori > gm and ori > sm.

In contrast to original episodes, gist modified episodes elicited activation in bilateral IPS, inferior frontal junction (IFJ), IFS (BA 45), and FG. Additionally, we found significant activation in right angular sulcus, inferior temporal sulcus, (posterior) middle temporal gyrus (pMTG), and posterior superior temporal sulcus (pSTS) (Table 1, Figure 3A). The reverse contrast (ori > gm) revealed reduced activation of several brain regions in gist modified episodes. These were bilateral pregenual anterior cingulate cortex (ACC), right superior frontal sulcus (SFS) and left midcingulate cortex (MCC) (Table 1, Figure 3C). Interestingly, this deactivation pattern was reminiscent of the one we previously reported for novel episodes compared to reactivated ones (Siestrup et al., 2022; Jainta et al., 2022).

For surface modified episodes compared to original ones, we did not detect significant activation after correction for multiple comparisons. There was subthreshold activation (uncorrected, $p < .001$) in hypothesized brain regions, which were the right IPS and an inferior frontal area (IFJ), as well as left FG. Additionally, activation clusters were located in right medial superior frontal gyrus (SFG; BA 8), caudate nucleus and anterior insula (Table 1, Figure 3B). The reverse contrast (ori > sm) did not yield significant activation with correction for multiple comparisons. Subthreshold activation clusters were found in bilateral pSTS, as well as in left postcentral sulcus, posterior MTG and middle occipital gyrus (Table 1).

Table 1. Whole-brain activation for contrasts of modified with original episodes at FDR $p < .05$/uncorrected at $p < .001$ (voxel level)

Localization	H	Cluster extent	MNI Coordinates x	y	z	*t*-value
gm > ori (FDR-corrected at $p < .05$)						
Posterior intraparietal sulcus	R	37	36	-55	47	4.83
	L	24	-27	-55	44	4.02
Inferior frontal junction	L	74	-42	-1	35	5.72
	R	205	39	5	32	6.47
Inferior frontal sulcus BA 45	R	l.m.	45	29	20	6.16
	L	10	-39	29	17	3.83
Angular sulcus	R	24	42	-64	29	4.11
Posterior superior temporal sulcus	R	37	48	-46	23	4.26
Posterior middle temporal gyrus	R	l.m.	45	-49	11	3.75
Inferior temporal sulcus	R	107	48	-46	-10	6.27
Fusiform gyrus	R	l.m.	36	-43	-16	5.04
Middle temporal gyrus	R	l.m.	63	-46	-7	3.88
Fusiform gyrus	L	76	-42	-46	-13	5.49
ori > gm (FDR-corrected at $p < .05$)						
Superior frontal sulcus	R	10	27	41	41	4.21
Midcingulate cortex	L	22	-3	-10	38	4.29
Middle occipital gyrus	L	39	-18	-97	5	5.40
Pregenual anterior cingulate cortex	L	252	-6	35	-1	5.79
	R	l.m.	9	41	-1	4.99
sm > ori (uncorrected at $p < .001$)						
Posterior intraparietal sulcus	R	14	33	-55	47	3.81
Medial superior frontal gyrus (BA 8)	R	20	3	29	44	4.01
Inferior frontal junction	R	20	39	5	35	4.43
Caudate nucleus	R	10	12	2	23	4.59
Anterior insula	R	19	33	26	-7	4.04
Fusiform gyrus	L	14	-30	-49	-16	3.95
ori > sm (uncorrected at $p < .001$)						
Postcentral sulcus	L	30	-48	-22	38	4.60

Posterior middle temporal gyrus	L	181	-54	-61	11	5.03
Posterior superior temporal sulcus	L	l.m.	-51	-46	11	4.55
Middle occipital gyrus	L	l.m.	-48	-79	8	3.60
Posterior superior temporal sulcus	R	34	48	-37	11	4.04

Note: H = Hemisphere, MNI = Montreal Neurological Institute, L = Left, R = Right, BA = Brodmann Area, l.m. = local maximum. Only clusters with a minimum extent of 10 voxels are reported.

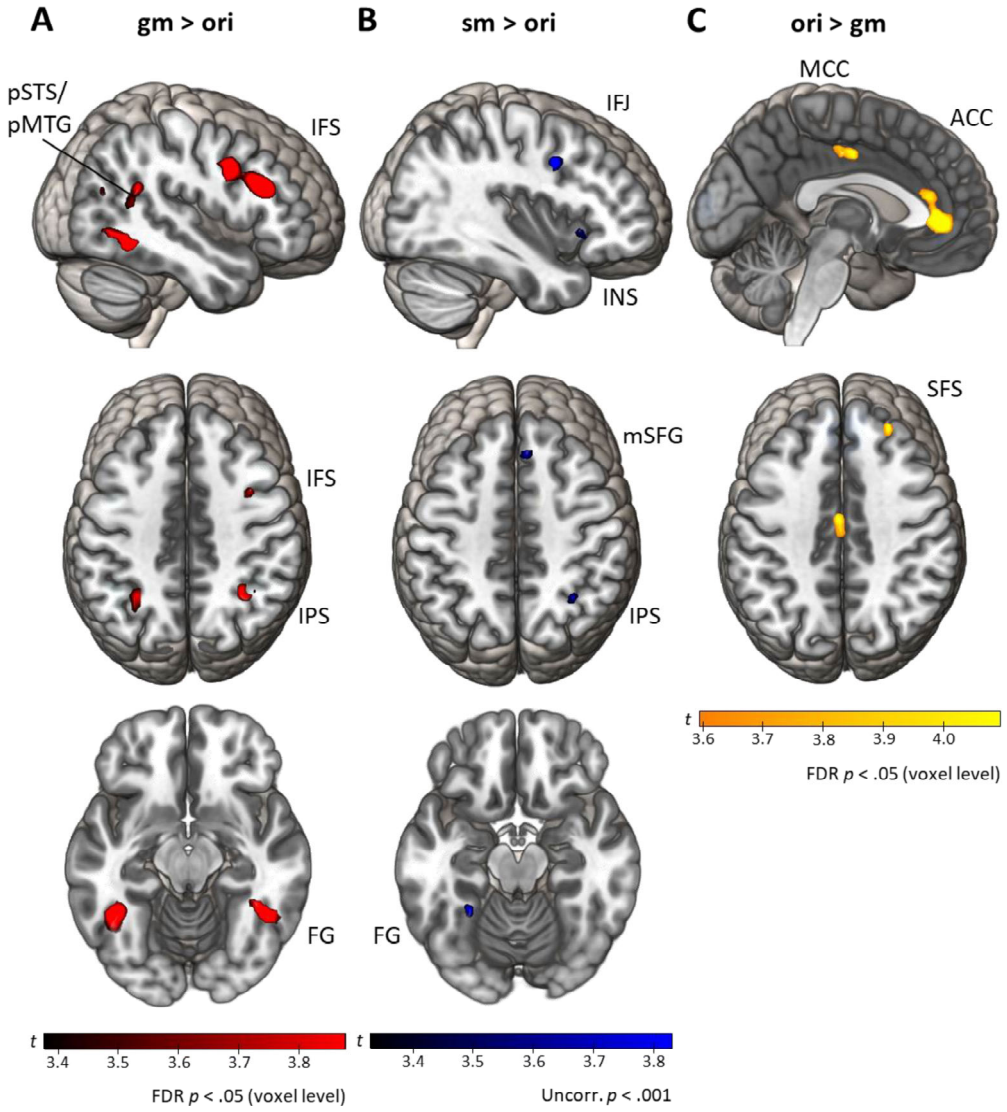


Figure 3. Whole-brain activation for episodic modifications compared to original episodes. (A) FDR-corrected ($p < .05$) *t*-map for gm > ori contrast. **(B)** Uncorrected ($p < .001$) *t*-map for sm > ori contrast. **(C)** FDR-corrected ($p < .05$) *t*-map for ori > gm contrast. pSTS = posterior superior temporal sulcus; pMTG = posterior middle temporal gyrus; IFS = inferior frontal sulcus; IPS = intraparietal sulcus; FG = fusiform gyrus; IFJ = inferior frontal junction; INS = insula; mSFG = medial superior frontal gyrus; MCC = midcingulate cortex; ACC = anterior cingulate cortex; SFS = superior frontal sulcus. Ori = original; sm = surface modification; gm = gist modification.

Additionally, we calculated the direct contrasts between both modification types (gm > sm, sm > gm). Compared to surface modifications, gist modifications elicited significantly higher activation in bilateral IFJ/IFS (BA 44, 45), pMTG, pSTS and FG. Additionally, activity was increased in left IPS and right precuneus and postcentral sulcus (Table 2, Figure 4A). Surface modifications yielded higher brain activation than gist modifications in bilateral pregenual ACC (pACC), MCC and left cuneus, as well as right SFS (Table 2, Figure 4B).

Table 2. Whole-brain activation for direct contrasts of different episodic modifications at FDR $p < .05$ (voxel level)

Localization	H	Cluster extent	MNI Coordinates x	y	z	*t*-value
gm > sm (FDR-corrected at $p < .05$)						
Posterior intraparietal sulcus	L	18	-33	-58	41	3.56
Precuneus	R	56	6	-55	41	3.66
Postcentral sulcus	R	29	57	-13	35	3.95
Inferior frontal junction	L	121	-39	5	29	6.30
Inferior frontal sulcus (BA 44)	L	l.m.	-42	20	26	5.39
Inferior frontal junction	R	105	39	14	26	5.22
Inferior frontal sulcus (BA 45)	R	l.m.	48	29	20	4.65
Posterior middle temporal gyrus	R	540	60	-61	8	7.17
Posterior superior temporal sulcus	R	l.m.	48	-43	17	5.07
Posterior middle temporal gyrus	L	382	-45	-64	14	5.25
Posterior superior temporal sulcus	L	l.m.	-54	-49	14	4.75
Fusiform gyrus	L	36	-42	-46	-13	5.70
	R	29	42	-49	-16	4.88
sm > gm (FDR-corrected at $p < .05$)						
Superior frontal sulcus	R	88	18	53	35	5.32
Midcingulate cortex	L + R	18	0	-16	32	3.80
Pregenual anterior cingulate cortex	R	584	6	35	11	8.88

	L	l.m.	-6	35	2	6.73
Cuneus/Calcarine sulcus	L	12	-6	-97	5	3.76

Note: H = Hemisphere, MNI = Montreal Neurological Institute, L = Left, R = Right, BA = Brodmann Area. l.m. = local maximum. Only clusters with a minimum extent of 10 voxels are reported.

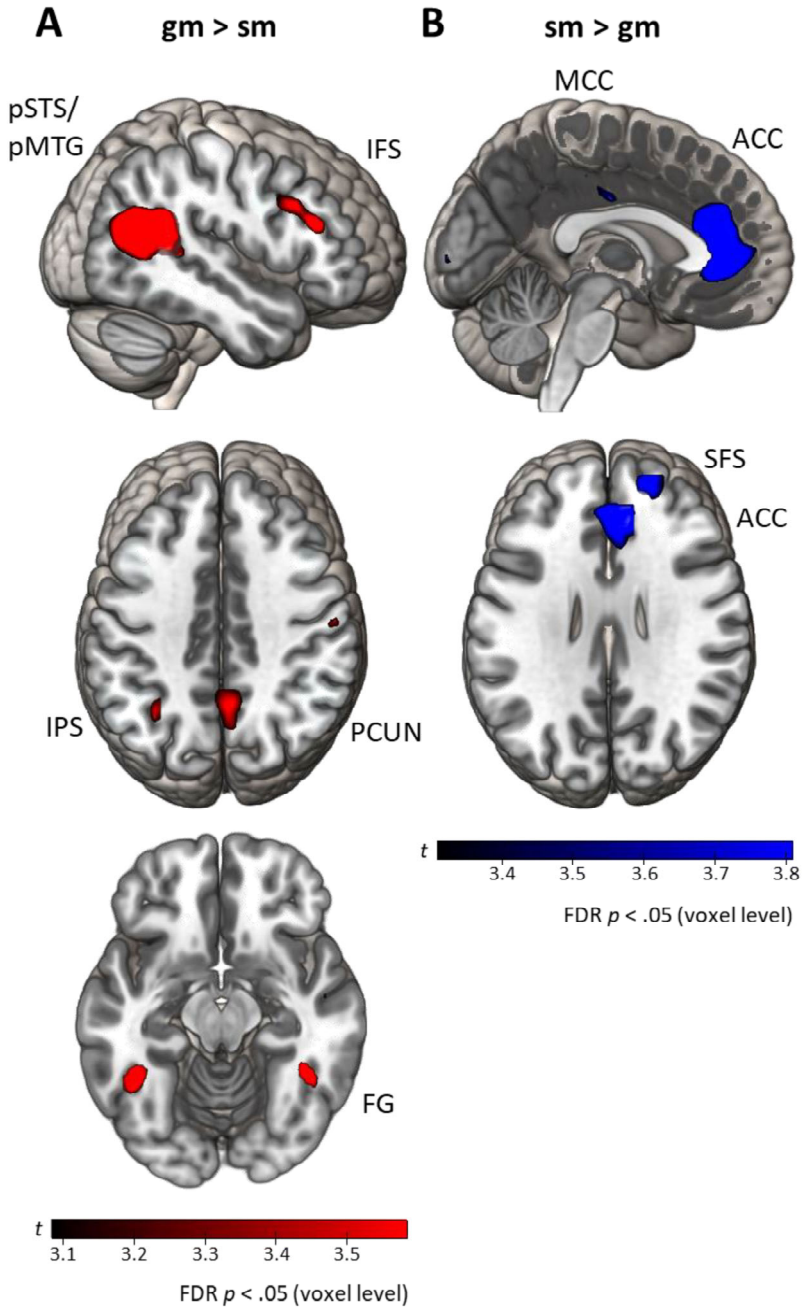


Figure 4. Whole-brain activation for direct contrasts of different episodic modifications. **(A)** FDR-corrected ($p < .05$) *t*-map for gm > sm contrast. **(B)** FDR-corrected ($p < .05$) *t*-map for sm > gm contrast. pSTS = posterior superior temporal sulcus; pMTG = posterior middle temporal gyrus; IFS = inferior frontal sulcus; IPS = intraparietal sulcus; PCUN = precuneus; FG = fusiform gyrus; MCC = midcingulate cortex; ACC = anterior cingulate cortex; SFS = superior frontal sulcus. Sm = surface modification; gm = gist modification.

Neural effects of episodic reactivation

Aside from our main research question, we aimed to replicate our previous findings concerning the neural effects of episodic recall compared to novel episodes. To this end, we calculated the conjunction of the three contrasts ori > nov, sm > nov and gm > nov. In fact, we detected activation in previously identified brain regions. These were bilateral pACC, LG, cuneus, precuneus, PCC, angular gyrus (AG), MFG and left insula (Siestrup et al., 2022; Jainta et al., 2022). Additionally, we found activation in bilateral SFS and MTG, as well as left STS and putamen (Table 3, Figure 5).

Table 3. Whole-brain activation for episodic recall at FDR $p < .001$ (voxel level)

Localization	H	Cluster extent	MNI Coordinates x	y	z	*t*-value
(ori > nov) ∩ (sm > nov) ∩ (gm > nov) (FDR-corrected at $p < .001$)						
Superior frontal gyrus	L	10	-9	26	59	4.23
	L	41	-9	44	44	4.60
Middle frontal gyrus	L	38	-36	23	50	5.09
	R	38	36	23	41	5.23
Angular gyrus	L	160	-51	-64	44	6.89
	R	114	51	-64	44	6.83
Pregenual anterior cingulate cortex	L	1379	-6	50	2	7.00
	R	l.m.	9	50	-1	6.21
Superior frontal sulcus	R	l.m.	21	47	32	6.60
Lingual gyrus	R (+ L)	3210	9	-82	-1	12.21
Cuneus	L + R	l.m.	0	-88	20	11.10
Calcarine sulcus	L (+ R)	l.m.	-6	-79	11	10.97
Precuneus, extending into posterior cingulate cortex	L + R	l.m.	0	-76	38	9.95
Superior temporal sulcus	L	72	-66	-31	-4	5.55
Middle temporal gyrus	L	l.m.	-63	-28	-13	4.33
	R	44	63	-22	-16	5.47
Anterior superior temporal sulcus	L	26	-51	8	-25	4.51
Putamen	L	24	-24	5	-4	4.28
Insula	L	l.m.	-36	11	-7	4.12
Anterior middle frontal gyrus	R	11	45	53	-4	4.28

Note: H = Hemisphere, MNI = Montreal Neurological Institute, L = Left, R = Right, BA = Brodmann Area, l.m. = local maximum. Only clusters with a minimum extent of 10 voxels are reported.

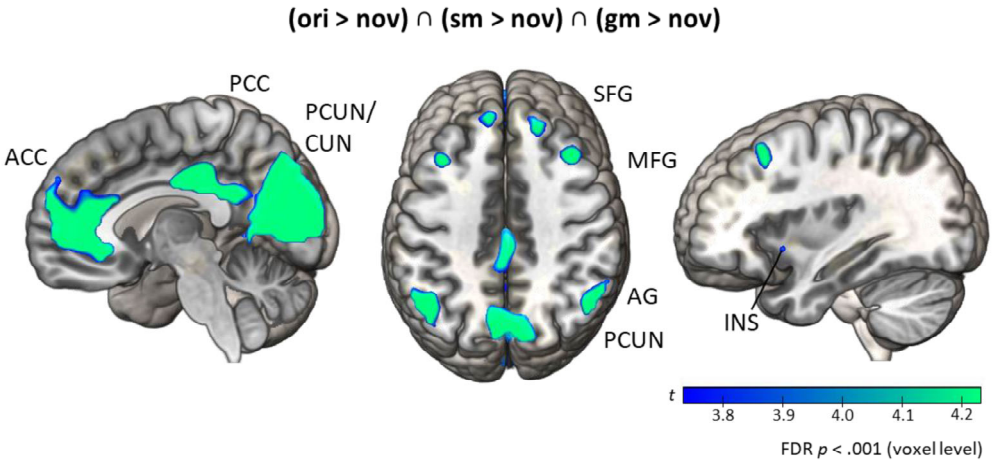


Figure 4. Whole-brain activation for episodic recall compared to novel videos. FDR-corrected ($p < .001$) *t*-map for conjunction of ori > nov, sm > nov and gm > nov contrasts. ACC = anterior cingulate cortex; PCC = posterior cingulate cortex; PCUN = precuneus; CUN = cuneus; SFG = superior frontal gyrus; MFG = middle frontal gyrus; AG = angular gyrus; INS = insula. Ori = original; nov = novel; sm = surface modification; gm = gist modification. For a smoother visualization of the conjunction, the *t*-map was resampled to a resolution of 1 mm³ voxels.

3.2.3 Region of interest analyses

To more specifically investigate which brain areas are in involved in the processing of sm and gm, we performed ROI analyses. We used functional ROIs (rIFS, OTC, IPS) of regions that responded to content modification, i.e., object exchanges, in our previous study (Siestrup et al., 2022), to test our hypotheses that (1) both types of modifications elicit activation in these areas and (2) that activation in these regions is stronger for gist than for surface modifications. Further ROIs were the anterior and posterior HPC to test whether gist and surface modifications differ regarding long-axis hippocampal activation. Mean contrast estimates were extracted from sm > ori and gm > ori contrasts.

In rIFS, there was significant activation for both, sm and gm in contrast to ori (sm: M = 0.403 ± 0.133, $t_{(36)}$ = 3.03, p = .005; gm: M = 0.829 ± 0.148, $t_{(36)}$ = 5.62, p < .001). A paired one-tailed t-test revealed that activation was greater for gm than for sm ($t_{(36)}$ = -3.27, p = .001). In the OTC ROI, gm showed significant activation (M = 0.383 ± 0.080, $t_{(36)}$ = 4.79, p < .001), while there was a non-significant trend for activation in sm (M = 0.147 ± 0.078, $t_{(36)}$ = 1.90, p = .066). Condition gm elicited significantly higher activation than sm ($t_{(36)}$ = -2.35, p = .012). Likewise, gm led to significant activation in the IPS ROI (M = 0.430 ± 0.132, $t_{(36)}$ = 3.25, p = .002), while this was not the case for sm (M = 0.138 ± 0.111, $t_{(36)}$ = 1.25, p = .219). The difference between the two modification types was significant ($t_{(36)}$ = -2.30, p = .014).

In anterior HPC, neither type of modification elicited significant activation (sm: M = 0.026 ± 0.033, $t_{(36)}$ = 0.78, p = .438; gm: M = 0.033 ± 0.044, $t_{(36)}$ = 0.75, p = 0.458) and they did not differ significantly ($t_{(36)}$ = -0.15, p = .442) (Figure 6A). In posterior HPC, we found a non-significant trend for activation for sm (M = 0.066 ± 0.036, $t_{(36)}$ = 1.84, p = .074), but not gm (M = -0.021 ± 0.041, $t_{(36)}$ = -0.51, p = 1). Activation was significantly stronger for sm than for gm ($t_{(36)}$ = 2.04, p = .025) (Figure 6B).

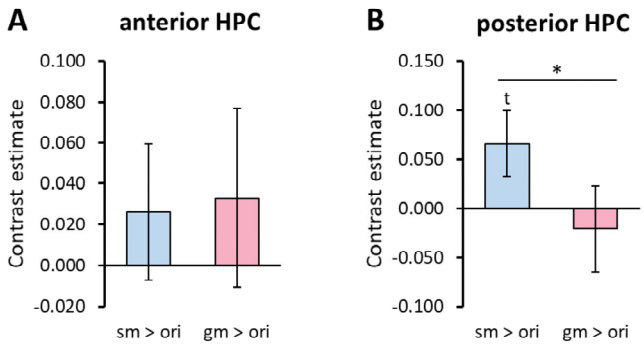


Figure 6. ROI analysis of bilateral anterior and posterior hippocampus (HPC). Contrast estimates were extracted from contrasts sm > ori and gm > ori. Statistics: One-sample t-tests and paired t-tests (one-tailed). t = p < .1, * = p < .05.

3.3 Post-fMRI memory test

In the post-fMRI memory test, participants were presented modified episodes from the fMRI session and additionally, their originally encoded counterparts. Episodes which had appeared in the original version during scanning were now presented in the original and, additionally, one modified version. After each video, participants had to rate how confident they were that they had originally encoded exactly this episode.

We calculated a non-parametric rmANOVA with the factors $\text{MODIFICATION}_{\text{FMRI}}$ (yes, no) and $\text{VERSION}_{\text{MT}}$ (sm, gm) to analyze AUC values as a measure of memory performance in the post-fMRI memory test. There was a significant effect of $\text{MODIFICATION}_{\text{FMRI}}$ ($F_{(1,36)} = 18.94$, $p < .001$, $\eta p^2 = .34$), as lower AUC values were achieved when participants had previously encountered modified episodes during the fMRI session ($M_{\text{yes}} = .906 \pm .010$; $M_{\text{no}} = .936 \pm .010$). Additionally, there was a significant effect of $\text{VERSION}_{\text{MT}}$ ($F_{(1,36)} = 239.97$, $p < .001$, $\eta p^2 = .87$), driven by higher AUC values for gm than for sm versions ($M_{\text{sm}} = .853 \pm .016$; $M_{\text{gm}} = .988 \pm .004$). Furthermore, the interaction between $\text{MODIFICATION}_{\text{FMRI}}$ and $\text{VERSION}_{\text{MT}}$ was significant ($F_{(1,36)} = 12.35$, $p = .001$, $\eta p^2 = .26$). Post hoc Wilcoxon tests revealed that AUC values were only reduced significantly by previous experience with modified versions for sm ($Z = -2.97$, $p = .006$) but not for gm ($Z = -1.00$, $p = .634$) (Figure 7A).

Notably, AUC scores obtained from the pilot study and the main study when only originals were considered from the fMRI session, i.e., when participants did not have any previous experience with modified episodes before the memory test, were highly similar (pilot study: $M_{\text{sm}} = .907 \pm .013$; $M_{\text{gm}} = .995 \pm .003$; main study: $M_{no\text{-sm}} = .882 \pm .019$; $M_{no\text{-gm}} = .990 \pm .006$).

To more specifically investigate the influence of modified episodic cueing on memory, we additionally analyzed the false alarm and hit alarm rates. Using a non-parametric rmANOVA, we found a significant effect of the factor $\text{MODIFICATION}_{\text{FMRI}}$ on false alarm rates

($F_{(1,36)}$ = 41.59, p < .001, ηp^2 = .54), which were higher when modified episodes had been experienced already during the fMRI session (M_{yes} = .188 ± .017; M_{no} = .118 ± .018). Additionally, there was a significant influence of VERSION_{MT} ($F_{(1,36)}$ = 473.63, p < .001, ηp^2 = .93), as participants had a higher tendency to falsely accept sm versions compared to gm versions as originals (M_{sm} = .291 ± .028; M_{gm} = .015 ± .006). Additionally, there was a significant interaction between both factors ($F_{(1,36)}$ = 18.16, p < .001, ηp^2 = .34). Post hoc Wilcoxon tests showed that there was only an increase in false alarms for sm (Z = -3.29, p = .003), but not for gm (Z = -1.47, p = .284) when participants had previously seen modified versions during scanning (Figure 7B). Hit rates were close to ceiling for all experimental conditions ($M_{yes\text{-}sm}$ = .929 ± .020; $M_{yes\text{-}gm}$ = .949 ± .014; $M_{no\text{-}sm}$ = .932 ± .023; $M_{no\text{-}gm}$ = .966 ± .014) and were neither significantly affected by the factor $\text{MODIFICATION}_{FMRI}$ ($F_{(1,36)}$ = 1.84, p = .184, ηp^2 = .05), VERSION_{MT} ($F_{(1,36)}$ = 0.02, p = .899, ηp^2 = .00) nor an interaction of both ($F_{(1,36)}$ = 0.23, p = .631, ηp^2 = .01) (Figure 7C).

For modified_{MT} videos, there was a significant effect of $\text{MODIFICATION}_{FMRI}$ on reaction times ($F_{(1,36)}$ = 4.43, p = .042, ηp^2 = .11), as participants responded faster when they had not experienced the modification during the fMRI session (M_{yes} = 471.383 ± 23.250 ms; M_{no} = 439.958 ± 20.392 ms). Additionally, there was a non-significant trend for an effect of VERSION_{MT} ($F_{(1,36)}$ = 3.57, p = .067, ηp^2 = .09), driven by the participants' tendency to respond faster for gm than for sm videos (M_{sm} = 475.620 ± 25.157 ms; M_{gm} = 435.721 ± 17.603 ms). There was no significant interaction ($F_{(1,36)}$ = 2.00, p = .166, ηp^2 = .05) (Figure 7D). Concerning reaction times for original_{MT} videos, we found no significant effect of the factor VERSION_{FMRI} ($F_{(1.59,57.23)}$ = 0.25, p = .727, ηp^2 = .01) (Figure 7E).

For novel videos, we found very low false alarm rates (M = .020 ± .017) and a mean reaction time of 527.041 ± 27.941 ms in the post-fMRI memory test.

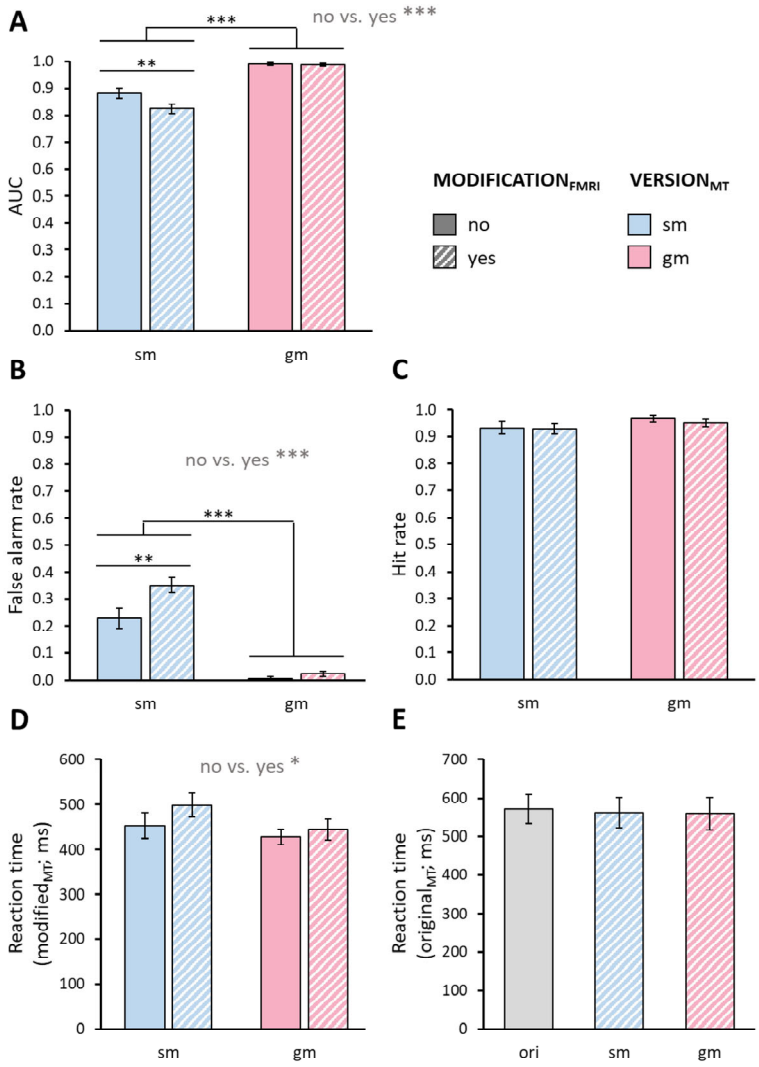


Figure 7. Behavioral results from post-fMRI memory test. Participants rated modified$_{MT}$ and original$_{MT}$ videos to decide whether they showed originally encoded episodes or not. (A) Area under the curve (AUC). Statistics: non-parametric rmANOVA and Wilcoxon signed rank tests (one-tailed). (B) False alarm rates for modified$_{MT}$ videos. Statistics: non-parametric rmANOVA and Wilcoxon signed rank tests (one-tailed). (C) Hit rates for original$_{MT}$ videos. Statistics: non-parametric rmANOVA. (D) Reaction times for modified$_{MT}$ videos. Statistics: rmANOVA. (E) Reaction times for original$_{MT}$ videos. Statistics: rmANOVA. Bar plots show means and standard errors. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

3.4 Rating task

In the rating task, participants had to rate how much the storyline of modified episodes deviated from the originally encoded version (from 1 = 0 % different to 6 = 100 % different). A non-parametric rmANOVA with the factors MODIFICATION$_{FMRI}$ (yes, no) and VERSION$_{MT}$ (sm, gm) confirmed a significant influence of VERSION$_{MT}$ on story-change ratings in the post-fMRI rating task ($F_{(1,36)} = 347.288$, $p < .001$, $\eta p^2 = .91$). As expected sm versions received lower ratings than gm versions ($M_{sm} = 1.363 \pm 0.058$; $M_{gm} = 4.897 \pm 0.093$), once again confirming our classification of modifications. Interestingly, there was also a significant effect of MODIFICATION$_{FMRI}$ ($F_{(1,36)} = 4.868$, $p = .034$, $\eta p^2 = .12$), as modified videos which had already been presented during the fMRI session received lower story-change ratings than those that had been presented in the original version before ($M_{yes} = 3.083 \pm 0.066$; $M_{no} = 3.177 \pm 0.069$). There was no significant interaction of both factors ($F_{(1,36)} = 1.791$, $p = .189$, $\eta p^2 = .05$).

A rmANOVA on reaction times in the rating task, i.e., how long it took participants to notice the change in modified videos, revealed a significant effect of VERSION$_{MT}$ ($F_{(1,36)} = 9.717$, $p = .004$, $\eta p^2 = .21$). Participants took longer to notice the modification in sm versions than in gm versions ($M_{sm} = 885.573 \pm 43.209$ ms; $M_{gm} = 798.305 \pm 32.695$ ms). Furthermore, there was a non-significant trend for an effect of MODIFICATION$_{FMRI}$ ($F_{(1,36)} = 3.937$, $p = .055$, $\eta p^2 = .10$), as participants tended to notice the modification sooner when they had already seen the same modified version during the fMRI session ($M_{yes} = 793.114 \pm 26.710$ ms; $M_{no} = 887.373 \pm 54.061$ ms). No significant interaction was found ($F_{(1,36)} = 0.876$, $p = .356$, $\eta p^2 = .02$).

4 Discussion

In the present study, we examined the impact of prediction errors on episodic memories. To this end, we used modified cues to induce prediction errors of different episodic relevance, altering the episodic gist or only surface properties. Surface modifications lead to significant activation in right IFS. In contrast, modifications that also changed the gist of the episode lead to overall more pronounced activation in the same area, and additional activity in OTC, IPS and pMTG. Posterior HPC responded to surface changes more strongly than to gist changes. Interestingly, there was a significant reduction of cingulate activity for gist modifications as compared to original or surface modified cueing. The day after the fMRI session, we found that a history with surface modified, but not with gist modified episodic cues increased the erroneous acceptance of modified stories as originals. Our findings demonstrate that the relevance of mnemonic PEs for the storyline of episodes is crucial for brain responses and influences on memory.

4.1 Neural responses to episodic modifications

We expected that both types of episodic modifications, characterized by the exchange of only one object, elicit brain activation in areas which we had previously identified for content-based PEs with intermediate episodic relevance (Siestrup et al., 2022). Indeed, gist modifications elicited brain activation in those hypothesized regions, namely OTC, including FG, posterior IPS and IFS. This was revealed by whole-brain and ROI analyses. For surface modifications, we found subthreshold whole-brain activation in right IFS, left FG and right posterior IPS. The ROI-analysis revealed significant activation in right IFS, as well as a trend for activation in OTC for surface modifications. As previously discussed (Siestrup et al., 2022), these brain areas are involved in the processing of object information in actions (El-Sourani et al., 2019; Lingnau & Downing, 2015; Wiggett & Downing, 2011; Reber et al., 2005; Grill-

Spector et al., 2001). Furthermore, lateral OTC, which was primarily activated by gist modifications, codes for conceptual action representations (Wurm & Caramazza, 2022). Activation in rIFS, OTC and IPS ROIs was significantly greater for gist than for surface modifications, as expected due to the gist modifications' comparably high relevance for episode content. In summary, we confirmed our general hypotheses concerning activation typical for PEs elicited by content modifications in an episodic cueing setting.

Additionally, we found subthreshold activation clusters for surface modifications in caudate nucleus, anterior insula and medial BA 8. These three areas, along with TPJ and IPS, were found to be co-activated when predictions about complex sequences of movements were violated (Schiffer & Schubotz, 2011). Even though the activations found in the present study did not remain significant after correction for multiple comparisons, the similarity to this previously identified network is striking. Caudate nucleus is known to respond to breaches of expectation (Delgado, Li, Schiller, & Phelps, 2008; Haruno & Kawato, 2006; Schiffer & Schubotz, 2011; Schiffer et al., 2012), while BA 8 and anterior insula activations were shown to reflect unexpectedness and uncertainty (Zaretsky, Mendelsohn, Mintz, & Hendler, 2010; Sarinopoulos et al., 2010; Schiffer, Krause, & Schubotz, 2013; Volz, Schubotz, & Von Cramon, 2003; Volz, Schubotz, & Von Cramon, 2005). In summary, the high degree of similarity between surface modified episodes and originals might have contributed to increased uncertainty, while gist modified episodes were more clearly distinguished from original ones.

As hypothesized, we found significant qualitative differences in brain activation between gist and surface modifications. For gist compared to surface modifications (and original episodes), there was increased activation in pSTS and pMTG. This finding is especially interesting as we have previously observed these areas to be activated specifically for structural episodic modifications, i.e., the exchange of adjacent action steps, but not for the exchange of an object with intermediate episodic relevance (Siestrup et al., 2022), which will be further

discussed below. Posterior temporal cortex, as part of the OTC (Wurm & Caramazza, 2022), is crucial for conceptual processing (Martin, 2007), and is thought to represent conceptual object properties (Fairhall & Caramazza, 2013). In the context of such concepts, especially left pMTG is suggested to be important for retrieval of semantic information and the comprehension and recognition of actions (Davey et al., 2016), together with other network components that were identified for gist modifications, namely STS, IPS and lateral OTC (James, VanDerKlok, Stevenson, & James, 2011). Davey et al. (2016) found that pMTG was functionally connected with IFG and IPS, and proposed that pMTG together with these regions is involved in adjusting semantic retrieval to different contexts. Interestingly, pMTG showed increased activation when goal-based expectations were violated (Jastorff, Clavagnier, Gergely, & Orban, 2011), which can also be transferred to prediction violations in the present study. Further, pMTG in conjunction with IFG has been suggested to have a role in the resolution of interfering episodic representations via semantic elaboration (Han, O'Connor, Eslick, & Dobbins, 2012). In that sense, pMTG could subserve the integration of semantic and episodic information, which might be particularly required in the case of gist modifications.

Furthermore, gist modifications yielded higher brain activation in precuneus compared to surface modifications. This area is involved in episodic memory retrieval, as well as the experience of self-perspective and sense of agency (Cavanna & Trimble, 2006). Increased activity for gist modifications here might reflect a strong dissociation between the self-experienced episode and the presented version of the episode. Interestingly, we previously identified this area for structure, but not for content modifications with intermediate episodic relevance (Siestrup et al., 2022). Aside from its role in the sequential organization of memories, precuneus was shown to be involved in their contextual reconstruction (Foudil, Kwok, & Macaluso, 2020). Therefore, activation in this area for gist modifications is a hint that the

exchange of a gist relevant object seems to be meaningful not only for episode content, but also structural and contextual (re)evaluations of that memory.

In summary, we saw that gist modifications elicited activation we identified to be specific for content, but also structure-based PEs in episodic memory in our previous work (Siestrup et al., 2022). Schiffer, Ahlheim et al. (2013) showed that adaptation of internal models following PEs occurred in exactly those areas that responded to gist modifications in the present study. In their experiment, episodes were modified by introducing a breach-point from which the action unfolded with new content *and* structural aspects. Thus, brain responses to Schiffer's gist modifications reflected more than mere object information processing, as the modified aspect was relevant for the integrity of an episode as a whole. Similarly, in the present study, integration of new episodic gist likely requires revision of object information *and* structure and context of events within an episode as the new storyline unfolds.

We had hypothesized to find increased ACC activation for gist modifications since it has been reported that more gist like memories depend strongly on ACC (e.g., Sekeres, Winocur, Moscovitch, et al., 2018). Surprisingly, we observed exactly the opposite: compared to original and surface modified episodes, pregenual ACC activation, as well as MCC activation, was reduced in gist modified episodes. Bonasia et al. (2018) found a similar response pattern in their study. Participants encoded episodes with a gist that was either congruent or incongruent with previous experiences. With increasing congruence, enhanced encoding-related activation in pregenual ACC, PCC, STG and AG was observed. This is consistent with our findings for gist modified episodes, and also for novel episodes we presented. For the latter, activation was decreased compared to previously experienced episodes in all four areas mentioned by the researchers. The authors argued that ACC and PCC were likely engaged due to their role during the activation of prior knowledge and semantic processing (Bonasia et al., 2018), thus linking new experience to prior ones. It has been suggested that ACC is responsible

for gist processing by providing a template that allows for prediction and interpretation of events (Sekeres, Winocur, & Moscovitch, 2018). In the case of gist modifications in the present study, the gist of the encoded episode was no longer intact, so predictions could potentially no longer rely on the established template. In addition, it is known that ACC is involved in conflict processing (Vassena, Holroyd, & Alexander, 2017; Botvinick, Cohen, & Carter, 2004). In gist modified episodes, conflict arising due to the mismatch with encoded episodes might be rapidly resolved, as the new gist defines a new episode. This also matches our behavioural findings in the post-fMRI memory test. Similarly, Webb et al., (2016) reported that retrieval of non-gist details was characterized by higher (dorsal) ACC activation than retrieval of gist details in a recognition memory task and argued that this finding was due to elevated cognitive efforts and conflict monitoring processes, as mediated by this brain area. In our previous study, pACC activation in response to episodic modifications was increased in the condition that was characterized by strong memory modification effects in the post-fMRI memory test and pACC activation increased over time for later false memories (Siestrup et al., 2023). In the present study, pACC activation was reduced in gist modified episodes, which did not trigger memory modification in the post-fMRI memory test. Therefore, we suggest that pACC might be involved in memory modification in response to mnemonic PEs. In summary, the relevance of PE within the episode is probably responsible for the impact on memory: when the overall storyline remains intact, a memory might be more likely to be (slightly) modified through PE, while a new, distinct episode is established when PE is gist relevant. This idea is similar to a computational approach of memory modification formulated by Gershman et al., (2017). They proposed that memories are modified when old and new information are inferred to share the same latent cause, while new memories are formed when sensory input is ascribed to a new latent cause.

Surface modifications elicited higher levels of activation in posterior HPC than gist modifications. This is consistent with the suggestion that posterior HPC mediates detailed memory representations while anterior HPC codes for gist (Poppenk, Evensmoen, Moscovitch, & Nadel, 2013; Moscovitch, Cabeza, Winocur, & Nadel, 2016; Sekeres, Winocur, Moscovitch, et al., 2018; Sekeres, Winocur, & Moscovitch, 2018). Accordingly, surface modifications likely elicited more elaboration of fine-grained details in reference to the original episode, as mediated by posterior HPC (Moscovitch et al., 2016; Robin & Moscovitch, 2017), while this was not required for gist modifications due to the clear contentual drift compared to the original. However, based on this account, gist modifications should have led to increased activation in anterior HPC, which we could neither confirm on the whole brain level nor on the basis of ROI analyses. According to a recently updated version of the Trace Transformation Theory (Sekeres, Winocur, & Moscovitch, 2018), gist like representations arise after some time as memories become less detailed and a representational shift from posterior to anterior HPC occurs. In the current study, we introduced surface and gist level modifications in a set of recently encoded, and thus, likely untransformed memories. The typical involvement of anterior HPC in gist memories might only arise after longer time intervals. Contrary to that assumption, Poppenk et al., (2008) found elevated anterior HPC activation when previously encoded sentences were modified in respect to their gist, but not when changing only verbatim detail. In their study, the modified episodes were presented on the same day the encoding took place, so without a temporal delay that would allow for memory transformation. Alternatively, some evidence suggests that especially anterior HPC is susceptible to repetition suppression, i.e., the decrease in activation due to a repeated presentation of the same stimulus (Reagh, Watabe, Ly, Murray, & Yassa, 2014). Therefore, repetition suppression in our study may have counteracted more subtle gist related effects in the anterior HPC.

4.2 Influence of episodic modifications on memory accuracy in post-fMRI memory test

We replicated our finding that the repeated experience of modified episodes during the fMRI session influences memory performance in a post-fMRI memory test. Importantly, in the current study, we found this effect only for surface, but not for gist modified episodes. Specifically, previous encounters with surface modified videos increased the participants' tendency to erroneously accept modified versions as truly encoded, as reflected by reduced AUC values and increased false alarm rates. Hit rates were not influenced and remained at a high level even after the repeated exposure to modified episodes, similar to what we showed before (Jainta et al., 2022; Siestrup et al., 2022; Siestrup et al., 2023). This finding is in line with reports that the inclusion of new information into an old memory is not necessarily accompanied by a deterioration of the original memory (St. Jacques, Olm, & Schacter, 2013; Sinclair & Barense, 2018). Together, the pattern of results in the post-fMRI memory test suggests that memory modification occurred due to the encoding of an additional memory trace which interfered with the original one, but not a replacement of original memory content (Siestrup et al., 2023).

We found the memory modification effect only for surface, but not for gist modifications, which was expected due to their differential relevance for the episodes. In line with that, participants were also faster to detect gist than surface modification in the rating task, i.e., when explicitly asked to rate how much the storyline of modified episodes deviated from the originally encoded version. It has previously been demonstrated that the experience of episode modifications can also improve memory, but only when such modifications are detected and remembered by participants (Wahlheim & Zacks, 2019). Furthermore, it is known that distinct items, i.e., those that violate the current context, are remembered better and that such distinctiveness reduced false memories (Hunt, 2013; Sommer & Sander, 2022). The fact

that gist modifications effectively constituted new episodes likely contributed to distinctiveness between original and modified versions, explaining the lack of confusion between them and, ultimately, the low rate of false alarms. This distinctiveness is probably also relevant for another finding from the post-fMRI memory test: memory performance was generally worse for episodes with a surface modification compared to those with a gist modification, irrespective of any previous experience with modified videos during scanning. This was evidenced by reduced AUC values and increased false alarm rates in the surface modification condition. Therefore, it was more difficult for participants to correctly differentiate modified from original versions when the modification was not gist relevant. This observation is in agreement with reports that non-central details are less likely to be remembered (Sekeres et al., 2016) and can be easily confused by misinformation (Reyna & Lloyd, 1997; Reyna et al., 2016).

Additionally, when modified episodes had been presented in the scanner, participants were slower to correctly judge the same as not part of the original episode repertoire in the post-fMRI memory test. Longer reaction times in memory tasks can be indicative of higher cognitive processing demands during retrieval (Larsen & Plunkett, 1987; Noppeney & Price, 2004). Potentially, it became more difficult for participants to correctly reject modified versions due to the encoding of alternative episode versions.

Overall, behavioral evidence we collected so far suggests that based on mnemonic PEs, alternative episode representations are encoded which then stand in conflict to the originally encoded ones. In the case of surface level PEs, these alternative representations interfered, leading to false memories, while in the case of gist level PEs, the two alternatives were clearly separated.

5 Conclusion

Our results indicate that the relevance of mnemonic PEs within episodes shapes neural responses and memory performance. Even though both gist and surface modifications of episodic cues were characterized by the exchange of only one object within an episode, their implications were highly specific: surface level PE increased uncertainty and conflict between alternative episode representations, whereas gist level PE induced the encoding of a new, distinct episode not conflicting with the original one. An important implication of our findings is that not all types of mnemonic PE influence memory in the same way, which needs to be considered in future research.

Data Availability Statement

The datasets generated for this study are available on request to the corresponding author.

Author Contributions

Sophie Siestrup: Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization

Ricarda I. Schubotz: Conceptualization, Methodology, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision, Funding acquisition

Acknowledgments

The authors thank Monika Mertens, Alea Bexten and Lana Steuernagel for their help during data collection. We thank Benjamin Jainta for helpful comments on an earlier draft of this manuscript and members of research unit FOR 2812 for valuable discussions.

Funding Information

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) – project number 419037023. The funder had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

Ethics Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Local Ethics Committee of the University of Münster. Participants signed an informed consent before participation.

References

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

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

Bonferroni, C. E. (1936). *Teoria Statistica Delle Classi e Calcolo Delle Probabilità, Vol. 8. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, (pp. 3–62).

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2022). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-022-02179-w

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164–169. https://doi.org/10.1111/1467-8721.00192

Brett, M., Anton, J. L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. Presented at the 8th International Conferance on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Available on CD-ROM in *Neuroimage*, Vol 16, No 2, Abstract 497. https://doi.org/10.1201/b14650-28

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*, 564–583. https://doi.org/10.1093/brain/awl004

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience and Biobehavioral Reviews*, *60*, 82–89. https://doi.org/10.1016/j.neubiorev.2015.11.011

Corbin, J. C., Reyna, V. F., Weldon, R. B., & Brainerd, C. J. (2015). How reasoning, judgment, and decision making are colored by gist-based intuition: A fuzzy-trace theory approach. *Journal of Applied Research in Memory and Cognition*, *4*, 344–355. https://doi.org/10.1016/j.jarmac.2015.09.001

Davey, J., Thompson, H. E., Hallam, G., Karapanagiotidis, T., Murphy, C., De Caso, I., et al. (2016). Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *Neuroimage*, *137*, 165–177. https://doi.org/10.1016/j.neuroimage.2016.05.051

Delgado, M. R., Li, J., Schiller, D., & Phelps, E. A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 3787–3800. https://doi.org/10.1098/rstb.2008.0161

El-Sourani, N., Trempler, I., Wurm, M. F., Fink, G. R., & Schubotz, R. I. (2019). Predictive

impact of contextual objects during action observation: Evidence from functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *32*, 326–337. https://doi.org/10.1162/jocn_a_01480

Epstein, R. A. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*, 954–978. https://doi.org/10.1080/13506280444000607

Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *Journal of Neuroscience*, *33*, 10552–10558. https://doi.org/10.1523/JNEUROSCI.0051-13.2013

Foudil, S. A., Kwok, S. C., & Macaluso, E. (2020). Context-dependent coding of temporal distance between cinematic events in the human precuneus. *Journal of Neuroscience*, *40*, 2129–2138. https://doi.org/10.1523/JNEUROSCI.2296-19.2020

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P, Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210. https://doi.org/10.1002/hbm.460020402

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Gershman, S. J., Monfils, M. H., Norman, K. A., & Niv, Y. (2017). The computational nature of memory modification. *ELife*, *6*, e23763. https://doi.org/10.7554/eLife.23763

Greene, N. R., & Naveh-Benjamin, M. (2021). The effects of divided attention at encoding on item and associative memory. *Memory and Cognition*, *31*, 1021–1035. https://doi.org/10.3758/BF03196123

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*, 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Guo, P., Li, Q., Wang, X., Li, X., Wang, S., Xie, Y., et al. (2020). Structural covariance changes of anterior and posterior hippocampus during musical training in young adults. *Frontiers in Neuroanatomy*, *14*, 20. https://doi.org/10.3389/fnana.2020.00020

Han, S., O'Connor, A. R., Eslick, A. N., & Dobbins, I. G. (2012). The role of left ventrolateral prefrontal cortex during episodic decisions: Semantic elaboration or resolution of episodic interference? *Journal of Cognitive Neuroscience*, *24*, 223–234. https://doi.org/10.1162/jocn_a_00133

Haruno, M., & Kawato, M. (2006). Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *Journal of Neurophysiology*, *95*, 948–959. https://doi.org/10.1152/jn.00382.2005

Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science*, *22*, 10–15. https://doi.org/10.1177/0963721412463228

Jainta, B., Siestrup, S., El-Sourani, N., Trempler, I., Wurm, M. F., Werning, M., et al. (2022). Seeing what i did (not): Cerebral and behavioral effects of agency and perspective on

episodic memory re-activation. *Frontiers in Behavioral Neuroscience*, *15*, 793115. https://doi.org/10.3389/fnbeh.2021.793115

James, T. W., VanDerKlok, R. M., Stevenson, R. A., & James, K. H. (2011). Multisensory perception of action in posterior temporal and parietal cortices. *Neuropsychologia*, *49*, 108–114. https://doi.org/10.1016/j.neuropsychologia.2010.10.030

Jastorff, J., Clavagnier, S., Gergely, G., & Orban, G. A. (2011). Neural mechanisms of understanding rational actions: Middle temporal gyrus activation by contextual violation. *Cerebral Cortex*, *21*, 318–329. https://doi.org/10.1093/cercor/bhq098

Larsen, S. F., & Plunkett, K. (1987). Remembering experienced and reported events. *Applied Cognitive Psychology*, *1*, 15–26. https://doi.org/10.1002/acp.2350010104

Lee, J. L. C., Nader, K., & Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in Cognitive Sciences*, *21*, 531–545. https://doi.org/10.1016/j.tics.2017.04.006

Li, X., Li, Q., Wang, X., Li, D., & Li, S. (2018). Differential age-related changes in structural covariance networks of human anterior and posterior hippocampus. *Frontiers in Physiology*, *9*, 518. https://doi.org/10.3389/fphys.2018.00518

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, *19*, 268–277. https://doi.org/10.1016/j.tics.2015.03.006

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*, *19*, 1233–1239. https://doi.org/10.1016/S1053-8119(03)00169-1

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annual Review of Psychology*, *67*, 105–134. https://doi.org/10.1146/annurev-psych-113011-143733

Nader, K. (2015). Reconsolidation and the dynamic nature of memory. *Cold Spring Harbor Perspectives in Biology*, *7*, a021782. https://doi.org/10.1101/cshperspect.a021782

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

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*, 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005

Noppeney, U., & Price, C. J. (2004). Retrieval of abstract semantics. *Neuroimage*, *22*, 164–170. https://doi.org/10.1016/j.neuroimage.2003.12.010

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36. https://doi.org/10.1016/S0079-6123(06)55002-2

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization

of the human hippocampus. *Trends in Cognitive Sciences*, *17*, 230–240. https://doi.org/10.1016/j.tics.2013.03.005

Poppenk, J., Walia, G., McIntosh, A. R., Joanisse, M. F., Klein, D., & Köhler, S. (2008). Why is the meaning of a sentence better remembered than its form? An fMRI study on the role of novelty-encoding processes. *Hippocampus*, *18*, 909–918. https://doi.org/10.1002/hipo.20453

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org.

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2005). Priming effects in the fusiform gyrus: Changes in neural activity beyond the second presentation. *Cerebral Cortex*, *15*, 787–795. https://doi.org/10.1093/cercor/bhh179

Reagh, Z. M., Watabe, J., Ly, M., Murray, E., & Yassa, M. A. (2014). Dissociated signals in human dentate gyrus and CA3 predict different facets of recognition memory. *Journal of Neuroscience*, *34*, 13301–13313. https://doi.org/10.1523/JNEUROSCI.2779-14.2014

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, *5*, 1–9. https://doi.org/10.1016/j.jarmac.2015.12.003

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, *9*, 95–123. https://doi.org/10.1016/S1041-6080(97)90002-9

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences*, *17*, 114–123. https://doi.org/10.1016/j.cobeha.2017.07.016

Sarinopoulos, I., Grupe, D. W., Mackiewicz, K. L., Herrington, J. D., Lor, M., Steege, E. E., & Nitschke, J. B. (2010). Uncertainty during anticipation modulates neural responses to aversion in human insula and amygdala. *Cerebral Cortex*, *20*, 929–940. https://doi.org/10.1093/cercor/bhp155

Schiffer, A. M., Ahlheim, C., Ulrichs, K., & Schubotz, R. I. (2013). Neural changes when actions change: Adaptation of strong and weak expectations. *Human Brain Mapping*, *34*, 1713–1727. https://doi.org/10.1002/hbm.22023

Schiffer, A. M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS One*, *7*, e36445. https://doi.org/10.1371/journal.pone.0036445

Schiffer, A. M., Krause, K. H., & Schubotz, R. I. (2013). Surprisingly correct: Unexpectedness of observed actions activates the medial prefrontal cortex. *Human Brain Mapping*, *35*, 1615–1629. https://doi.org/10.1002/hbm.22277

Schiffer, A. M., & Schubotz, R. I. (2011). Caudate nucleus signals for breaches of expectation in a movement observation paradigm. *Frontiers in Human Neuroscience*, *5*, 38. https://doi.org/10.3389/fnhum.2011.00038

Schubotz, R. I. (2015). Prediction and expectation. In A. W. Toga (Ed.), *Brain Mapping: An Encyclopedic Reference* (Vol. 3, pp. 295–302). Academic Press: Elsevier. https://doi.org/10.1016/B978-0-12-397025-1.00205-0

Schubotz, R. I., & Von Cramon, D. Y. (2009). The case of pretense: Observing actions and

inferring goals. *Journal of Cognitive Neuroscience*, *21*, 642–653. https://doi.org/10.1162/jocn.2009.21049

Scully, I. D., Napper, L. E., & Hupbach, A. (2017). Does reactivation trigger episodic memory change? A meta-analysis. *Neurobiology of Learning and Memory*, *142*, 99–107. https://doi.org/10.1016/j.nlm.2016.12.012

Sekeres, M. J., Bonasia, K., St-Laurent, M., Pishdadian, S., Winocur, G., Grady, C., & Moscovitch, M. (2016). Recovering and preventing loss of detailed memory: Differential rates of forgetting for detail types in episodic memory. *Learning and Memory*, *23*, 72–82. https://doi.org/10.1101/lm.039057.115

Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters*, *680*, 39–53. https://doi.org/10.1016/j.neulet.2018.05.006

Sekeres, M. J., Winocur, G., Moscovitch, M., Anderson, J. A. E., Pishdadian, S., Martin Wojtowicz, J., et al. (2018). Changes in patterns of neural activity underlie a time-dependent transformation of memory in rats and humans. *Hippocampus*, *28*, 745–764. https://doi.org/10.1002/hipo.23009

Siestrup, S., Jainta, B., Cheng, S., & Schubotz, R. I. (2023). Solidity meets surprise: Cerebral and behavioral effects of learning from episodic prediction errors. *Journal of Cognitive Neuroscience*, *35*, 291–313. https://doi.org/10.1162/jocn_a_01948

Siestrup, S., Jainta, B., El-Sourani, N., Trempler, I., Wurm, M. F., Wolf, O. T., et al. (2022). What happened when? Cerebral processing of modified structure and content in episodic cueing. *Journal of Cognitive Neuroscience*, *34*, 1287–1305. https://doi.org/10.1162/JOCN_A_01862

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error triggers episodic memory updating. *Learning & Memory*, *25*, 369–381. https://doi.org/10.1101/lm.046912.117.25

Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neurosciences*, *42*, 727–739. https://doi.org/10.1016/j.tins.2019.08.007

Sommer, V. R., & Sander, M. C. (2022). Contributions of representational distinctiveness and stability to memory performance and age differences. *Aging, Neuropsychology, and Cognition*, *29*, 443–462. https://doi.org/10.1080/13825585.2021.2019184

St. Jacques, P. L., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-Induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 19671–19678. https://doi.org/10.1073/pnas.1319630110

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25. 10.1146/annurev.psych.53.100901.135114

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al.. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289. https://doi.org/10.1006/nimg.2001.0978

Vassena, E., Holroyd, C. B., & Alexander, W. H. (2017). Computational models of anterior

cingulate cortex: At the crossroads between prediction and effort. *Frontiers in Neuroscience*, *11*, 316. https://doi.org/10.3389/fnins.2017.00316

Volz, K. G., Schubotz, R. I., & Von Cramon, D. Y. (2003). Predicting events of varying probability: Uncertainty investigated by fMRI. *Neuroimage*, *19*, 271–280. https://doi.org/10.1016/S1053-8119(03)00122-8

Volz, K. G., Schubotz, R. I., & Von Cramon, D. Y. (2005). Variants of uncertainty in decision-making and their neural correlates. *Brain Research Bulletin*, *67*, 403–412. https://doi.org/10.1016/j.brainresbull.2005.06.011

Wahlheim, C. N., & Zacks, J. M. (2019). Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General*, *148*, 30–50. https://doi.org/10.1037/xge0000458

Webb, C. E., Turney, I. C., & Dennis, N. A. (2016). What's the gist? The influence of schemas on the neural correlates underlying true and false memories. *Neuropsychologia*, *93*, 61–75. https://doi.org/10.1016/j.neuropsychologia.2016.09.023

Wiggett, A. J., & Downing, P. E. (2011). Representation of action in occipito-temporal cortex. *Journal of Cognitive Neuroscience*, *23*, 1765–1780. https://doi.org/10.1162/jocn.2010.21552

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, *17*, 766–780. https://doi.org/10.1017/S1355617711000683

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings ofthe SIGCHI Conference on Human Factors in Computing Systems* (pp. 143–146). New York: Association for Computing Machinery. https://doi.org/10.1145/1978942.1978963

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revesited - Again. *Neuroimage*, *2*, 173–181. https://doi.org/10.1006/nimg.1995.1023

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends in Cognitive Sciences*, *26*, 103–116. https://doi.org/10.1016/j.tics.2021.10.003

Zaretsky, M., Mendelsohn, A., Mintz, M., & Hendler, T. (2010). In the eye of the beholder: Internally driven uncertainty of danger recruits the amygdala and dorsomedial prefrontal cortex. *Journal of Cognitive Neuroscience*, *22*, 2263–2275. https://doi.org/10.1162/jocn.2009.21402

4 General Discussion and Future Directions

4.1 Summary of the Presented Studies

Previous research has shown that memories are not exact copies of past experiences, but can be changed on the basis of new relevant information. Such modifications are likely triggered by mnemonic PEs, i.e., when there is a mismatch between predictions derived from memories and actual situational input. The aim of the presented studies was to investigate the neural underpinnings of mnemonic PEs in the context of episodic memories and their influence on memory. To this end, a naturalistic episodic modification paradigm was developed and brain responses to modified episodic cues were investigated using fMRI.

In **Study 1** (Siestrup et al., 2022), the focus was on characterizing brain responses for two different types of mnemonic PEs, either based on episode structure or content information, and to assess their consequences for episodic memories. In two initial encoding sessions, participants encoded short toy stories from demo videos by re-enacting the stories themselves. During two further sessions which were conducted on separate days, participants performed a retrieval task in order to further consolidate memories. On the last day of the study, participants went through an fMRI session during which original or slightly modified episode videos were presented. Modifications either affected the structure of episodes (change of two adjacent action steps) or their content (change of an object). Afterwards, participants completed a post-fMRI memory test during which memory for originally encoded episodes was probed. Content and structure modification commonly activated superior parietal areas as well as right ventrolateral prefrontal cortex (vlPFC). Compared to structure modifications, content modifications elicited elevated brain activity in posterior parietal, temporo-occipital and parahippocampal areas. Structure modifications led to enhanced neural responses in right dorsal premotor, posterior temporal and medial parietal regions. After the pre-exposure with modified episodes, i.e., the

experience of PEs, participants showed an increased tendency to accept the same modified versions as originally encoded in the post-fMRI memory test. In the case of structure modifications, participants' tendency to reject original episodes increased as well. However, overall, hit rates for originally encoded episodes were near ceiling level. In summary, it was successfully demonstrated that different types of mnemonic PEs, based on structure or content information, are (1) processed differently by the brain and (2) influence subsequent memories.

Study 2 (Siestrup, Jainta, Cheng, & Schubotz, 2023) aimed to investigate the impact of consolidation through retrieval practice on neural processing of episodic memories, especially in the case of PE. Importantly, **Study 2** was based on the same data as **Study 1**. After encoding, participants went through two retrieval sessions. Each episode was either retrieved two or eight times and in a spaced or massed fashion, i.e., during both retrieval sessions or only during one. FMRI and post-fMRI memory test were conducted as described above. The re-analysis of the behavioral data, this time using rating scores instead of hit rates and false alarm rates, revealed the same pattern as found in **Study 1**, namely increased acceptance of modified episodes after repeated experience of PE during the fMRI session and decreased acceptance of original videos (in the presence of an overall ceiling effect). Memory performance was better for episodes previously retrieved eight compared to two times, with and without pre-experience with modified episodes in the scanner. The spacing of retrieval practice did not affect memory performance, but brain activation during episodic cueing was increased in superior parietal lobe (SPL), PCC and precuneus after spaced compared to massed retrieval practice. In the case of modified episodes, two previous retrieval opportunities lead to more brain activation in pre- and subgenual ACC than eight. Furthermore, later false alarms in the memory test were characterized by increasing brain activation over the course of the fMRI experiment in several brain regions, including ACC and hippocampal complex. Taken together, **Study 2** demonstrated that (1) different memory consolidation approaches in the form of retrieval

practice protocols differently influence neural processing and retention of episodic memories and (2) identified a network of brain regions which dynamically determine the formation of false memories in response to mnemonic PEs.

Lastly, **Study 3** (Siestrup & Schubotz, 2023) aimed to shed more light on the differential processing of mnemonic PEs, this time characterized by their episodic relevance. Two types of content modifications were used: some challenged the storyline of episodes (so-called gist modifications) while others did not (so-called surface modifications). On three consecutive days, participants went through encoding, fMRI and post-fMRI memory test. As expected, surface as well as gist modifications triggered activation in brain areas which responded to content modifications in **Study 1**. This activation was generally stronger for gist modifications, which, additionally, also activated posterior temporal regions and precuneus. Increased activation in posterior hippocampus for surface compared to gist modifications confirmed the suggested role of this region for detail-processing in episodic memory. Interestingly, gist modified episodes were characterized by reduced activation in pregenual ACC. In the post-fMRI memory test, previous experience with surface modified, but not gist modified episodes increased the participants' tendency to erroneously endorse the same modified episodes as originally encoded. In summary, **Study 3** demonstrated that mnemonic PEs of different episodic relevance (1) are processed differently by the brain and (2) differentially influence subsequent memory.

4.2 Mnemonic Prediction Errors in the Brain

4.2.1 Neural Responses to Mnemonic Prediction Errors

PEs are not only surprise signals, but carry representational content (den Ouden et al., 2012). That is, PEs elicit enhanced neuronal processing (den Ouden et al., 2012; Sayood, 2018), leading to measurably increased brain activation in relevant brain areas (Bubic et al., 2009; Gläscher et al., 2010). Accordingly, it can be expected that in the context of episodic memory,

different types of mnemonic PEs are represented in distinct brain areas. To our knowledge, we were the first to systematically compare the neural processing of different types of mnemonic PEs in episodic memory. Importantly, encoding and retrieval of episodic memories are usually intertwined. For example, new encoding is associated with retrieval due to the embedding of new information in existing knowledge structures, and retrieval is accompanied by re-encoding of recalled information and/or the constructive embedding of current contextual details in a memory representation (Finn, 2017; Scully & Hupbach, 2020; Xue, 2022). For this reason, neural activation we detected in response to episodic modifications cannot be attributed to only one of the two processes. Rather, brain responses share features of retrieval, which is required for making top-down predictions, and new encoding of informative, i.e., unpredicted, input.

Common Neural Responses

Several brain areas have been suggested to subserve the processing of PEs in general, for example striatal, (medial) prefrontal, and parietal regions (D'Astolfo & Rief, 2017; Kastner, Kube, Villringer, & Neumann, 2017; Schiffer et al., 2012; Schiffer & Schubotz, 2011; Wang et al., 2017). Additionally, similar brain structures are deemed crucial for episodic encoding and retrieval, like hippocampus, and prefrontal and parietal cortices (Rugg et al., 2015). For this reason, we expected that some brain regions would commonly respond to mnemonic PEs in our paradigm.

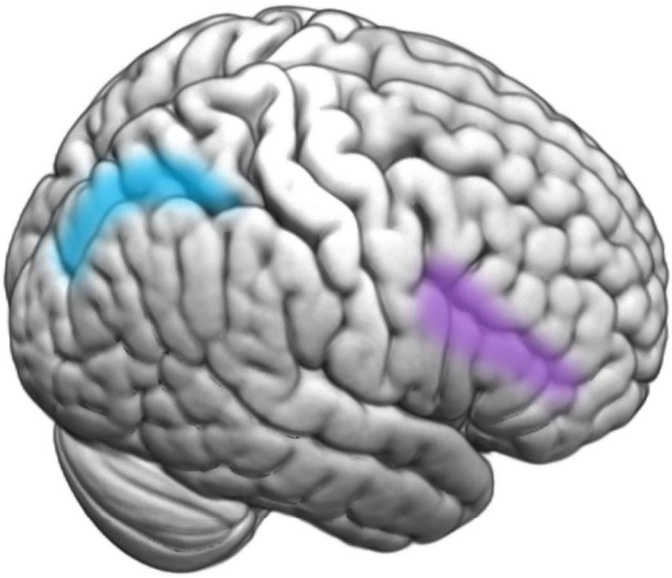
In fact, we found that vlPFC, more specifically a region comprising inferior frontal sulcus (IFS)/inferior frontal gyrus (IFG) (BA 44 and 45), generally responded to mnemonic PEs, predominantly in the right hemisphere, in **Studies 1** and **3** (and in an unpublished analysis of data from Jainta et al., 2022). In line with this, previous research has demonstrated that lateral PFC responds to violated predictions (Bubic et al., 2009; Corlett et al., 2004; Rushworth, Noonan, Boorman, Walton, & Behrens, 2011; Schiffer, Krause, & Schubotz, 2013; Schiffer &

Schubotz, 2011; Schultz & Dickinson, 2000) and activation in right lateral PFC has even been deemed a "reliable signature for the presence of prediction error" (Corlett et al., 2004, p. 877). Our results also relate to the assumption that lateral PFC in general is involved in the establishment of predictions derived from long-term memories (den Ouden et al., 2012; Schubotz, 2015) and the finding that vlPFC activity increases for highly informative or inconsistent detail in observed actions (El-Sourani, Trempler, Wurm, Fink, & Schubotz, 2019; Hrkać, Wurm, Kühn, & Schubotz, 2015; Wurm & Schubotz, 2012). Furthermore, vlPFC, especially the pars triangularis of IFG (BA 45), subserves the selection between competing memories (Kuhl, Bainbridge, & Chun, 2012; Martin, 2007; Schlichting, Mumford, & Preston, 2015), which is accomplished by promoting the encoding of relevant information (Kuhl et al., 2012), and top-down (post)retrieval processing including the selective access of semantic memory (Martin, 2007; Wimber, Alink, Charest, Kriegeskorte, & Anderson, 2015). On this basis, lateral PFC is involved in updating of mnemonic associations (Kluen, Dandolo, Jocham, & Schwabe, 2019; Long et al., 2016) and "goal-directed memory changes" (Xue, 2022, p. 8), suggesting that this region might be involved in regulatory processes of memory modification after PE. While the above mentioned roles are sometimes reported for bilateral IFG (Schlichting et al., 2015), they are more often demonstrated for the left hemisphere (Kluen et al., 2019; Kuhl et al., 2012; Martin, 2007). In contrast, we report predominantly right activation in this area in response to PEs. This might reflect a dominance for retrieval compared to encoding-related processes in our studies, as right and left PFC have been linked to these roles, respectively (Finn, 2017; Hayes, Ryan, Schnyer, & Nadel, 2004). See also Figure 1 for a graphic summary of results.

Additionally, we found posterior intraparietal sulcus (pIPS) activation in response to all different types of modification in our studies (**Studies 1** and **3,** and in an unpublished analysis of data from Jainta et al., 2022). This brain area has been linked to complex cognitive processes,

like reasoning, memory, and action observation (Richter et al., 2019). For example, superior parietal cortex is involved in the formation of episodic memories (Rugg et al., 2015; Uncapher & Wagner, 2009) and their strategic retrieval, including recall of source and item information. The contribution of this brain area to these tasks has been ascribed to its role in attention (Ciaramelli, Grady, & Moscovitch, 2008). Namely, SPL is involved in the reorientation of attention to unexpected or otherwise salient stimuli (Corbetta & Shulman, 2002; Molenberghs, Mesulam, Peeters, & Vandenberghe, 2007; Tamber-Rosenau, Esterman, Chiu, & Yantis, 2011) and regulates top-down attention in mnemonic processes. This is important when postretrieval monitoring is necessary to differentiate truly memorized content from such content that is not part of memories (Cabeza et al., 2011; Ciaramelli et al., 2008). Taken together, our findings from **Study 3** further supported the interpretation of pIPS activation we reported in **Study 1**. Namely, we suggest that superior parietal regions are involved in the processing of PEs in episodic memory, possibly by guiding updating processes (Siestrup et al., 2022; see also Figure 1 for a graphic summary).

In summary, vlPFC and pIPS likely have a general role in the processing of episodic mismatch and subsequent mnemonic selection processes.



Ventrolateral PFC (IFS/IFG)

Relevant functional roles:

- Establishment of predictions and signaling of prediction errors
- Retrieval and post-retrieval processing
- Resolving conflict between competing memories and updating of mnemonic associations

pIPS/SPL

Relevant functional roles:

- Encoding and retrieval of episodic memories
- Reorientation of attention to unexpected stimuli
- Strategic evaluation of memory content

Bubic et al., 2009; Corlett et al., 2004; den Ouden et al., 2012; El-Sourani et al., 2019; Finn, 2017; Hayes et al., 2004; Hrkać et al., 2015; Kluen et al., 2019; Kuhl et al., 2012; Long et al., 2016; Martin, 2007; Rushworth et al., 2011; Schiffer, Krause, et al., 2013; Schiffer & Schubotz, 2011; Schlichting et al., 2015; Schubotz, 2015; Schultz & Dickinson, 2000; Wimber et al., 2015; Wurm & Schubotz, 2012; Cabeza et al., 2011; Ciaramelli et al., 2008; Corbetta & Shulman, 2002; Molenberghs et al., 2007; Rugg et al., 2015; Tamber-Rosenau et al., 2011; Uncapher & Wagner, 2009

Figure 1. Schematic graphic summary of brain areas which commonly responded to mnemonic prediction errors. Ventrolateral prefrontal cortex (PFC), specifically inferior frontal sulcus (IFS)/inferior frontal gyrus (IFG), and posterior intraparietal sulcus (pIPS)/superior parietal lobe (SPL) responded to mnemonic prediction errors in **Studies 1** and **3**. Proposed relevant functions of the two areas within the framework of prediction error processing and, potentially, memory modification, are summarized on the right.

Differential Neural Signatures

In **Study 1**, we were able to show that structure and content modifications elicit differential neural processing[6]. Specifically, structure modifications yielded increased activation in right premotor regions, posterior temporal and parietal cortex, and precuneus, which we interpreted as a network operating in the updating of predictive models due to an unexpected new structure in episodes (Siestrup et al., 2022). In an active serial prediction task, Bubic and colleagues (2009) also reported activation in premotor cortex for sequence violations

[6] Please note that the findings reported here for structure and content modifications were overall replicated in an unpublished analysis of the data reported in Jainta et al., 2022.

and likewise concluded that these brain areas are involved in the establishment of predictive models of sequences. Interestingly, their findings mirror the right hemisphere dominance we also observed, and the authors relate this to the need for increased information retrieval and high processing load (Bubic et al., 2009).

Content modifications elicited activation in parietal, parahippocampal and temporo-occipital regions, which we related to the processing of new object information within episodes (Siestrup et al., 2022). Interestingly, Gläscher et al. (2010), who investigated prediction violations in the transition between probabilistic states implemented via fractal images, found a neural response pattern that is highly comparable to the one we showed for content related PEs. Similarly, serial PEs seem to share some neuronal underpinnings in different tasks (Bubic et al., 2009). These commonalities show that to some extent, our findings can be extended from the context of episodic memory to more general PE responses, thus suggesting a ubiquitous role of memory in general in the generation of predictive models.

In **Study 3**, we were able to extent our findings of differential neural processing to PEs of different episodic, and thus predictive, relevance. For both modification types, we found brain activation in areas that responded to content modifications in **Study 1**. In general, brain responses were weaker for surface modifications than gist modifications. Exceptions were posterior hippocampus, midcingulate cortex and pregenual ACC, where surface modifications elicited more brain activation.

The finding that content-specific PE responses were not as pronounced for surface modifications than for gist modifications was expected, as it can be assumed that more relevant pieces of information within an episode are processed more strongly. Reasons for this may either be initial stronger encoding of the original episode or increased processing at retrieval, which we cannot clearly differentiate with our paradigm. In terms of predictive relevance, it makes sense that gist modifications are a more meaningful learning signal, and thus trigger a

stronger PE signal. In fact, subjective PE ratings (on a scale from 1 [least unexpected] to 6 [most unexpected]) were significantly higher for gist than for surface modifications ($M_{surface}$ = 2.742 ± 0.174, M_{gist} = 4.307 ± 0.146, $F_{(1,36)}$ = 49.08, p < .001, ηp^2 = 0.58, unpublished data). However, it should be noted that PE ratings and story-change ratings were highly correlated (r = 0.58, p < .001, unpublished data), which might either hint at a natural confound or methodological problems, as discussed below. Interestingly, it has been shown previously that brain activation attenuated in exactly the regions which responded to gist modifications when direct predictability via first-order information was given (Ahlheim, Schiffer, & Schubotz, 2016), suggesting that gist modifications disrupt high predictability, as expected.

Notably, (pregenual) ACC activation was specifically reduced for gist modified episodes, which shows that it cannot ubiquitously be assumed that stronger PE signals yield stronger brain responses. As discussed below, PEs through gist modifications did not increase false alarm responses in the memory test. Therefore, it could be that reduced ACC activation prevented the formation of false memories, potentially, as suggested by us, via the rapid resolvation of mnemonic conflict (Siestrup & Schubotz, 2023).

When relating the findings from **Study 1** to those of **Study 3**, it is striking that several areas which exclusively responded to structure modifications in **Study 1** also responded to gist modifications in **Study 3** (posterior superior temporal sulcus/middle temporal gyrus, precuneus). This is interesting, since gist modifications were also content-based modifications, i.e., operationalized through the exchange of a single object. However, in **Study 1**, content modifications were designed to be of intermediate relevance for the episode, i.e., they should have some impact but the episode as a whole should still remain the same. An explorative *post-hoc* rating of content modifications (from **Study 1**) with a small independent sample of participants (n = 6) revealed that story changes were indeed of intermediate impact compared to surface and gist modifications (mean story-change ratings, on a scale from 1 to 6: $M_{surface}$ =

1.36, $M_{content}$ = 2.37, M_{gist} = 4.90, unpublished data). This observation further supports our conclusion from **Study 3** that gist modifications not only require cognitive re-evaluation of memory content, but also structural, contextual and semantic aspects due the complex impact on the storyline (Siestrup & Schubotz, 2023), highlighting the fact that predictions rely on the interaction of different memory systems (Henson & Gagnepain, 2010).

In summary, we successfully answered our first research question how different types of episodic PEs are processed by the brain. In line with our hypotheses, we demonstrated that some brain regions, vlPFC and pIPS, responded to all types of PEs investigated, while other brain activations were specifically connected to the type and informativeness of modified details. However, we also showed that different types of PEs can recruit the same brain regions, like structure and gist (i.e., content) modifications in **Study 1** and **3**, respectively. Thus, while qualitative differences in neuronal processing of PE types definitely exist, more work is needed to decisively determine factors that further influence their neural signatures.

4.2.2 The Role of Hippocampus in Mnemonic Prediction Error Processing

The hippocampal complex is a crucial brain region for episodic memory (Rugg et al., 2015), as well as learning (Suzuki, 2007). This area is not only involved in the formation and retrieval of true but also false memories (Dennis, Bowman, & Turney, 2015). Furthermore, the hippocampus was found to be a key region in the neural processing of mnemonic PEs, earning it the reputation of a mismatch detector (Duncan et al., 2012; Kumaran & Maguire, 2007).

Surprisingly, our findings concerning hippocampal complex were rather mixed over the different studies. While in **Study 1**, content modifications elicited elevated activation in parahippocampal gyrus (PHG), **Study 2** revealed that in the spaced modification subconditions, there was a mean deactivation compared to original episodes within a hippocampal region of interest (ROI). Further, in **Study 3**, only surface modifications yielded a meaningful increase in (posterior) hippocampal activation, while a mean decrease was found in this ROI for gist

modifications. Additional explorative ROI analyses in hippocampus proper revealed that a (descriptive) average deactivation for modified versus original episodes was actually common in our studies (e.g., for the aggregated mod > ori contrast from **Study 1**: M = -0.058 ± 0.031; unpublished data). Notably, others could also not demonstrate PE signals in hippocampus proper (D'Astolfo & Rief, 2017; Sinclair et al., 2021), specifically not in passive paradigms, i.e., when participants did not have to make active predictions (D'Astolfo & Rief, 2017). Accordingly, active predictions might promote the establishment of detectable mnemonic mismatch in hippocampus (e.g., Long et al., 2016; but see also Kumaran & Maguire, 2007). Similar to our findings, Giovanello, Schnyer, and Verfaellie, 2004 reported reduced hippocampal activation for mismatched vs. matched associations, which they suggested might reflect an increased focus on reinstatement of old associations. Indeed, the fact that hippocampus is crucial for predictive processing, but also encoding and retrieval of episodic memories (Barron, Auksztulewicz, & Friston, 2020) might help to understand our findings better. Modified episodes were contrasted with original episodes to isolate neural PE signals. However, hippocampus was likely heavily engaged for original episodes as well, due to ongoing episodic retrieval, contextual re-encoding (Finn, 2017; Scully & Hupbach, 2020; Xue, 2022) and online comparison of predictions and sensory input, which can complicate the identification of PE responses (but see e.g., Kumaran & Maguire, 2006, for a similar baseline condition). Another possibility is that a stronger deviation between new information and the context in which it appears would lead to increased hippocampal activation (Thakral, Yu, & Rugg, 2015), which is suggested by our finding that novel episodes compared to original ones elicited a hippocampal response (Jainta et al., 2022).

One interesting observation from **Study 1** is that PHG signaled for content PEs. In fact, others have also demonstrated the involvement of this region in PE processing (D'Astolfo & Rief, 2017; Schiffer, Ahlheim, et al., 2013). Additionally, PHG is involved in item memory

(Asperholm, Högman, Rafi, & Herlitz, 2019), which might have been especially relevant in our paradigm, as episodes were effectively comprised of sequences of different toy items. Thus, depending on the nature of episodic memories under investigation, not only hippocampus proper might serve as a mismatch detector, but this function might be expanded to other structures of hippocampal complex, like PHG.

In summary, our findings do not support the proposed ubiquitous role of hippocampus as a detector of mnemonic PEs. Even though this was not a specific research question of this thesis, this is a highly insightful finding.

4.2.3 The Influence of Consolidation

One factor that might influence brain responses to mnemonic PEs is memory solidity (Exton-McGuinness et al., 2015; R. S. Fernández et al., 2016; Schiffer, Ahlheim, et al., 2013). Aside from sleep, retrieval practice can contribute to memory consolidation (Antony et al., 2017), a phenomenon termed testing effect (Rowland, 2014). However, depending on which theory of systems consolidation is considered, predictions for the neural processing of consolidated memories vary strikingly. For example, according to the standard theory of consolidation, hippocampal involvement would decrease for more consolidated memories, while neocortical activation, especially in mPFC, would increase. In contrast, other theories like multiple trace theory or contextual binding theory would predict increased hippocampal activation with ongoing consolidation, due to the constant re-encoding of memory traces (Bosshardt et al., 2005; Euston et al., 2012; Yonelinas, Ranganath, Ekstrom, & Wiltgen, 2019). Furthermore, while there are first findings indicating that internal model solidity, or, in other words, memory solidity influences the neural processing of PEs (Schiffer, Ahlheim, et al., 2013), systematic investigations are still lacking.

In **Study 2**, we therefore systematically varied consolidation strategies by applying different retrieval practice protocols. Participants either went through different amounts of

additional retrievals (two vs. eight) and retrieval practice was carried out during only one or two sessions (massed vs. spaced). Note that while the beneficial effect of more retrieval opportunities on memory solidity seems to be rather ubiquitous (e.g., Gerbier & Toppino, 2015), the benefit of spaced over massed retrieval is reported more inconsistently and might be more dependent on specific features of the paradigm (Dempster, 1989).

For the general neural processing of episodes, we only found an effect of spaced vs. massed retrieval, which was characterized by elevated activation in areas involved in episodic memory (Siestrup et al., 2023), as previously reported by others (Li & Yang, 2020; Takashima et al., 2007). Unfortunately, due to the lack of a behavioral effect, it is difficult to interpret these findings in reference to memory solidity. Based on other researchers' conclusions, one might carefully assume that increased brain activation in the spaced conditions represents more stable memories (Ezzyat, Inhoff, & Davachi, 2018; Li & Yang, 2020; Takashima et al., 2007) and to show the behavioral effect as well, our experimental paradigm might require careful revision (c.f. Gerbier & Toppino, 2015).

A more straightforward observation was made for the effect of fewer vs. more retrieval opportunities in our study. Here, we found that for modified (versus original) episodes, activation in ACC was increased for less frequently retrieved vs. more frequently retrieved episodes. From the behavioral results, we could additionally conclude that memories which had been retrieved more often were more solid. This observation is in line with our suggestion that (pregenual) ACC activation might be related to false memory formation after PE. However, our findings contrast previous reports of increased brain activation for more consolidated memories (Bosshardt et al., 2005; Euston et al., 2012; Takashima et al., 2006 [specific for pregenual ACC]; Wiklund-Hörnqvist, Stillesjö, Andersson, Jonsson, & Nyberg, 2021), specifically in the

case of PE (Schiffer, Ahlheim, et al., 2013)[7]. In this context, it is worth noting that in many studies, consolidation is operationalized through a time lag between encoding and final test (several days or weeks), as it is assumed that much sleep equals much consolidation (Bosshardt et al., 2005; Takashima et al., 2006). Importantly, memory performance has been observed to decrease over longer time intervals (Bosshardt et al., 2005; Li & Yang, 2020; Tompary & Davachi, 2017; Zhan, Guo, Chen, & Yang, 2018), so that it is questionable whether observed changes in neural processing can be truly attributed to increased consolidation. Rather, an alternative explanation might be more appropriate. Accordingly, Bosshardt and colleagues (2005) proposed that increased neural activation (e.g., in ACC) after long time lags represents elevated processing demands during attempted retrieval. This is in line with our interpretation from **Study 2**, namely that less stable memories might be more effortful to retrieve, giving rise to more neural activation for less frequently retrieved memories. In relation to this, we suggested that PEs might serve as a stronger learning signal in the case of weaker mnemonic representations, i.e., the predictive model might be more easily influenced (Schiffer, Ahlheim, et al., 2013; Siestrup et al., 2023). Unfortunately, it was not possible to directly validate this hypothesis through the differentiation of later false alarms and correct rejections in **Study 2,** potentially due to the heavily decreased power in the statistical model, resulting from the exclusion of several participants from this specific analysis (n = 22).

As a last note, our findings are not in line with the standard theory of consolidation, which predicts an increase in mPFC/ACC activation, as this region is believed to take over the role of hippocampus in memory retrieval after successful consolidation (Euston et al., 2012). However, since we did not find any significant influence of retrieval frequency on hippocampus

[7] Please note, however, that in the study by Schiffer, Ahlheim, et al., (2013), there were only 15 minutes between (differently strong) initial encoding and the final exposure during fMRI scanning. Thus, the time interval was likely not sufficient to achieve consolidation. Consequently, their results are more likely to reflect immediate effects of model solidity via encoding strength, which might not be equal to model solidity from consolidation of long-term memories.

and outcomes for spaced vs. massed retrieval were mixed as well, our findings do not allow concrete conclusions about which theory of consolidation is most suitable in light of our findings, and in general.

Taken together, **Study 2** delivered first evidence for answering our second research question, namely how different memory consolidation approaches influence neural processing (in the case of PE). The results indicate that brain activation increased for less solid memories, potentially due to increased processing demands. However, consolidation of episodic memories is a complex and multifaceted topic which can be operationalized in a variety of ways. Thus, more research is necessary to complete our understanding of this issue.

4.3 The Role of Mnemonic Prediction Errors in Memory Modification

4.3.1 Behavioral Evidence

It has previously been demonstrated that in some cases, mnemonic PEs can lead to the modifications of previously formed memories (e.g., R. S. Fernández et al., 2016), and the episodic memory system is known to be especially vulnerable for such interferences (Martin-Ordas & Call, 2013). While sometimes, no influence of PEs on subsequent memory can be detected (e.g., Hermann, Wahlheim, Alexander, & Zacks, 2021), others have replicably demonstrated that PEs elicit episodic memory modification (Sinclair & Barense, 2018; Sinclair et al., 2021). Similarly, we were able to demonstrate in all of our studies that repeated encounters with modified episodes selectively increased subsequent false acceptance of modified episodes as originally encoded (**Studies 1**, **2** and **3**; see also Jainta et al., 2022 for another replication of these findings with the same paradigm). Notably, the correct recognition of originally encoded episodes largely remained intact.

There is still an ongoing discussion in the scientific community on what exactly happens when a memory is modified. It has been suggested that old mnemonic content is replaced or overwritten entirely by new information (Elsey et al., 2018; Richards & Frankland, 2017; Xue,

2022), while sometimes the original memory is only weakened (e.g., Kim et al., 2014). However, the mere inability to recall an original memory does not necessarily imply that it has been erased. Rather, it might have become unavailable under certain contextual conditions as the result of memory modification (Elsey et al., 2018; Tronson & Taylor, 2007). Others proposed that memory modification is characterized by the integration of new information without a deterioration of old memory content (Bryant, Nadel, & Gómez, 2019; Elsey et al., 2018; Sinclair & Barense, 2018; Sinclair et al., 2021). The latter suggestion seems to be more in line with the adaptive idea of memory modification, because also old memory content can, under certain circumstances, still be of predictive relevance (Bein et al., 2021). In support of this, our studies demonstrated that after PE, participants tended to endorse both, original but also modified episodes as originally encoded. Thus, it is likely that two alternatives of the same episode were generated and likewise supported by the internal predictive model. Accordingly, Schiffer, Ahlheim, et al. (2013) showed that when evidence for different alternatives is available, predictive models represent such balanced states. Thus, when even more evidence for the new episode alternative accumulates, i.e., when modified episodes are presented more often, a more striking decrease in original episode endorsement might be observed in our paradigm as well. Furthermore, in all of our studies, we observed the tendency that participants took longer to correctly differentiate original and modified episodes in the memory test when they had experienced PEs in the preceding fMRI session. This finding can be interpreted as further evidence for competing memory alternatives (Anderson, 1983; Kuhl et al., 2012).

Furthermore, we also showed that different types of mnemonic PEs influence subsequent memory differently. For example, only structure modifications in **Study 1** yielded a significant reduction in the correct recognition of original episodes. Gist modifications in **Study 3** were the only ones for which we could not show the memory modification effect. Furthermore, **Study 2** revealed that additional consolidation can, to a certain extent, protect

memories from the influence of PEs. In line with these observations, it is known that certain factors influence how readily a memory can be changed. This includes the strength of the original memory or predictive model (Exton-McGuinness et al., 2015; R. S. Fernández et al., 2016; Schiffer, Ahlheim, et al., 2013) and the details represented by it (Reichardt et al., 2020), but also the strength of new learning or PE (Milton et al., 2023; Wichert, Wolf, & Schwabe, 2013). Accordingly, it has been proposed that when the mismatch between prediction and reality is moderate, memories undergo modification. In contrast, when the mismatch is striking, a new mnemonic representation is formed that is clearly separable from the previous one (Gershman, Monfils, Norman, & Niv, 2017; Gershman, Radulescu, Norman, & Niv, 2014; Milton et al., 2023). Meaningful mismatches might be more salient, allowing us to re-orient our attention to changed features, which, in turn, can facilitate the formation of clearly separate memories. For example, it has been demonstrated that subjective awareness of change in episodic association can reduce memory modification (Wahlheim, Smith, & Delaney, 2019; Wahlheim & Zacks, 2019).

Our finding that surface, but not gist modifications induced memory modifications can be nicely integrated into this framework. Gist modifications were, by definition, highly impactful. In contrast to surface modifications, they presumably triggered a strong PE signal, as evidenced by pronounced brain responses. These brain responses involved elevated pIPS involvement, which, as discussed above, might be attributed to attentional re-orientation (Corbetta & Shulman, 2002; Molenberghs et al., 2007; Tamber-Rosenau et al., 2011). Additionally, elevated baseline attention for gist details might have increased the neural PE signal (den Ouden et al., 2012; Schubotz, 2015; Stefanics et al., 2014). Thus, the participants' attention might have been drawn to differences between original and modified episodes, decreasing the chance of false memory formation. The finding that more consolidated memories were better protected from interference in our study further adds to this idea. Notably, our

findings do not imply that in case of gist modifications or especially solid memories, internal models were not adapted. For example, it was previously demonstrated that even when memories remain intact, predictions are updated (Schiffer, Ahlheim, et al., 2013). Thus, in cases when PEs are especially impactful (due to the nature of the violated information or the previous memory strength), it might be more beneficial for internal models to clearly distinguish between different alternatives, as they might selectively inform predictions in highly specific contexts. In contrast, when the mismatch is less impactful, the internal model might integrate different alternatives, as the best strategy for future predictions might be a more flexible reliance on prior knowledge.

As a last note, it is not entirely clear why especially structure modifications influenced the correct recognition of original episodes. It could be that sequence information is more readily updated than other types of episodic details. Accordingly, Yazin, Das, Banerjee, and Roy (2021) recently demonstrated that structural PEs strikingly weakened old mnemonic associations while at the same time promoting new ones.

Taken together, our highly replicable findings provide further evidence for two of our research questions. Namely, we demonstrated that mnemonic PEs can induce modification in episodic memory, which is further modulated by (1) PE type and (2) memory solidity.

4.3.2 Neural Evidence

Static Brain Responses

So far, I presented neural responses to mnemonic PEs and argued that they likely play a role in the integration of new information into predictive models. But how can we be sure that brain activation in response to episodic modifications can actually be linked to the formation of false memories? Several studies have investigated brain activation during the encoding of later false memories. For example, increased activity in MTL, visual areas and ventromedial PFC

during the encoding of misinformation was shown to be greater for later false compared to later true memories (Gonsalves et al., 2004; Okado & Stark, 2005; St. Jacques, Olm, & Schacter, 2013). In contrast, increased activation in ACC, PCC, precuneus and cuneus during a misinformation phase protected original memories from interference (Baym & Gonsalves, 2010).

To test for these effects in our data, we conducted additional analyses that related the behavioral findings from the post-fMRI memory test to brain responses during PE processing. Unfortunately, directly contrasting brain responses to later false alarms (fa; i.e., false memories) with those of later correct rejections (cr; i.e., true memories) did not yield significant results after correction for multiple comparisons (**Study 1**[8]). When inspecting the mentioned contrast (fa > cr) at subthreshold level ($p < 0.001$, uncorrected; unpublished data from **Studies 1** and **2**), two interesting clusters were found in right lingual gyrus/calcarine sulcus and bilateral cuneus (Appendix, Table A1). These two areas were recently reported to be activated by low visual predictability in complex action sequences due to increased reliance on exploratory visual gain. The authors interpreted this activation as reflecting the updating of predictive models in upstream areas (Pomp et al., 2021). Relating this to our findings, it is possible that when unpredicted new information is processed in a rather exploratory mode, this might lead to a form of model updating that fosters the formation of false memories.

Interestingly, brain activation for correct rejections versus false alarms (unpublished data from **Studies 1** and **2,** see Appendix, Table A1) largely mirrored the activation pattern we identified for modification responses in general (pIPS, IFS/IFG [BA 44 and 45], fusiform gyrus [FG]). There was also activation in bilateral anterior insula and left PHG, which are known to be involved in mismatch and error processing (Brázdil et al., 2002; Klein, Ullsperger, & Danielmeier, 2013; Schiffer, Krause, et al., 2013). This finding can be taken as a hint that

[8] The analysis was only conducted with **Study 1/2** data, as in **Study 3**, there were only few false alarms in total.

stronger PE responses in the brain (with specific focus on mismatch/error processing) can avoid the formation of false memories due to PEs. This assumption is also supported by our finding that gist modifications elicited stronger neural responses than surface modifications, but only surface modifications lead to the formation of false memories in the post-fMRI memory test. Notably, as previously suggested (section 4.2.2), PHG rather than hippocampus proper seemed to be involved in mismatch processing in our studies. Furthermore, our findings lend support to the suggested role of IFG (specifically BA 45) in resolving competition between different mnemonic alternatives through which they can be kept separate (Schlichting et al., 2015). Thus, the above suggested involvement of vlPFC in regulatory processes of memory modification after PE (section 4.2.1) can be narrowed down even further. In that sense, strong activation in this area might reduce the susceptibility to memory modification. A similar role can be assumed for pIPS. Interestingly, our findings from both contrasts (fa > cr, cr > fa) are not in line with the aforementioned observations by Baym and Gonsalves (2010), indicating that brain responses that lead to later false or true memories are specific for the type of misinformation paradigm.

Dynamic Brain Responses

Next, we were also interested in the dynamic change in brain responses, i.e., their increase over time, in the case of later false memories. This approach might seem counterintuitive at first, since often when the same stimulus is encountered repeatedly, brain activations are observed to decrease. This phenomenon is called repetition suppression, and is assumed to reflect more efficient processing of repeated stimuli (Henson, Shallice, & Dolan, 2000; Martin, 2007). For example, hippocampal activation often decreases with ongoing learning repetitions of the same material (Brodt et al., 2016; Himmer et al., 2019). Similarly, it was demonstrated that activation in PE-sensitive brain areas, including hippocampal complex, attenuates with repeated experience of the same PE, which can be interpreted as a sign of

predictive model adaptation (Schiffer, Ahlheim, et al., 2013; Schiffer et al., 2012). However, under certain circumstances, brain responses can also increase with repeated exposure, which has been termed a repetition enhancement effect (Henson et al., 2000). For example, when new information is encountered repeatedly, such enhancements have been observed in FG, inferior parietal lobe and precuneus. It was suggested that these increases in brain activation reflect the gradual establishment of new mnemonic representations (Brodt et al., 2016; Henson et al., 2000; Himmer et al., 2019). For this reason, we expected that in the case of later false memories, brain responses increase with repeated exposure to PE, specifically in areas involved in the formation of episodic memories. In fact, we demonstrated that many areas of the episodic memory network increased in activity for later false alarms, including inferior parietal areas and precuneus (**Study 2**), in line with previous findings. Also, hippocampal complex (including PHG) and ACC were among the regions where activity increases were detected. This finding contrasts reports of decreasing activity with stimulus repetition in both areas (Himmer et al., 2019; Kuhl, Dudukovic, Kahn, & Wagner, 2007). However, the amount of hippocampal activation is typically linked to current encoding strength (Brodt et al., 2016), and ACC is involved in updating reactivated memories (Xue, 2022), especially in the case of overlapping information (G. Fernández, 2017), and learning from (prediction) errors (Rushworth et al., 2011; Vassena, Holroyd, & Alexander, 2017). Considering the previously discussed findings from our other studies, it can be proposed that these two areas might play a key role in the ongoing formation of PE-based false memories. In that sense, they might mediate the relearning of conflicting, yet similar, alternative episodes. Notably, pronounced activity increases were specifically observed for later false alarms, and only to a reduced extent for later correct rejections and original episodes (unpublished data, see Appendix, Table A2).

Interestingly, no significant repetition suppression could be observed for later false alarms (unpublished data). This finding indicates that when false memories were formed, the

PE-inducing information remained informative even after several repetitions of the same modification. For later correct rejections, activity decreases were only found in some clusters without correction for multiple comparisons ($p < .001$). These were located in bilateral FG and left PHG (unpublished data, see Appendix, Table A3). This finding was further supported by the ROI analysis, which evidenced a significant decrease in PHG. The strongest decrease in activation was found for original videos, i.e., those that were just repeated the way they had been initially encoded, in bilateral occipitotemporal cortex (including FG [unpublished data, see Appendix, Table A3]). These observations nicely fit the report by Henson and colleagues (2000), who showed that repetition suppression in FG occurred only for familiar stimuli, but not for unfamiliar ones, as new mnemonic representations still needed to be formed. However, our results are inconsistent with Schiffer, Ahlheim, and colleagues' (2013) observations of decreasing brain activation with repeated exposure to the same PE in a very similar paradigm. Notably, there are some important differences between experiments which could account for this discrepancy. One key characteristic of our studies was the high personal involvement during initial encoding, while participants in Schiffer, Ahlheim, and colleagues' (2013) study only watched videos. Further, our experimental paradigm was stretched out over several days, allowing for consolidation through sleep, and additional consolidation was achieved via active retrieval practice. In contrast, Schiffer, Ahlheim, et al. (2013) did not allow for additional consolidation, as participants were transferred to the MRI scanner directly after first encoding. All of these aspects could have led to a higher initial model solidity in our studies, so that more counter evidence might be required to adapt the model to a stage where PEs become less informative and thus a decrease in brain responses can be observed.

In summary, we gathered valuable first evidence regarding our research question which (dynamic) brain activation in response to mnemonic PEs characterize false memory formation. Specifically, we found hints that specific brain responses during PE processing can predict later

false or true memories. The results indicate that especially pronounced neural mismatch signals can prevent the formation of false memories after PE. Furthermore, we were the first to demonstrate that during the formation of false memories through PE, brain activity specifically increases in the episodic memory network as the episode is relearned. Together, neural and behavioral evidence suggests that in case of later false alarms, a new, competing alternative of the original episode was formed.

4.3.3 Mechanisms of Modification

Our experiments did not aim to specifically differentiate by which underlying mechanism (e.g., reconsolidation vs. interference of memory traces) memories are modified in response to PEs (c.f. Elsey et al., 2018). However, given the current controversial discussion in literature, I will briefly discuss our findings in this context.

Today, a very fashionable explanation of memory modification is reconsolidation (but see Scully & Hupbach, 2020). Briefly, within this framework, PE serves as a reminder that reactivates an old memory, making it labile again so that new information can be incorporated. According to this view, the original memory is permanently changed. For reconsolidation to occur, new protein synthesis is necessary, which is why the process can only be completed after a longer time interval (Elsey et al., 2018; R. S. Fernández et al., 2016). Several studies demonstrating memory modification effects in humans have attributed these findings to reconsolidation (e.g., Hupbach et al., 2007; Hupbach et al., 2009; Sinclair & Barense, 2018).

In case of our experiments, several required criteria to exclusively demonstrate reconsolidation were not met (Elsey et al., 2018). For example, in **Studies 1** and **2**, the post-fMRI memory test was conducted immediately after participants exited the scanner. This time interval would be too short for reconsolidation to be measured (Elsey et al., 2018). Nevertheless, we did observe a memory modification effect, which consequently, most likely cannot be attributed to reconsolidation. When the post-fMRI memory test was postponed to the

subsequent day in **Study 3**, we replicated our results, indicating that behavioral effects in our studies might overall not be based on reconsolidation (alone). Another observation from our studies that is not in line with the reconsolidation framework is that the original memory was not lost after reactivation. On the contrary, memory for originally encoded episodes was hardly affected by PEs.

As discussed above, our behavioral as well as neural findings hint at the establishment of alternative, competing memory representations through PEs. Consequently, in line with previous findings and theoretical considerations, our results might be explained by the interference of memory traces rather than reconsolidation (Alberini & Ledoux, 2013; Brewin, 2015; Gisquet-Verrier & Riccio, 2018; Kiley & Parks, 2022; Klingmüller et al., 2017; Lee, 2010; Sederberg et al., 2011; Yassa & Reagh, 2013), which is consistent with the suggestion that PEs trigger the formation of new memory contents instead of the replacement of old ones (Bein et al., 2021). Notably, we found hints for a long-lasting effect of PEs on episodic memory, as predicted by reconsolidation theory (Elsey et al., 2018; Tronson & Taylor, 2007), since explorative repetition of the post-fMRI memory test approximately one week after the fMRI session still revealed the tendency for a memory modification effect (i.e., increased acceptance of modified episodes after PE; $t_{(31)} = 1.40$, $p = .085$). However, such an effect can also be explained by a permanent interference of memory traces, as for example predicted by the temporal context model (Sederberg et al., 2011) or other frameworks that suggest fixed neural co-activation of associated memories (e.g., Otgaar et al., 2017).

Importantly, while our memory modification effects can likely not be attributed to reconsolidation (alone), this does not imply that reconsolidation is not relevant in the modification of human episodic memory. Instead, there are likely several complementary processes, including interference effects, which influence episodic memory after PE (Elsey et al., 2018; Frankland, Josselyn, & Köhler, 2019; Moscovitch & Gilboa, 2021; Sinclair &

Barense, 2018). However, what remains unresolved to date is which memory modifying mechanisms operate under which conditions, how exactly they might interact and how memory traces are influenced in detail (Elsey et al., 2018; Greve, Abdulrahman, & Henson, 2018; Scully & Hupbach, 2020).

4.4 Critical Evaluation and Methodological Considerations

Even though our findings are overall highly insightful, some aspects of our work require critical evaluation. These include the design of the paradigm, data analyses and interpretations. First, I will discuss the reliability of basic assumptions on which we based our work and interpretations. Afterwards, I will reflect on specific methodological decisions and potential problems (and proposed solutions) in more detail.

4.4.1 Basic Assumptions and Interpretations

The focus of the present work was on mnemonic PEs. In our studies, participants never needed to make active predictions, so it could be questioned whether PEs were actually present. From a predictive coding perspective, it is clear that predictions and PEs do not need to be experienced consciously (Lupyan & Clark, 2015; Schubotz, 2015) to influence our behavior. Additionally, while it is true that violating active predictions might further boost neural PE signals (D'Astolfo & Rief, 2017) and behavioral influences (Brod et al., 2018), passive PEs can likewise elicit characteristic neural responses (D'Astolfo & Rief, 2017; Kumaran & Maguire, 2006) and lead to memory modification (Sinclair & Barense, 2018; Sinclair et al., 2021; Yazin et al., 2021). Furthermore, it could be argued that episodic modifications in the present studies merely elicited novelty, but not PE responses. While it is true that modifications conveyed new information, it must be considered that this was delivered within a familiar context, i.e., the previously encoded episode. Thus, according to the predictive coding framework, predictions were violated, giving rise to PEs (Barto et al., 2013; Reichardt et al., 2020).

From the arguments presented above, it becomes clear that we can assume that PEs were induced through modified episodes. However, it must also be considered whether aside from PEs, there are alternative explanations for our findings. For example, how do we know that neural responses were actually specific responses to PEs, and not unrelated encoding of new information? In fact, we presented multiple times that brain responses to familiar episodes, including modified ones, were qualitatively different from mere novelty responses (**Study 1**, **Study 3**, Jainta et al., 2022). The similarity of brain activation patterns for original and modified episodes in contrast to novel ones highlights the fact that modified details were processed in reference to originally encoded episodes. Concerning behavioral findings from the post-fMRI memory test, it needs to be considered whether modification effects could be attributed to familiarity or recency effects rather than PE. While this possibility cannot be ruled out with absolute certainty, it does not seem very likely since novel episodes were only rarely mistaken for originally encoded episodes in the post-fMRI memory test. Like modified episodes, they had been presented during the fMRI session for the first time and thus were encoded equally recently and were comparably familiar as modified episode versions.

Furthermore, we argue that through PEs, different alternatives of the same episode were encoded which then likely interfered with each other. In the case when no memory modification occurred, namely in the gist modification condition, we can, strictly speaking, not verify whether participants actually encoded the second alternative as well, since we did not specifically ask them to recall it. Thus, it is theoretically possible that they immediately forgot about the gist modified alternative, which would explain the lack of confusion between alternatives in the memory test. However, given the powerful role of PEs in new learning (Reichardt et al., 2020), this explanation seems highly unlikely. Further, many of our findings suggest that gist modified episodes were encoded aside from original alternative. First of all, brain responses to gist modified episodes were pronounced in areas that are involved in the

encoding of object information (Siestrup et al., 2022). Additionally, reaction times in the post-fMRI memory test increased after the experience of gist modifications, indicating mnemonic conflict (Anderson, 1983; Kuhl et al., 2012). Lastly, participants showed a tendency to detect gist modification faster in the rating task after pre-exposure in the scanner, suggesting that modifications were already familiar.

Taken together, it can be concluded that the general operationalization of the presented studies and the main interpretations are overall reliable and can be well justified with reference to the predictive coding framework.

4.4.2 Detailed Methodological Decisions

A major aim of the presented work was to develop a naturalistic paradigm to investigate the influence of PEs on episodic memory modification. To this end, complex toy stories served as episodes and participants were highly personally involved in encoding them. This is already a vast improvement compared to previous work, where sometimes only still images or simple (sequences of) associations were encoded, merely by passively presenting the material to participants (e.g., Greve et al., 2017; Kim et al., 2014). However, several aspects of the paradigm still lack naturalism. For example, to ensure good encoding, episodes were repeated multiple times and participants often needed more than one attempt to imitate and/or describe the story. However, one key characteristic of episodic memory is that it is usually based on one single experience (Cheng & Werning, 2016). Furthermore, modified versions were also presented repeatedly, while it has been suggested that learning from PE might only need a single exposition (Greve et al., 2017). In summary, our paradigm could be further improved to address memory modification under natural conditions even better.

Additionally, some features of the post-fMRI memory test might need to be adjusted for future application. First, it is theoretically possible that the observed decrease in memory performance after viewing modified videos was caused not by PEs, but the lack of additional

presentations of original episodes as in the comparative condition. To exclude this possibility, we would have needed to include a set of episodes that is encoded, *not* shown during the fMRI session, and appears again during the memory test. Due to the time constraints in the encoding session(s), this was not possible. However, in the gist modification conditions, memory was not negatively influenced by the presentation of modified versions, that is, no additional presentation of originals was necessary to protect memories. Furthermore, memory performances in the behavioral pilot experiment and the fMRI experiment of **Study 3** were highly similar when only conditions without PE experience were considered. Thus, it is unlikely that our behavioral effects were caused by a boost in memory performance through additional presentations of original episodes, but rather, as we argue, by the influence of PEs.

Additionally, the operationalization of the memory test limits the degree to which we can understand what actually happened to the memories which underwent modification. Participants were presented original and modified episodes and had to answer after each video whether this was the episode they had originally encoded. In case both versions were accepted, we could not differentiate whether there was a dominant mnemonic representation, which could rather be achieved with a forced-choice design (e.g., Long et al., 2016), or a cued recall test (e.g., Sinclair et al., 2021). Nevertheless, a clear advantage of our approach was that with the question we asked, it was possible to demonstrate that participants actually attributed modified episodes to the original episode repertoire. Still, one might argue that this finding only demonstrates simple source confusion. However, novel videos were rarely attributed to the wrong source (i.e., the encoding session), which highlights that for wrong source attribution to occur, different alternatives of the same episodes likely became associated with each other, which is an indication of memory modification (Otgaar et al., 2017; Sederberg et al., 2011). An additional advantage of our memory test was that it allowed us to show that memory for the original episode was still largely intact. Many other researchers neglect the fate of original

memories, which is problematic for understanding the mechanisms behind memory modification (Elsey et al., 2018).

For an explorative investigation of long-term effects, the memory test was repeated after approximately one week (**Studies 1** and **2**, unpublished data). While we found a tendency for a long-lasting effect as discussed above, these results have an important limitation. This is due to the fact that the memory test was executed in exactly the same way twice, including the exact same stimuli and even the same order of video presentation. Therefore, participants had experienced modified versions of *all* episodes at the time of the second test. If PE signals actually triggered one-shot learning (Greve et al., 2017), all episodes would have been influenced to a certain degree at this point, making the results of the second memory test unreliable. To reliably test for a long-term effect, it would be necessary to either split the group of participants (i.e., half of the participants take part in the early memory test, the other half in the late test), or divide the set of episodes over two testing sessions (e.g., Bosshardt et al., 2005).

For future work, it should also be considered that a subjective evaluation of PE strength, as applied in **Study 3**, might not be the optimal approach. Even though others have previously used such individual PE ratings (e.g., Sinclair et al., 2021), participants seemed confused by the task in our experiment. For this reason, they might have used the story-change ratings to guide PE ratings, leading to a high correlation of the two. With a rating approach, it is also not possible to reflect on information-theoretical surprisal, just subjective surprise (Clark, 2013), which might not always adequately reflect the true informational value of PEs. Due to these weaknesses, PE ratings were not further included into any analyses in the present work. Thus, obtaining objective measures of PE strength, for example using pupillometry (Brod et al., 2018), would be a meaningful improvement of our paradigm to further elucidate the suggested influence of PE strength on memory modification (Milton et al., 2023).

4.5 Outlook

With our here presented work, we have already provided valuable new insights into the neural processing and behavioral consequences of mnemonic PEs. However, much more research will be needed to complete the understanding of this complex topic. The following section gives an overview over several general ideas for future research. Subsequently, I will present a concrete example of a possible future study applying representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008).

4.5.1 General Ideas for Future Research

In the present studies, participants only went through one fMRI session, so that we could only evaluate brain activation during the processing of modified episodes. However, for completely understanding under which conditions false memories arise after PE, future study protocols should also include fMRI recordings during initial encoding. Previous work has demonstrated that not only brain activation during the encoding of misinformation, but also during initial encoding can be predictive of the fate of memories. Accordingly, it has been suggested that when the original piece of information receives more neural processing, e.g., in MTL, the memory will be protected from interference. In contrast, when a higher level of neural processing occurs during the encoding of misinformation, false memories are formed (Dennis et al., 2015; Okado & Stark, 2005). Thus, relating the strength of original encoding to the strength of encoding of modified episodes and behavioral outcomes from a memory test might help to further define which conditions promote the formation of false memories in response to mnemonic PEs.

Furthermore, it would be valuable to translate our paradigm to additional modalities aside from fMRI. For example, using EEG, it would be possible to investigate the precise timing underlying neural processing of mnemonic PEs, for example by focusing on the P300 component which is known to signal for unexpected stimuli (Hoy, Steiner, & Knight, 2021;

Luck, 2006; Singh et al., 2018; Stefanics et al., 2014). Even a combination of fMRI and EEG could be implemented within our paradigm (Huster, Debener, Eichele, & Herrmann, 2012). Additionally, some conclusions from our studies could be further validated using TMS to influence brain activation. For example, it could be tested whether mPFC (including ACC) activation is related to the formation of false memories after PE, as our findings would predict. First hints for this relationship have already been demonstrated. For example, it has previously been described that perturbating mPFC processing using TMS resulted in fewer false recognitions of critical lures (Berkers et al., 2017).

Another important next step will be to investigate whether our findings can be generalized to different types of stimuli (e.g., auditory ones) and PEs. By doing so, it will be possible to understand whether our findings might reflect universal processes in the modification of episodic memory, or rather a special case. A first step could be to investigate whether the impacts we showed for content-based gist and surface modifications (**Study 3**) could be translated to PEs that are based on episode structure, like in **Study 1**.

Additionally, the influence of PE strength could be further investigated by systematically varying prior strength (e.g., Ortiz-Tudela et al., 2023), and operationalizing PE impact in a continuous, or at least multi-step, manner. Computational modeling approaches could help to understand the fate of memory traces depending on PE strength even further. For example, it is “well-established anecdotally” (Milton et al., 2023, p. 101) that memory modification occurs when the mismatch between prediction and sensory input is mild, while distinct memories are formed when the mismatch is striking. Computational models could help to understand how this principle specifically relates to PE strength and PE signals in the brain, for example with the use of Hopfield networks (Hopfield, 1982; Zotow, Bisby, & Burgess, 2020).

Furthermore, future research could shed more light on the influence of consolidation. For example, more variations of our retrieval practice sessions could be probed. Specifically, spacing out retrieval sessions even more and/or distributing the material over more retrieval sessions (Dobson, Perez, & Linderholm, 2017; Roediger & Butler, 2011) might yield the benefit in mnemonic retention we could not show yet. Successful consolidation is characterized by increased functional connectivity, for example between mPFC and hippocampus and other areas, such as precuneus (Antony et al., 2017; G. Fernández, 2017; Sterpenich et al., 2009). Also, there are hints that functional connectivity between hippocampus and other cortical areas influence the processing of mnemonic PEs (Sinclair et al., 2021). Therefore, future work could address how functional connectivity after consolidation influences PE processing.

Overall, understanding the influence of PEs on memory is not only of interest to learn more about the functioning of our memory system. In the long run, building upon and further adapting the presented paradigm might yield valuable findings with direct practical implications. For example, results might find application in legal or clinical settings. This could include the scientific evaluation of eye witness testimony reliability, or the targeted use of memory modification techniques in psychotherapy (Brewin, 2015; Krawczyk et al., 2017).

4.5.2 Tracking Mnemonic Representations After Prediction Error: A Proposed Future Study

Research Question

What becomes clear from our studies, but also previous work, is that it is difficult to judge what exactly happens to a memory when it is modified. For example, the main suggestion we derived from our findings, namely that two alternative versions of an episode are encoded and subsequently confused, could reflect the outcome of two processes. Either, completely separate memory traces could be formed which then interfere with each other, or information

from different sources could be combined in a single representation (Krawczyk et al., 2017). Importantly, behavioral tests (Elsey et al., 2018; Tronson & Taylor, 2007) and, potentially, univariate fMRI analyses (Sinclair et al., 2021), cannot resolve this issue. For this reason, it is needed to track mnemonic representations at different stages of the episodic modification paradigm and compare similarities in neural activation patterns using RSA. To this end, neuroimaging data needs to be collected during all experimental phases, initial encoding, modification phase, and final memory test[9].

Background

According to Ritvo et al., (2019), there are two neural mechanisms which can help to avoid confusion and competition between similar memories. On the one hand, neural representations can be differentiated from each other. While overlapping memory features are weakened, the focus is then on distinctive characteristics of single memories (Stawarczyk, Wahlheim, Etzel, Snyder, & Zacks, 2020). This principle of neural differentiation is well established in humans (Xue, 2022) and it has been shown that PE strength is related to the degree of neural differentiation that is achieved (Kim, Norman, & Turk-Browne, 2017). Accordingly, when differentiation is incomplete or weak, memories might interfere (Sommer & Sander, 2022). On the other hand, memories can be integrated with each other, so that a single representation is formed (Ritvo et al., 2019). Ideally, this shared representation contains information that relates both memories to each other, for example, the temporal relationship of the different experiences and underlying source information (Wahlheim & Zacks, 2019; Scully & Hupbach, 2020; Stawarczyk et al., 2020). Indeed, several studies have demonstrated that during the successful retrieval of a target memory, the neural activation pattern of a competitor memory is reinstated as well (Kuhl et al., 2012; Tompary & Davachi, 2017; Ye, Shi, Li, Chen,

[9] Importantly, I assume that most experimental protocols, including the introduction of modifications and the final memory test (previously referred to as post-fMRI memory test), remain the same in this proposed study.

& Xue, 2020). However, when some important detail is missing from the combined representation, e.g., source information, correct recall might be impaired. Thus, both incomplete differentiation or integration could lead to memory modification and explain our current findings. But how could we use measures of representational similarity to further investigate which mechanism underlies false memory formation after PE?

Operationalization and Possible Outcomes

According to Ritvo et al., (2019), only strong reactivation triggers memory integration, while moderate reactivation triggers differentiation. Therefore, the first step would be to evaluate how strongly the original memory is reactivated before the modification is introduced in the second scanning session. To measure this, the representational similarity of the episode part *before* the modification occurs could be compared between initial encoding (scanning session 1) and modification phase (scanning session 2).

Next, the representational similarity between original and modified alternatives of the same episode, as presented during the final memory test (scanning session 3), would be compared and related to the behavioral outcome of said test. In a situation where a strong reactivation of the original episode (scanning sessions 1 & 2) is observed, together with a high similarity of original and modified representations (scanning session 3) and a correct rejection of the modified version in the memory test, memories would likely be successfully integrated with each other. When, instead, the similarity in scanning session 3 is low and a false alarm is observed in the memory test, this is an indication of an incomplete integration. Conversely, when only moderate reactivation is found (scanning sessions 1 & 2), there is low similarity in scanning session 3 and a correct rejection occurs, memories would likely be successfully differentiated. Otherwise, when high similarity (scanning session 3) and a false alarm are observed, differentiation would probably be incomplete.

Apart from a more explorative searchlight approach, several ROIs would be of interest for this analysis. One area of interest would be hippocampus, for which much work concerning memory representations already exists (e.g., Horner & Doeller, 2017; Sinclair et al., 2021; Xue, 2022). Concerning its role in pattern separation and pattern completion (Ngo et al., 2021), which likely play a role in differentiation and integration of memories, hippocampal activation patterns would be especially interesting. Another important ROI would be mPFC (including ACC), which is often implicated in mnemonic integration (Ritvo et al., 2019; Stawarczyk et al., 2020; Tompary & Davachi, 2017; Ye et al., 2020).

What would need to be clarified is how this idea of incomplete differentiation or integration can be related to the suggestion that memory modification is actually beneficial, as it allows us to make better predictions (Exton-McGuinness et al., 2015; R. S. Fernández et al., 2016). Importantly, missing completeness does not necessarily equal failure. One could speculate that from a predictive coding point of view, this might rather be the most parsimonious solution to when there is no clear predictive benefit of encoding specific differences between episodes. If, for example, it is signaled that telling specific memories apart is highly relevant for future predictions, for example because one becomes subjectively aware of a modification (Wahlheim et al., 2019), this might foster completion of either differentiation or integration. However, this hypothesis needs thorough evaluation, for example with the help of computational models.

5 Conclusion

Episodic memories are not exact copies of past experiences, but can change on the basis of new, relevant information. A recent line of research has identified mnemonic PEs as a potential trigger for such modifications.

The three presented studies provide further, highly replicable evidence that different types of mnemonic PEs can contribute to memory modification. This effect manifested through the increased tendency to endorse modified episodes as originally encoded after PE. Furthermore, for the first time, the present studies systematically evaluated neural signatures of different types of PEs in episodic memories. Results demonstrated that two brain regions, namely vlPFC (IFS/IFG) and pIPS/SPL, commonly responded to mnemonic PEs, implicating a general role of these two regions in the processing of episodic mismatch and mnemonic selection processes. Furthermore, different types of PEs recruited different brain regions, depending on the piece of episodic information that was violated and memory solidity. Additionally, for the first time, it was demonstrated that activation in the episodic memory network, including hippocampus and ACC, specifically increased over time when false memories were formed. Overall, our findings hint at a special role for (pregenual) ACC in memory modification through mnemonic PEs, potentially related to the relearning of alternative, competing episodes.

In conclusion, the presented studies provide valuable new insights into the neural processing of mnemonic PE and false memory formation. Specifically, our behavioral as well as neural findings suggest that, in the case of false memories, a second episode alternative is encoded which subsequently competes with the original one. A pronounced (neural) mismatch signal, e.g., in pIPS and vlPFC, might prevent the formation of false memories by allowing a clear mnemonic distinction of episodes. Furthermore, our new, naturalistic paradigm provides a viable starting point that future episodic memory research can build upon.

References

Abe, N. (2012). Neuroimaging studies of false memory: A selective review. *Psychologia*, *55*, 131–145. https://doi.org/10.2117/psysoc.2012.131

Ahlheim, C., Schiffer, A. M., & Schubotz, R. I. (2016). Prefrontal cortex activation reflects efficient exploitation of higher-order statistical structure. *Journal of Cognitive Neuroscience*, *28*, 1909–1922. https://doi.org/10.1162/jocn

Alberini, C. M., & Ledoux, J. E. (2013). Memory reconsolidation. *Current Biology*, *23*, R746–R750. https://doi.org/10.1016/j.cub.2013.06.046

Alkire, M. T., Haier, R. J., Fallon, J. H., & Cahill, L. (1998). Hippocampal, but not amygdala, activity at encoding correlates with long-term, free recall of nonemotional information. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 14506–14510. https://doi.org/10.1073/pnas.95.24.14506

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295. https://doi.org/10.1016/B978-1-4832-1446-7.50016-9

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

Asperholm, M., Högman, N., Rafi, J., & Herlitz, A. (2019). What did you do yesterday? A meta-analysis of sex differences in episodic memory. *Psychological Bulletin*, *145*, 785–821. https://doi.org/10.1037/bul0000197

Atance, C. M., & O'Neill, D. K. (2005). The emergence of episodic future thinking in humans. *Learning and Motivation*, *36*, 126–144. https://doi.org/10.1016/j.lmot.2005.02.003

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology*, *192*, 101821. https://doi.org/10.1016/j.pneurobio.2020.101821

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology*, *4*, 907. https://doi.org/10.3389/fpsyg.2013.00907

Baym, C. L., & Gonsalves, B. D. (2010). Comparison of neural activity that leads to true memories, false memories, and forgetting: An fMRI study of the misinformation effect. *Cognitive, Affective and Behavioral Neuroscience*, *10*, 339–348. https://doi.org/10.3758/CABN.10.3.339

Bein, O., Duncan, K., & Davachi, L. (2020). Mnemonic prediction errors bias hippocampal states. *Nature Communications*, *11*, 3451. https://doi.org/10.1038/s41467-020-17287-1

Bein, O., Plotkin, N. A., & Davachi, L. (2021). Mnemonic prediction errors promote detailed memories. *Learning and Memory*, *28*, 422–434. https://doi.org/10.1101/LM.053410.121

Berkers, R. M. W. J., van der Linde, 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*, *88*, 42–52. https://doi.org/10.1016/j.cortex.2016.12.015

Bosshardt, S., Degonda, N., Schmidt, C. F., Boesiger, P., Nitsch, R. M., Hock, C., & Henke, K. (2005). One month of human memory consolidation enhances retrieval-related hippocampal activity. *Hippocampus*, *15*, 1026–1040. https://doi.org/10.1002/hipo.20105

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *347*, 3–47. https://doi.org/10.1016/0273-2297(90)90003-M

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology*, *71*, 81–129. https://doi.org/10.1006/jecp.1998.2464

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, *2*, 331–350. https://doi.org/10.1016/0010-0285(71)90019-3

Brázdil, M., Roman, R., Falkenstein, M., Daniel, P., Jurák, P., & Rektor, I. (2002). Error processing - Evidence from intracerebral ERP recordings. *Experimental Brain Research*, *146*, 460–466. https://doi.org/10.1007/s00221-002-1201-y

Brewin, C. R. (2015). Reconsolidation versus retrieval competition: Rival hypotheses to explain memory change in psychotherapy. *Behavioral and Brain Sciences*, *38*, 21–22. https://doi.org/10.1017/S0140525X14000144

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The influence of prior knowledge on memory: A developmental cognitive neuroscience perspective. *Frontiers in Behavioral Neuroscience*, *7*, 139. https://doi.org/10.3389/fnbeh.2013.00139

Brodt, S., Pöhlchen, D., Flanagin, V. L., Glasauer, S., Gais, S., & Schönauer, M. (2016). Rapid and independent memory formation in the parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 13251–13256. https://doi.org/10.1073/pnas.1605719113

Bryant, N. B., Nadel, L., & Gómez, R. L. (2019). Associations between sleep and episodic memory updating. *Hippocampus*, *30*, 794–805. https://doi.org/10.1002/hipo.23178

Bubic, A., von Cramon, D. Y., Jacobsen, T., Schröger, E., & Schubotz, R. I. (2009). Violation of expectation: Neural correlates reflect bases of prediction. *Journal of Cognitive Neuroscience*, *21*, 155–168. https://doi.org/10.1162/jocn.2009.21013

Buckner, R. L., & Carroll, D. C. (2006). Self-projection and the brain. *Trends in Cell Biology*, *11*, 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends in Cognitive Sciences*, *16*, 338–352. https://doi.org/10.1016/j.tics.2012.04.008

Cabeza, R., Locantore, J. K., & Anderson, N. D. (2003). Lateralization of prefrontal activity during episodic memory retrieval: Evidence for the production-monitoring hypothesis. *Journal of Cognitive Neuroscience*, *15*, 249–259. https://doi.org/10.1162/089892903321208187

Cabeza, R., Mazuz, Y. S., Stokes, J., Kragel, J. E., Woldorff, M. G., Ciaramelli, E., . . . Moscovitch, M. (2011). Overlapping parietal activity in memory and perception: Evidence for the attention to memory model. *Journal of Cognitive Neuroscience*, *23*, 3209–3217. https://doi.org/10.1162/jocn_a_00065

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. *Learning and Memory*, *18*, 523–528. https://doi.org/10.1101/lm.2135211

Cheng, S., & Werning, M. (2013). Composition and replay of mnemonic sequences: The contributions of REM and slow-wave sleep to episodic memory. *Behavioral and Brain Sciences*, *36*, 610–611. https://doi.org/10.1017/S0140525X13001234

Cheng, S., & Werning, M. (2016). What is episodic memory if it is a natural kind? *Synthese*, *193*, 1345–1385. https://doi.org/10.1007/s11229-014-0628-6

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience and Biobehavioral Reviews*, *60*, 82–89. https://doi.org/10.1016/j.neubiorev.2015.11.011

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*, 177–184. https://doi.org/10.1016/j.conb.2007.03.005

Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46*, 1828–1851. https://doi.org/10.1016/j.neuropsychologia.2008.03.022

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2015). Radical predictive processing. *Southern Journal of Philosophy*, *53*, 3–27. https://doi.org/10.1111/sjp.12120

Conway, M. A. (2008). Exploring episodic memory. In E. Dere, A. Easton, L. Nadel, & J. P. Huston (Eds.), *Handbook of episodic memory* (1st ed., pp. 19–29). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S1569-7339(08)00202-6

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, *126*, 393–413. https://doi.org/10.1037/0096-3445.126.4.393

Conway, M. A., & Loveday, C. (2015). Remembering, imagining, false memories & personal meanings. *Consciousness and Cognition*, *33*, 574–581. https://doi.org/10.1016/j.concog.2014.12.002

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215. https://doi.org/10.1038/nrn755

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A. E., . . . Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal

associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron*, *44*, 877–888. https://doi.org/10.1016/j.neuron.2004.11.022

D'Astolfo, L., & Rief, W. (2017). Learning about expectation violation from prediction error paradigms - A meta-analysis on brain processes following a prediction error. *Frontiers in Psychology*, *8*, 1253. https://doi.org/10.3389/fpsyg.2017.01253

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, *136*, 87–102. https://doi.org/10.1037/a0017937

Daselaar, S. M., Prince, S. E., & Cabeza, R. (2004). When less means more: Deactivations during encoding that predict subsequent memory. *Neuroimage*, *23*, 921–927. https://doi.org/10.1016/j.neuroimage.2004.07.031

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*, 693–700. https://doi.org/10.1016/j.conb.2006.10.012

Davis, H. P., & Squire, L. R. (1984). Protein synthesis and memory: A review. *Psychological Bulletin*, *96*, 518–559. https://doi.org/10.1037/0033-2909.96.3.518

De Oliveira Alvares, L., Crestani, A. P., Cassini, L. F., Haubrich, J., Santana, F., & Quillfeldt, J. A. (2013). Reactivation enables memory updating, precision-keeping and strengthening: Exploring the possible biological roles of reconsolidation. *Neuroscience*, *244*, 42–48. https://doi.org/10.1016/j.neuroscience.2013.04.005

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*, 309–330. https://doi.org/10.1007/BF01320097

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, 548. https://doi.org/10.3389/fpsyg.2012.00548

Dennis, N. A., Bowman, C. R., & Turney, I. C. (2015). Functional neuroimaging of false memories. In D. R. Addis, M. Barense, & A. Duarte (Eds.), *The Wiley handbook on the cognitive neuroscience of memory* (1st ed., pp. 150–171). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118332634.ch8.

Dobson, J. L., Perez, J., & Linderholm, T. (2017). Distributed retrieval practice promotes superior recall of anatomy information. *Anatomical Sciences Education*, *10*, 339–347. https://doi.org/10.1002/ase.1668

Dodson, C. S., & Schacter, D. L. (2002). The cognitive neuropsychology of false memories: Theory and data. In A. D. Baddely, M. D. Kopelman, & B. A. Wilson (Eds.), *The handbook of memory disorders* (2nd ed., pp. 343–362). Chichester, UK: John Wiley & Sons, Ltd.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86. https://doi.org/10.1146/annurev.psych.55.090902.142050

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus*, *22*, 389–398. https://doi.org/10.1002/hipo.20933

El Haj, M., Colombel, F., Kapogiannis, D., & Gallouj, K. (2020). False memory in Alzheimer's disease. *Behavioural Neurology*, *2020*, 5284504. https://doi.org/10.1155/2020/5284504

Elsey, J. W. B., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin*, *144*, 797–848. https://doi.org/10.1037/bul0000152

El-Sourani, N., Trempler, I., Wurm, M. F., Fink, G. R., & Schubotz, R. I. (2019). Predictive impact of contextual objects during action observation: Evidence from functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *32*, 326–337. https://doi.org/10.1162/jocn_a_01480

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, *76*, 1057–1070. https://doi.org/10.1016/j.neuron.2012.12.002

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories—The role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011

Ezzyat, Y., Inhoff, M. C., & Davachi, L. (2018). Differentiation of human medial prefrontal cortex activity underlies long-term resistance to forgetting in memory. *Journal of Neuroscience*, *38*, 10244–10254. https://doi.org/10.1523/JNEUROSCI.2290-17.2018

Fernández, G. (2017). The medial prefrontal cortex is a critical hub in the declarative memory system. In N. Axmacher & B. Rasch (Eds.), *Cognitive neuroscience of memory consolidation* (1st ed., pp. 45–56). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-45066-7_6

Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of prediction error. *Neuroscience and Biobehavioral Reviews*, *68*, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004

Finn, B. (2017). A framework of episodic updating: An account of memory updating after retrieval. *Progress in Brain Research*, *232*, 173–211. https://doi.org/10.1016/bs.plm.2017.03.006

Fletcher, P. C., Firth, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S. J., & Dolan, R. J. (1995). The mind's eye—activation of the precuneus in memory related imagery. *Neuroimage*, *2*, 195–200. https://doi.org/10.1006/nimg.1995.1025

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory Insights from functional neuroimaging. *Brain*, *124*, 849–881. https://doi.org/10.1093/brain/124.5.849

Frankland, P. W., Josselyn, S. A., & Köhler, S. (2019). The neurobiological foundation of memory retrieval. *Nature Neuroscience*, *22*, 1576–1585. https://doi.org/10.1038/s41593-019-0493-1

Franks, J. J., & Bransford, J. D. (1970). Abstraction of visual patterns. *Journal of Experimental Psychology*, *90*, 65–74. https://doi.org/10.1037/h0031349

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*, 1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*, 1019–1021. https://doi.org./ 10.1038/s41593-018-0200-7

Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, *4*, 49–59. https://doi.org/10.1016/j.tine.2015.01.001

Gershman, S. J., Monfils, M. H., Norman, K. A., & Niv, Y. (2017). The computational nature of memory modification. *eLife*, *6*, e23763. https://doi.org/10.7554/eLife.23763

Gershman, S. J., Radulescu, A., Norman, K. A., & Niv, Y. (2014). Statistical computations underlying the dynamics of memory updating. *PLoS Computational Biology*, *10*, 1003939. https://doi.org/10.1371/journal.pcbi.1003939

Giovanello, K. S., Schnyer, D. M., & Verfaellie, M. (2004). A critical role of the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. *Hippocampus*, *14*, 5–8. https://doi.org/10.1002/hipo.10182

Gisquet-Verrier, P., & Riccio, D. C. (2018). Memory integration: An alternative to the consolidation/reconsolidation hypothesis. *Progress in Neurobiology*, *171*, 15–31. https://doi.org/10.1016/j.pneurobio.2018.10.002

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*, 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Gonsalves, B., Reber, P. J., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Paller, K. A. (2004). Neural evidence that vivid imagining can lead to false remembering. *Psychological Science*, *15*, 655–660. https://doi.org/10.1111/j.0956-7976.2004.00736.x

Greve, A., Abdulrahman, H., & Henson, R. N. (2018). Neural differentiation of incorrectly predicted memories. *Frontiers in Human Neuroscience*, *12*, 278. https://doi.org/10.3389/fnhum.2018.00278

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, *148*, 325–341. https://doi.org/10.1037/xge0000498

Hasselmo, M. E. (2012). *How we remember: Brain mechanisms of episodic memory*. Cambridge, MA: MIT Press. https://doi.org/10.5860/choice.49-5040

Hayama, H. R., Vilberg, K. L., & Rugg, M. D. (2012). Overlap between the neural correlates of cued recall and source memory: Evidence for a generic recollection network? *Journal of Cognitive Neuroscience*, *24*, 1127–1137. https://doi.org/10.1162/jocn_a_00202

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

Heins, N., Pomp, J., Kluger, D. S., Trempler, I., Zentgraf, K., Raab, M., & Schubotz, R. I. (2020). Incidental or intentional? Different brain responses to one's own action sounds in hurdling vs. tap dancing. *Frontiers in Neuroscience*, *14*, 483. https://doi.org/10.3389/fnins.2020.00483

Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event- related functional magnetic resonance imaging study. *The Journal of Neuroscience*, *19*, 3962–3972. https://doi.org/10.1523/JNEUROSCI.19-10-03962.1999

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*, 1315–1326. https://doi.org/10.1002/hipo.20857

Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, *287*, 1269–1272. https://doi.org/10.1126/science.287.5456.1269

Hermann, M. M., Wahlheim, C. N., Alexander, T. R., & Zacks, J. M. (2021). The role of prior-event retrieval in encoding changed event features. *Memory and Cognition*, *49*, 1387–1404. https://doi.org/10.3758/s13421-021-01173-2

Himmer, L., Schönauer, M., Heib, D. P. J., Schabus, M., & Gais, S. (2019). Rehearsal initiates systems memory consolidation, sleep makes it last. *Science Advances*, *5*, eaav1695. https://doi.org/10.1126/sciadv.aav1695

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

Horner, A. J., & Doeller, C. F. (2017). Plasticity of hippocampal memories in humans. *Current Opinion in Neurobiology*, *43*, 102–109. https://doi.org/10.1016/j.conb.2017.02.004

Howe, M. L. (1998). When distinctiveness fails, false memories prevail. *Journal of Experimental Child Psychology*, *71*, 170–177. https://doi.org/10.1006/jecp.1998.2469

Hoy, C. W., Steiner, S. C., & Knight, R. T. (2021). Single-trial modeling separates multiple overlapping prediction errors during reward processing in human EEG. *Communications Biology*, *4*, 910. https://doi.org/10.1038/s42003-021-02426-1

Hrkać, M., Wurm, M. F., Kühn, A. B., & Schubotz, R. I. (2015). Objects mediate goal integration in ventrolateral prefrontal cortex during action observation. *PLoS One*, *10*, e0134316. https://doi.org/10.1371/journal.pone.0134316

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 580–593. https://doi.org/10.1002/wcs.142

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). A subtle reminder triggers integration of new information. *Learning & Memory*, *14*, 47–53. https://doi.org/10.1101/lm.365707.effects

Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: Updating or source confusion? *Memory*, *17*, 502–510. https://doi.org/10.1080/09658210902882399

Huster, R. J., Debener, S., Eichele, T., & Herrmann, C. S. (2012). Methods for simultaneous EEG-fMRI: An introductory review. *Journal of Neuroscience*, *32*, 6053–6060. https://doi.org/10.1523/JNEUROSCI.0447-12.2012

Inostroza, M., & Born, J. (2013). Sleep for preserving and transforming episodic memory. *Annual Review of Neuroscience*, *36*, 79–102. https://doi.org/10.1146/annurev-neuro-062012-170429

Jainta, B., Siestrup, S., El-Sourani, N., Trempler, I., Wurm, M. F., Werning, M., . . . Schubotz, R. I. (2022). Seeing what I did (not): Cerebral and behavioral effects of agency and perspective on episodic memory re-activation. *Frontiers in Behavioral Neuroscience*, *15*, 793115. https://doi.org/10.3389/fnbeh.2021.793115

Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral Cortex*, *17*, 2507–2515. https://doi.org/10.1093/cercor/bhl156

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Johnson, M. K., Raye, C. L., Mitchell, K. J., & Ankudowich, E. (2012). The cognitive neuroscience of true and false memories. In R. F. Belli (Ed.), *True and false recovered memories: Toward a reconciliation of the debate* (1st ed., pp. 15–52). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-1195-6_2

Judge, M. E., & Ouartermain, D. (1982). Characteristics of retrograde amnesia following reactivation of memory in mice. *Physiology & Behavior*, *28*, 585–590. https://doi.org/10.1016/0031-9384(82)90034-8

Kafkas, A., & Montaldi, D. (2018). Expectation affects learning and modulates memory experience at retrieval. *Cognition*, *180*, 123–134. https://doi.org/10.1016/j.cognition.2018.07.010

Kapur, S., Craik, F. I. M., Jones, C., Brown, G. M., Houle, S., & Tulving, E. (1995). Functional role of the prefrontal cortex in memory retrieval: A PET study. *NeuroReport*, *6*, 1880–1884. https://doi.org/ 10.1097/00001756-199510020-00014

Kastner, L., Kube, J., Villringer, A., & Neumann, J. (2017). Cardiac concomitants of feedback and prediction error processing in reinforcement learning. *Frontiers in Neuroscience*, *11*, 598. https://doi.org/10.3389/fnins.2017.00598

Kiley, C., & Parks, C. M. (2022). Mechanisms of memory updating: State dependency vs. reconsolidation. *Journal of Cognition*, *5*, 7. https://doi.org/10.5334/JOC.198

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 8997–9002. https://doi.org/10.1073/pnas.1319438111

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

Kirwan, C. B., & Stark, C. E. L. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning and Memory*, *14*, 625–633. https://doi.org/10.1101/lm.663507

Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Error awareness and the insula: Links to neurological and psychiatric diseases. *Frontiers in Human Neuroscience*, *7*, 14. https://doi.org/10.3389/fnhum.2013.00014

Klingmüller, A., Caplan, J. B., & Sommer, T. (2017). Intrusions in episodic memory: Reconsolidation or interference? *Learning and Memory*, *24*, 216–224. https://doi.org/10.1101/lm.045047.117

Kluen, L. M., Dandolo, L. C., Jocham, G., & Schwabe, L. (2019). Dorsolateral prefrontal cortex enables updating of established memories. *Cerebral Cortex*, *29*, 4154–4168. https://doi.org/10.1093/cercor/bhy298

Krawczyk, M. C., Fernández, R. S., Pedreira, M. E., & Boccia, M. M. (2017). Toward a better understanding on the role of prediction error on memory processes: From bench to clinic. *Neurobiology of Learning and Memory*, *142*, 13–20. https://doi.org/10.1016/j.nlm.2016.12.011

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis–connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro .06.004.2008

Kuhl, B. A., Bainbridge, W. A., & Chun, M. M. (2012). Neural reactivation reveals mechanisms for updating memory. *The Journal of Neuroscience*, *32*, 3453–3461. https://doi.org/10.1523/JNEUROSCI.5846-11.2012

Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nature Neuroscience*, *10*, 908–914. https://doi.org/10.1038/nn1918

Kumaran, D., & Maguire, E. A. (2006). An unexpected sequence of events: Mismatch detection in the human hippocampus. *PLoS Biology*, *4*, 2372–2382. https://doi.org/10.1371/journal.pbio.0040424

Kumaran, D., & Maguire, E. A. (2007). Match–mismatch processes underlie human hippocampal responses to associative novelty. *Journal of Neuroscience*, *27*, 8517–8524. https://doi.org/10.1523/JNEUROSCI.1677-07.2007

Kveraga, K., Ghuman, A. S., Kassam, K. S., Aminoff, E. A., Hämäläinen, M. S., Chaumon, M., & Bar, M. (2011). Early onset of neural synchronization in the contextual associations network. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 3389–3394. https://doi.org/10.1073/pnas.1013760108

Lee, J. L. C. (2010). Memory reconsolidation mediates the updating of hippocampal memory content. *Frontiers in Behavioral Neuroscience*, *4*, 168. https://doi.org/10.3389/fnbeh.2010.00168

Li, C., & Yang, J. (2020). Role of the hippocampus in the spacing effect during memory retrieval. *Hippocampus*, *30*, 703–714. https://doi.org/10.1002/hipo.23193

Loftus, E. F. (1996). Memory distortion and false memory creation. *Bulletin of the American Academy of Psychiatry and the Law*, *24*, 281–295.

Loftus, E. F. (1997). Creating false memories. *Scientific American*, *277*, 70–75. https://doi.org/10.1038/scientificamerican0997-70.

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

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *Journal of Neuroscience*, *36*, 12677–12687. https://doi.org/10.1523/JNEUROSCI.1850-16.2016

Luck, S. J. (2006). *An introduction to event-related potential technique* (1st ed.). Cambridge, MA: The MIT Press.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*, 279–284. https://doi.org/10.1177/0963721415570732

Maguire, E. A., Henson, R. N. A., Mummery, C. J., & Frith, C. D. (2001). Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. *NeuroReport*, *12*, 441–444. https://doi.org/10.1097/00001756-200103050-00004

Manns, J. R., & Squire, L. R. (2002). The medial temporal lobe and memory for facts and events. In A. D. Baddely, M. D. Kopelman, & B. A. Wilson (Eds.), *The handbook of memory disorders* (2nd ed., pp. 81–100). Chichester, UK: John Wiley & Sons, Ltd.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Martin-Ordas, G., & Call, J. (2013). Episodic memory: A comparative approach. *Frontiers in Behavioral Neuroscience*, *7*, *1–13*. https://doi.org/10.3389/fnbeh.2013.00063

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Milton, A. L., Das, R. K., & Merlo, E. (2023). The challenge of memory destabilisation: From prediction error to prior expectations and biomarkers. *Brain Research Bulletin*, *194*, 100–104. https://doi.org/10.1016/j.brainresbull.2023.01.010

Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science*, *160*, 554–555. https://doi.org/10.1126/science.160.3827.554

Molenberghs, P., Mesulam, M. M., Peeters, R., & Vandenberghe, R. R. C. (2007). Remapping attentional priorities: Differential contribution of superior parietal lobule and intraparietal sulcus. *Cerebral Cortex*, *17*, 2703–2712. https://doi.org/10.1093/cercor/bhl179

Moscovitch, M. (2003). Memory consolidation. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (1st ed., pp. 1–11). New York, NY: Nature Publishing Group.

Moscovitch, M., & Gilboa, A. (2021). Systems consolidation, transformation and reorganization: Multiple trace theory, trace transformation theory and their competitors. In M.J. Kahana, & A.D. Wagner (Eds.), *The Oxford handbook of human memory*. PsyArXiv. https://doi.org/10.31234/osf.io/yxbrs

Moscovitch, M., Winocur, G., Ryan, L., & Nadel, L. (2008). Functional neuroanatomy of remote, episodic memory. In E. Dere, A. Easton, L. Nadel, & J. P. Huston (Eds.), *Handbook of episodic memory* (1st ed., pp. 239–269). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S1569-7339(08)00214-2

Nadel, L., Hupbach, A., Hardt, O., & Gomez, R. (2008). Episodic memory: Reconsolidation. In E. Dere, A. Easton, L. Nadel, & J. P. Huston (Eds.), *Handbook of episodic memory* (1st ed., pp. 43–56). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S1569-7339(08)00204-X

Nader, K. (2015). Reconsolidation and the dynamic nature of memory. In Giese, K., & Radwanska, K. (Eds.), *Novel mechanisms of memory* (1 ed., pp. 1–20). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-24364-1_1

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

Nave, K., Deane, G., Miller, M., & Clark, A. (2020). Wilding the predictive brain. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*, e1542. https://doi.org/10.1002/wcs.1542

Ngo, C. T., Michelmann, S., Olson, I. R., & Newcombe, N. S. (2021). Pattern separation and pattern completion: Behaviorally separable processes? *Memory and Cognition*, *49*, *193–205*. https://doi.org/10.3758/s13421-020-01072-y

Nyberg, L., Tulving, E., Habib, R., Nilsonn, L.-G., Kapur, S., Houle, S., . . . McIntosh, A. R. (1995). Functional brain maps of retrieval mode and recovery of episodic information. *NeuroReport*, *7*, *249–252*.

Okado, Y., & Stark, C. E. L. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning and Memory*, *12*, 3–11. https://doi.org/10.1101/lm.87605

Ortiz-Tudela, J., Nolden, S., Pupillo, F., Ehrlich, I., Schommartz, I., Turan, G., & Shing, Y. L. (2023). Not what U expect: Effects of prediction errors on episodic memory. *Journal of Experimental Psychology: General.* https://doi.org/10.1037/xge0001367

Otgaar, H., Muris, P., Howe, M. L., & Merckelbach, H. (2017). What drives false memories in psychopathology? A case for associative activation. *Clinical Psychological Science*, *5*, 1048–1069. https://doi.org/10.1177/2167702617724424

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*, 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Pedreira, M. E., Pérez-Cuesta, L. M., & Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learning and Memory*, *11*, 579–585. https://doi.org/10.1101/lm.76904

Pillemer, D. B., Steiner, K. L., Kuwabara, K. J., Thomsen, D. K., & Svob, C. (2015). Vicarious memories. *Consciousness and Cognition*, *36*, 233–245. https://doi.org/10.1016/j.concog.2015.06.010

Poe, G. R., Walsh, C. M., & Bjorness, T. E. (2010). Both duration and timing of sleep are important to memory consolidation. *Sleep*, *33*, 1277–1278. https://doi.org/10.1093/sleep/33.10.1277

Pomp, J., Heins, N., Trempler, I., Kulvicius, T., Tamosiunaite, M., Mecklenbrauck, F., . . . Schubotz, R. I. (2021). Touching events predict human action segmentation in brain and

behavior. *Neuroimage*, *243*, 118534. https://doi.org/10.1016/j.neuroimage.2021.118534

Quent, J. A., Greve, A., & Henson, R. N. (2022). Shape of U: The nonmonotonic relationship between object–location memory and expectedness. *Psychological Science*, *33*, 2084–2097. https://doi.org/10.1177/09567976221109134

Ranganath, C. (2010). Binding items and contexts: The cognitive neuroscience of episodic memory. *Current Directions in Psychological Science*, *19*, 131–137. https://doi.org/10.1177/0963721410368805

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87. https://doi.org/10.1038/4580

Reichardt, R., Polner, B., & Simor, P. (2020). Novelty manipulations, memory performance, and predictive coding: The role of unexpectedness. *Frontiers in Human Neuroscience*, *14*, 152. https://doi.org/10.3389/fnhum.2020.00152

Rekkas, P. V, & Constable, R. T. (2005). Evidence that autobiographic memory retrieval does not become independent of the hippocampus: An fMRI study contrasting very recent with remote events. *Journal of Cognitive Neuroscience*, *17*, 1950–1961. https://doi.org/10.1162/089892905775008652

Reyna, V. F. (1995). Interference effects in memory and reasoning. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (1st ed., pp. 29–59). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012208930-5/50003-9

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75. https://doi.org/10.1016/1041-6080(95)90031-4

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, *5*, 1–9. https://doi.org/10.1016/j.jarmac.2015.12.003

Richards, B. A., & Frankland, P. W. (2017). The persistence and transience of memory. *Neuron*, *94*, 1071–1084. https://doi.org/10.1016/j.neuron.2017.04.037

Richter, M., Amunts, K., Mohlberg, H., Bludau, S., Eickhoff, S. B., Zilles, K., & Caspers, S. (2019). Cytoarchitectonic segregation of human posterior intraparietal and adjacent parieto-occipital sulcus and its relation to visuomotor and cognitive functions. *Cerebral Cortex*, *29*, 1305–1327. https://doi.org/10.1093/cercor/bhy245

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic plasticity: How memory retrieval drives learning. *Trends in Cognitive Sciences*, *23*, 726–742. https://doi.org/10.1016/j.tics.2019.06.007

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27. https://doi.org/10.1016/j.tics.2010.09.003

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*, 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*, 1432–1463. https://doi.org/10.1037/a0037559

Rugg, M. D. (2002). Functional neuroimaging of memory. In A. D. Baddely, M. D. Kopelman, & B. A. Wilson (Eds.), *The handbook of memory disorders* (2nd ed., pp. 57–80). Chichester, UK: John Wiley & Sons, Ltd.

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, *119*, 2073–2083. https://doi.org/10.1093/brain/119.6.2073

Rugg, M. D., Johnson, J. D., Park, H., & Uncapher, M. R. (2008). Encoding-retrieval overlap in human episodic memory: A functional neuroimaging perspective. *Progress in Brain Research*, *169*, 339–352. https://doi.org/10.1016/S0079-6123(07)00021-0

Rugg, M. D., Johnson, J. D., & Uncapher, M. R. (2015). Encoding and retrieval in episodic memory: Insights from fMRI. In D. R. Addis, M. Barense, & A. Duarte (Eds.), *The Wiley handbook on the cognitive neuroscience of memory* (1st ed., pp. 84–107). Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/9781118332634.ch5

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

Rushworth, M. F. S., Noonan, M. A. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, *70*, 1054–1069. https://doi.org/10.1016/j.neuron.2011.05.014

Ryan, L., Hoscheidt, S., & Nadel, L. (2008). Perspectives on episodic and semantic memory retrieval. In E. Dere, A. Easton, L. Nadel, & J. P. Huston (Eds.), *Handbook of episodic memory* (1st ed., pp. 5–18). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S1569-7339(08)00201-4

Ryan, L., Nadel, L., Keil, K., Putnam, K., Schnyer, D., Trouard, T., & Moscovitch, M. (2001). Hippocampal complex and retrieval of recent and very remote autobiographical memories: Evidence from functional magnetic resonance imaging in neurologically intact people. *Hippocampus*, *11*, 707–714. https://doi.org/10.1002/hipo.1086

Sara, S. J. (2000). Retrieval and reconsolidation: Toward a neurobiology of remembering. *Learning & Memory*, *7*, 73–84. https://doi.org/10.1101/lm.7.2.73

Sayood, K. (2018). Information theory and cognition: A review. *Entropy*, *20*, 706. https://doi.org/10.3390/e20090706

Schacter, D. L. (1995). Memory distortion: History and current status. In D. L. Schacter (Ed.), *Memory distortion: How minds, brains, and societies reconstruct the past* (1st ed., pp. 1–46). Cambridge, MA: Harvard University Press.

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

Schacter, D. L. (2012). Constructive memory: Past and future. *Dialogues in Clinical Neuroscience*, *14*, 7–18. https://doi.org/10.31887/dcns.2012.14.1/dschacter

Schacter, D. L. (2019). Implicit memory, constructive memory, and imagining the future: A career perspective. *Perspectives on Psychological Science*, *14*, 256–272. https://doi.org/10.1177/1745691618803640

Schacter, D. L., & Addis, D. R. (2007a). Constructive memory: The ghosts of past and future. *Nature*, *445*, 27. https://doi.org/10.1038/445027a

Schacter, D. L., & Addis, D. R. (2007b). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 773–786. https://doi.org/10.1098/rstb.2007.2087

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, *8*, 657–661. https://doi.org/10.1038/nrn2213

Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*, *467–474*. https://doi.org/10.1016/j.tics.2011.08.004

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, D. L., & Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron*, *44*, 149–160. https://doi.org/10.1016/j.neuron.2004.08.017

Schiffer, A. M., Ahlheim, C., Ulrichs, K., & Schubotz, R. I. (2013). Neural changes when actions change: Adaptation of strong and weak expectations. *Human Brain Mapping*, *34*, 1713–1727. https://doi.org/10.1002/hbm.22023

Schiffer, A. M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS One*, *7*, 1–11. https://doi.org/10.1371/journal.pone.0036445

Schiffer, A. M., Krause, K. H., & Schubotz, R. I. (2013). Surprisingly correct: Unexpectedness of observed actions activates the medial prefrontal cortex. *Human Brain Mapping*, *35*, 1615–1629. https://doi.org/10.1002/hbm.22277

Schiffer, A. M., & Schubotz, R. I. (2011). Caudate nucleus signals for breaches of expectation in a movement observation paradigm. *Frontiers in Human Neuroscience*, *5*, 38. https://doi.org/10.3389/fnhum.2011.00038

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 8151. https://doi.org/10.1038/ncomms9151

Schubotz, R. I. (2015). Prediction and expectation. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference* (Vol. 3, pp. 295–302). Cambridge, MA: Academic Press: Elsevier. https://doi.org/10.1016/B978-0-12-397025-1.00205-0

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473–500. https://doi.org/10.1146/annurev.neuro.23.1.473

Scully, I. D., & Hupbach, A. (2020). Different reactivation procedures enable or prevent episodic memory updating. *Hippocampus*, *30*, 806–814. https://doi.org/10.1002/hipo.23159

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin and Review*, *18*, 455–468. https://doi.org/10.3758/s13423-011-0086-9

Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learning and Memory*, *21*, 580–584. https://doi.org/10.1101/lm.035493.114

Shimamura, A. P. (2008). A neurocognitive approach to metacognitive monitoring and control. In J. Dunlosky & R. Bjork (Eds.), *Handbook of memory and metacognition* (1st ed., pp. 373–390). Mahwah, NJ: Psychology Press.

Shimamura, A. P. (2011). Episodic retrieval and the cortical binding of relational activity. *Cognitive, Affective, & Behavioral Neuroscience*, *11*, 277–291. https://doi.org/10.3758/s13415-011-0031-4

Shimamura, A. P. (2014). Remembering the past: Neural substrates underlying episodic encoding and retrieval. *Current Directions in Psychological Science*, *23*, 257–263. https://doi.org/10.1177/0963721414536181

Siestrup, S., Jainta, B., Cheng, S., & Schubotz, R. I. (2023). Solidity meets surprise: Cerebral and behavioral effects of learning from episodic prediction errors. *Journal of Cognitive Neuroscience*, *35*, 291–313. https://doi.org/10.1162/jocn_a_01948

Siestrup, S., Jainta, B., El-Sourani, N., Trempler, I., Wurm, M. F., Wolf, O. T., . . . Schubotz, R. I. (2022). What happened when? Cerebral processing of modified structure and content in episodic cueing. *Journal of Cognitive Neuroscience*, *34*, 1287–1305. https://doi.org/10.1162/JOCN_A_01862

Siestrup, S., & Schubotz, R. I. (2023). Minor changes change memories: FMRI and behavioral reflections of episodic prediction errors. *Manuscript Submitted for Publication*.

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning and Memory*, *25*, 369–381. https://doi.org/10.1101/lm.046912.117

Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neurosciences*, *42*, 727–739. https://doi.org/10.1016/j.tins.2019.08.007

Sinclair, A. H., Manalili, G. M., Brunec, I. K., Adcock, R. A., & Barense, M. D. (2021). Prediction errors disrupt hippocampal representations and update episodic memories. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2117625118. https://doi.org/10.1073/pnas.2117625118

Singh, A. K., Chen, H. T., Cheng, Y. F., King, J. T., Ko, L. W., Gramann, K., & Lin, C. T. (2018). Visual appearance modulates prediction error in virtual reality. *IEEE Access*, *6*, 24617–24624. https://doi.org/10.1109/ACCESS.2018.2832089

Sommer, V. R., & Sander, M. C. (2022). Contributions of representational distinctiveness and stability to memory performance and age differences. *Aging, Neuropsychology, and Cognition*, *29*, 443–462. https://doi.org/10.1080/13825585.2021.2019184

Squire, L. R. (1992a). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, *4*, 232–243. https://doi.org/10.1162/jocn.1992.4.3.232

Squire, L. R. (1992b). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*, 195–231. https://doi.org/10.1037/0033-295X.99.2.195

St. Jacques, P. L., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 19671–19678. https://doi.org/10.1073/pnas.1319630110

Stawarczyk, D., Wahlheim, C. N., Etzel, J. A., Snyder, A. Z., & Zacks, J. M. (2020). Aging and the encoding of changes in events: The role of neural activity pattern reinstatement. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 29346–29353. https://doi.org/10.1073/pnas.1918063117

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience*, *8*, 666. https://doi.org/10.3389/fnhum.2014.00666

Sterpenich, V., Albouy, G., Darsaud, A., Schmidt, C., Vandewalle, G., Dang Vu, T. T., . . . Maquet, P. (2009). Sleep promotes the neural reorganization of remote emotional memory. *Journal of Neuroscience*, *29*, 5143–5152. https://doi.org/10.1523/JNEUROSCI.0561-09.2009

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*, 299–351. https://doi.org/10.1017/S0140525X07001975

Suddendorf, T., & Corballis, M. C. (2008). Episodic memory and mental time travel. In E. Dere, A. Easton, L. Nadel, & J. P. Huston (Eds.), *Handbook of episodic memory* (1st ed., pp. 31–42). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S1569-7339(08)00203-8

Sugar, J., & Moser, M. B. (2019). Episodic memory: Neuronal codes for what, where, and when. *Hippocampus*, *29*, 1190–1205. https://doi.org/10.1002/hipo.23132

Suzuki, W. A. (2007). Integrating associative learning signals across the brain. *Hippocampus*, *17*, 842–850. https://doi.org/10.1002/hipo

Takashima, A., Nieuwenhuis, I. L. C., Rijpkema, M., Petersson, K. M., Jensen, O., & Fernández, G. (2007). Memory trace stabilization leads to large-scale changes in the retrieval network: A functional MRI study on associative memory. *Learning and Memory*, *14*, 472–479. https://doi.org/10.1101/lm.605607

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., . . . Fernández, G. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 756–761. https://doi.org/10.1073/pnas.0507774103

Tamber-Rosenau, B. J., Esterman, M., Chiu, Y. C., & Yantis, S. (2011). Cortical mechanisms of cognitive control for shifting attention in vision and working memory. *Journal of Cognitive Neuroscience*, *23*, 2905–2919. https://doi.org/10.1162/jocn.2011.21608

Thakral, P. P., Yu, S. S., & Rugg, M. D. (2015). The hippocampus is sensitive to the mismatch in novelty between items and their contexts. *Brain Research*, *1602*, 144–152. https://doi.org/10.1016/j.brainres.2015.01.033

Tompary, A., & Davachi, L. (2017). Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron*, *96*, 228-241.e5. https://doi.org/10.1016/j.neuron.2017.09.005

Tranel, D., & Damasio, A. R. (2002). Neurobiological foundations of human memory. In A. D. Baddely, M. D. Kopelman, & B. A. Wilson (Eds.), *The handbook of memory disorders* (2nd ed., pp. 17–56). Chichester, UK: John Wiley & Sons, Ltd.

Tribus, M. (1961). *Thermodynamics and thermostatics: An introduction to energy, information and states of matter, with engineering applications*. New York, NY: Van Nostrand.

Tronson, N. C., & Taylor, J. R. (2007). Molecular mechanisms of memory reconsolidation. *Nature Reviews Neuroscience*, *8*, 262–275. https://doi.org/10.1038/nrn2090

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (2nd ed., pp. 381–403). Cambridge, MA: Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21037-7

Tulving, E. (1983). *Elements of episodic memory*. New York, NY: Oxford University Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*, 1–12. https://doi.org/10.1037/h0080017

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In H. S. Terrace, & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (1st ed., pp. 3–56). New York: Oxford University Press.

Tulving, E., Kapur, S., Craik, F. I. M., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Science of the United States of America*, *91*, 2016–2020. https://doi.org/10.1073/pnas.91.6.2016

Turan, G., Ehrlich, I., Shing, Y. L., & Nolden, S. (2023). From generating to violating predictions: The effects of prediction error on episodic memory. PsyArXiv. https://doi.org/10.31234/osf.io/zm29a

Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: Insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiology of Learning and Memory*, *91*, 139–154. https://doi.org/10.1016/j.nlm.2008.10.011

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122–129. https://doi.org/10.1037/h0022014

Vassena, E., Holroyd, C. B., & Alexander, W. H. (2017). Computational models of anterior cingulate cortex: At the crossroads between prediction and effort. *Frontiers in Neuroscience*, *11*, 316. https://doi.org/10.3389/fnins.2017.00316

Wagner, A. D. (2002). Cognitive control and episodic memory: Contributions from prefrontal cortex. In L. R. Squire & D. L. Schacter (Eds.), *Neuropsychology of memory* (3rd ed., pp. 174–192). New York, NY: Guilford.

Wahlheim, C. N., Smith, W. G., & Delaney, P. F. (2019). Reminders can enhance or impair episodic memory updating: A memory-for-change perspective. *Memory*, *27*, 849–867. https://doi.org/10.1080/09658211.2019.1582677

Wahlheim, C. N., & Zacks, J. M. (2019). Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General*, *148*, 30–50. https://doi.org/10.1037/xge0000458

Wang, Y., Ma, N., He, X., Li, N., Wei, Z., Yang, L., . . . Zhang, X. (2017). Neural substrates of updating the prediction through prediction error during decision making. *Neuroimage*, *157*, 1–12. https://doi.org/10.1016/j.neuroimage.2017.05.041

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin*, *121*, 331–354. https://doi.org/10.1037/0033-2909.121.3.331

Wichert, S., Wolf, O. T., & Schwabe, L. (2013). Updating of episodic memories depends on the strength of new learning after memory reactivation. *Behavioral Neuroscience*, *127*, 331–338. https://doi.org/10.1037/a0032028

Wiklund-Hörnqvist, C., Stillesjö, S., Andersson, M., Jonsson, B., & Nyberg, L. (2021). Retrieval practice facilitates learning by strengthening processing in both the anterior and posterior hippocampus. *Brain and Behavior*, *11*, e01909. https://doi.org/10.1002/brb3.1909

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*, 582–589. https://doi.org/10.1038/nn.3973

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: Contextual information modulates action recognition. *Neuroimage*, *59*, 1551–1559. https://doi.org/10.1016/j.neuroimage.2011.08.038

Xue, G. (2022). From remembering to reconstruction: The transformative neural representation of episodic memory. *Progress in Neurobiology*, *219*, 102351. https://doi.org/10.1016/j.pneurobio.2022.102351

Yassa, M. A., & Reagh, Z. M. (2013). Competitive trace theory: A role for the hippocampus in contextual interference during retrieval. *Frontiers in Behavioral Neuroscience*, *7*, 107. https://doi.org/10.3389/fnbeh.2013.00107

Yazin, F., Das, M., Banerjee, A., & Roy, D. (2021). Contextual prediction errors reorganize naturalistic episodic memories in time. *Scientific Reports*, *11*, 12364. https://doi.org/10.1038/s41598-021-90990-1

Ye, Z., Shi, L., Li, A., Chen, C., & Xue, G. (2020). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *eLife*, *9*, e57023. https://doi.org/10.7554/eLife.57023

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience*, *20*, 364–375. https://doi.org/10.1038/s41583-019-0150-4

Zacks, J. M., Bezdek, M. A., & Cunningham, G. E. (2021). Knowledge and the reliability of constructive memory. *Memory*, *30*, 22–25. https://doi.org/10.1080/09658211.2020.1871022

Zhan, L., Guo, D., Chen, G., & Yang, J. (2018). Effects of repetition learning on associative recognition over time: Role of the hippocampus and prefrontal cortex. *Frontiers in Human Neuroscience*, *12*, 277. https://doi.org/10.3389/fnhum.2018.00277

Zotow, E., Bisby, J. A., & Burgess, N. (2020). Behavioral evidence for pattern separation in human episodic memory. *Learning and Memory*, *27*, 301–309. https://doi.org/10.1101/LM.051821.120

Appendix

Table A1. Peak activations from second-level whole-brain analyses of contrasts between later false alarms and correct rejections at FDR $p < .05$/uncorrected at $p < .001$ (voxel level)

Localization	H	Cluster extent	MNI Coordinates			*t*-value
			x	*y*	*z*	
fa > cr (uncorrected at $p < .001$)						
Cuneus	R + L	42	0	-88	26	3.87
Calcarine sulcus	R	60	15	-82	14	4.35
Lingual gyrus	R	l.m.	9	-76	-1	4.20
cr > fa (FDR-corrected at $p < .05$)						
Posterior intraparietal sulcus	R	199	30	-61	44	4.29
Superior parietal lobe	R	l.m.	30	-61	62	4.06
Posterior intraparietal sulcus	L	253	-27	-73	29	5.22
Superior parietal lobe	L	l.m.	-27	-61	47	5.18
Inferior frontal sulcus (BA 44/45)	L	85	-42	8	32	5.63
	R	219	42	8	29	5.93
Middle occipital gyrus	R	35	36	-76	-1	4.43
	L	22	-33	-88	-1	3.76
Occipitotemporal cortex	L	133	-42	-61	-7	5.17
	R	160	48	-55	-10	5.63
Inferior frontal gyrus (BA 47)	R	71	42	38	-13	4.57
Anterior insula	R	l.m.	30	26	-1	3.86
	L	28	-33	17	-7	3.58
Fusiform gyrus	R	27	33	-55	-13	3.62
Parahippocampal gyrus	R	l.m.	36	-49	-7	3.62

Note: H = Hemisphere, MNI = Montreal Neurological Institute, L = Left, R = Right, BA = Brodmann Area. l.m. = local maximum, fa = false alarms, cr = correct rejections. Only clusters with a minimum extent of 10 voxels are reported. Unpublished data from GLM3 (Siestrup et al., 2023).

Table A2. Peak activations from second-level whole-brain analyses of parametric effect (increase) for later correct rejections and original episodes at FDR $p < .05$ (voxel level)

Localization	H	Cluster extent	MNI Coordinates			*t*-value
			x	*y*	*z*	
Parametric modulator cr - increasing (FDR-corrected at $p < .05$; not significant at $p < .01$)						
Precentral gyrus	R	16	24	-16	62	3.33
Midcingulate cortex	R	829	6	-22	38	5.49

Central sulcus/precentral gyrus	L	l.m.	-42	-22	56	4.71
Precentral gyrus	R	l.m.	6	-31	53	4.62
Midcingulate cortex	R	16	6	2	38	3.49
Central sulcus	R	14	39	-10	38	3.41
Parieto-occipital fissure	R + L	20	0	-79	38	3.20
Superior frontal gyrus	R	168	24	53	32	4.39
Superior frontal sulcus	R	l.m.	24	35	32	3.48
Superior frontal gyrus	L	120	-30	44	32	3.95
Middle frontal gyrus	L	l.m.	-36	29	38	3.90
Superior frontal gyrus	R	42	30	62	8	3.81
Supramarginal gyrus	R	13	60	-58	32	3.44
Caudate nucleus	R	11	18	14	20	3.21
	L	15	-18	23	2	3.63
	R	80	15	23	-7	4.53
Superior temporal sulcus	R	598	60	-22	8	5.87
Heschl's gyrus	R	l.m.	36	-28	11	5.59
Precentral gyrus	R	l.m.	60	5	11	4.54
Superior temporal sulcus	L	592	-63	-31	2	5.92
Precentral gyrus	L	l.m.	-54	-1	11	4.93
Supramarginal gyrus	L	l.m.	-51	-52	32	4.63
Superior temporal gyrus/lateral sulcus	L	l.m.	-54	-37	14	4.60
Hippocampus	R	10	27	-34	5	3.40
Lingual gyrus	L	624	-9	-67	-4	5.25
	R	l.m.	12	-64	-4	4.72
Cerebellum	R	l.m.	48	-70	-31	4.58
Inferior frontal gyrus (BA 47)	L	30	-42	41	-16	3.72
Middle temporal gyrus	L	23	-54	-28	-19	4.80
Cerebellum	L	157	-36	-79	-31	4.24
	L	28	-42	-61	-37	3.35
Parametric modulator ori - increasing (FDR-corrected at $p < .05$; also significant at $p < .01$)						
Superior temporal sulcus	R	35	60	-22	2	7.73

Note: H = Hemisphere, MNI = Montreal Neurological Institute, L = Left, R = Right, BA = Brodmann Area. l.m. = local maximum, cr = correct rejections, ori = originals. Only clusters with a minimum extent of 10 voxels are reported. Unpublished data from GLM3 (Siestrup et al., 2023).

Table A3. Peak activations from second-level whole-brain analyses of parametric effect (decrease) for later correct rejections and original episodes at FDR $p < .05$/uncorrected at $p < .001$ (voxel level)

Localization	H	Cluster extent	MNI Coordinates			*t*-value
			x	*y*	*z*	
Parametric modulator cr - decreasing (uncorrected at $p < .001$)						
Parieto-occipital fissure	R	14	21	-55	23	4.20
Parahippocampal gyrus	L	9	-30	-43	-10	4.28
Fusiform gyrus	R	144	36	-55	-13	5.48
	L	16	-39	-58	-13	4.07
Parametric modulator ori - decreasing (FDR-corrected at $p < .05$)						
Superior parietal lobe	L	10	-30	-49	59	3.74
Postcentral sulcus	R	11	33	-40	56	3.72
Posterior intraparietal sulcus	R	38	30	-73	35	4.43
Occipitotemporal cortex	L	272	-39	-79	14	5.14
Fusiform gyrus	L	l.m.	-36	-52	-13	4.81
Occipitotemporal cortex	R	644	48	-70	5	5.43
Fusiform gyrus	R	l.m.	39	-46	-16	5.00

Note: H = Hemisphere, MNI = Montreal Neurological Institute, L = Left, R = Right. l.m. = local maximum, cr = correct rejections, ori = originals. Only clusters with a minimum extent of 9 voxels are reported. Unpublished data from GLM3 (Siestrup et al., 2023).

Abbreviations

ACC	anterior cingulate cortex
BA	Brodmann area
cr	correct rejection(s)
EEG	electroencephalography
fa	false alarm(s)
FG	fusiform gyrus
fMRI	functional magnetic resonance imaging
IFG	inferior frontal gyrus
IFS	inferior frontal sulcus
mPFC	medial prefrontal cortex
MRI	magnetic resonance imaging
MTL	medial temporal lobe
PCC	posterior cingulate cortex
PE	prediction error
PFC	prefrontal cortex
PHG	parahippocampal gyrus
pIPS	posterior intraparietal sulcus
ROI	region of interest
RSA	representational similarity analysis
SPL	superior parietal lobe
TMS	transcranial magnetic stimulation
vlPFC	ventrolateral prefrontal cortex
vPPC	ventral posterior parietal cortex

List of Figures

Figure 1. Schematic graphic summary of brain areas which commonly responded to mnemonic prediction errors.. 135

Curriculum Vitae

Personal Details

Name	Sophie Siestrup
Date of Birth	13/02/1994
Place of Birth	Steinfurt

Education

Since 04/2020	**Doctoral student at the Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience (OCC)** *University of Münster*
10/2016 – 02/2019	**Master of Science in Biosciences (1.0)** *University of Münster*
10/2013 – 09/2016	**Bachelor of Science in Biosciences (1.2)** *University of Münster*
09/2004 – 07/2013	**Abitur (1.0)** *Städtisches Gymnasium Borghorst*

Work Experience

Since 09/2019	**Research associate and doctoral student** *Institute of Psychology, Biological Psychology,* *University of Münster*
03/2019 – 06/2019	**Research associate** *Institute for Neuro- and Behavioural Biology,* *Department of Behavioural Biology,* *University of Münster*
07/2017 – 04/2018	**Student assistant** *Institute for Neuro- and Behavioural Biology,* *Department of Behavioural Biology,* *University of Münster*

Teaching Experience

Winter Semester 2020/2021, 2021/2022, 2022/2023	Seminar “Surprise, surprise! - Neuronal foundations of expectation” for M.Sc. students in Psychology (focus area Cognitive Neuroscience) *University of Münster*

Summer Semester 2018, 2019	Project module for B.Sc. students in Biosciences *University of Münster*
Winter Semester 2014/2015	Chemistry study group for two-subject Bachelor's degree program students in Biology *University of Münster*

Supervised Theses

Summer Semester 2022	Jana Jamuna Halscheid, Master of Science (Psychology)
Winter Semester 2020/2021	Yuyi Xu, Master of Science (Psychology)

Publications

Siestrup, S., & Schubotz, R. I. (2023). Minor changes change memories: FMRI and behavioral reflections of episodic prediction errors. *Manuscript Submitted for Publication*

Siestrup S., Jainta B., Cheng S., & Schubotz R. I. (2023). Solidity meets surprise: Cerebral and behavioral effects of learning from episodic prediction errors. *Journal of Cognitive Neuroscience*, *35*, 291–313. https://doi.org/10.1162/jocn_a_01948

Siestrup, S., Jainta., B., El-Sourani, N., Trempler, I., Wurm, M. F., Wolf, O. T., . . . Schubotz, R. I. (2022). What happened when? Cerebral processing of modified structure and content in episodic cueing. *Journal of Cognitive Neuroscience*, *34*, 1287–1305. https://doi.org/10.1162/jocn_a_01862

Jainta, B., **Siestrup, S.**, El-Sourani, N., Trempler, I., Wurm, M. F., Werning, M., . . . Schubotz, R. I. (2022). Seeing what I did (not): Cerebral and behavioral effects of agency and perspective on episodic memory re-activation. *Frontiers in Behavioral Neuroscience, 15*, 793115. https://doi.org/10.3389/fnbeh.2021.793115

Melotti, L., **Siestrup, S.**, Peng, M., Vitali, V., Dowling, D., von Kortzfleisch, V. T., . . . Richter, S. H. (2021). Individuality, as well as genotype, affects characteristics and temporal consistency of courtship songs in male mice. *Animal Behaviour*, *180*, 179–196. https://doi.org/10.1016/j.anbehav.2021.08.003

Krakenberg, V., **Siestrup, S.**, Palme, R., Kaiser, S., Sachser, N., & Richter, S. H. (2020). Effects of different social experiences on emotional state in mice. *Scientific Reports*, *10*, 15255 https://doi.org/10.1038/s41598-020-71994-9

Bodden, C., **Siestrup, S.**, Palme, R., Kaiser, S., Sachser, N., & Richter, S. H. (2018). Evidence-based severity assessment: Impact of repeated versus single open-field testing on welfare in C57BL/6J mice. *Behavioural Brain Research*, *336*, 261–268. https://doi.org/10.1016/j.bbr.2017.08.029

Declarations

Declarations by the candidate documenting **open-science activities** and **on the consideration of ethical aspects** as part of the doctoral process and disclosure of **personal contribution** for manuscripts completed by two or more authors (cumulative dissertations)

Doctoral candidate: Sophie Siestrup

Title of dissertation: The Influence of Mnemonic Prediction Errors on Brain Activity and Episodic Memory - a Perspective on Memory Modification

1. Documentation of **open-science activities**

Manuscript 1

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data	X		https://osf.io/m7dcu/?view_only=575d6ed3fbf544ada3bcb0519c86f94b
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication	X		https://doi.org/10.1162/jocn_a_01862

Manuscript 2

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication		X	

Manuscript 3

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication		X	

2. Declaration on **the consideration of ethical aspects**

Study number	**Source (manuscript / chapter of dissertation):** e.g., study 1 in paper 2, study 1 described in chapter 4	**Was the study reviewed by an ethics commission?**	
		yes	**No**
1	Manuscript 1 & 2	X	
2	Manuscript 3	X	

3. Declaration of one's **personal contribution** to the submitted academic manuscripts by two or more authors

Manuscript 1

Title	What Happened When? Cerebral Processing of Modified Structure and Content in Episodic Cueing		
Author(s)	Sophie Siestrup, Benjamin Jainta, Nadiya El-Sourani, Ima Trempler, Moritz F. Wurm, Oliver T. Wolf, Sen Cheng, Ricarda I. Schubotz		
Publication status:	not yet submitted		(please mark with X)
	submitted		
	in review		
	in revision		
	accepted		
	published	X	
Journal	Journal of Cognitive Neuroscience		
Year of publication	2022		

Description of your own contribution in the case of joint authorship:

- partly responsible for the study's conception and design
- partly responsible for data collection
- processing, analyzing and interpreting data
- mainly responsible for drafting and revising the manuscript
- marked as corresponding author

Manuscript 2

Title	Solidity Meets Surprise: Cerebral and Behavioral Effects of Learning from Episodic Prediction Errors		
Author(s)	Sophie Siestrup, Benjamin Jainta, Sen Cheng, Ricarda I. Schubotz		
Publication status:	not yet submitted		(please mark with X)
	submitted		
	in review		
	in revision		
	accepted		
	published	X	
Journal	Journal of Cognitive Neuroscience		
Year of publication	2023		

Description of your own contribution in the case of joint authorship:

- partly responsible for the study's conception and design
- partly responsible for data collection (same data as for Manuscript 1)
- partly responsible for processing, analyzing and interpreting data
- partly responsible for drafting and revising the manuscript
- marked as corresponding author

Manuscript 3

Title	Minor Changes Change Memories: FMRI and Behavioral Reflections of Episodic Prediction Errors		
Author(s)	Sophie Siestrup, Ricarda I. Schubotz		
Publication status:	not yet submitted		(please mark with X)
	submitted		
	in review	X	
	in revision		
	accepted		
	published		
Journal	Journal of Cognitive Neuroscience		
Year of publication			

Description of your own contribution in the case of joint authorship:

- partly responsible for the study's conception and design
- collecting, processing, analyzing and interpreting data
- mainly responsible for drafting and revising the manuscript
- marked as corresponding author

Place, Date

Signature of the doctoral candidate