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, Neuro-Behavioral 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

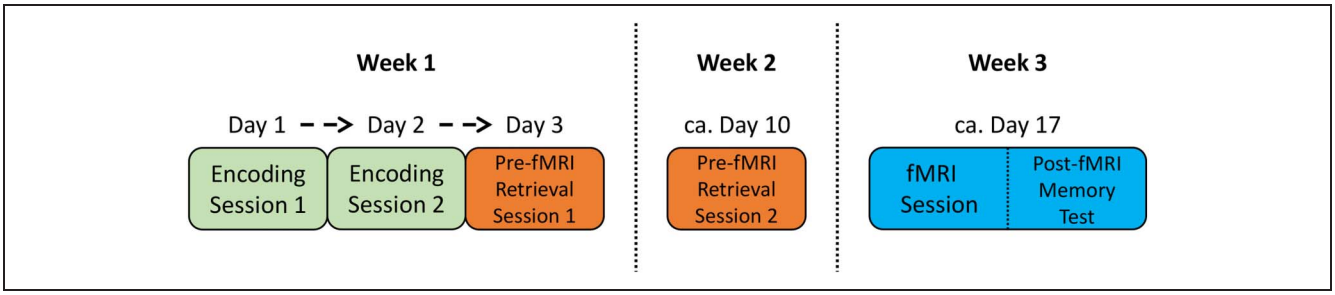

Figure 1. Schematic overview of the experimental procedure. During the first week, participants underwent two encoding sessions and one pre-fMRI retrieval session on three consecutive days. The first retrieval session only contained videos belonging to the spaced condition. Approximately 1 week later, 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.

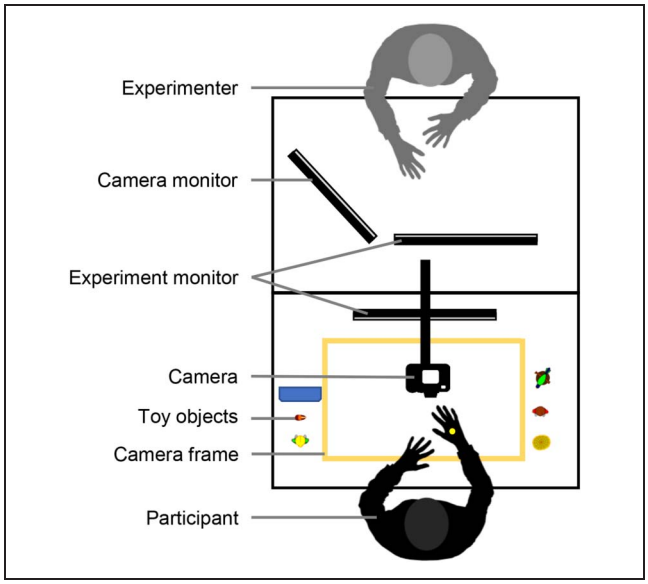

Figure 2. Encoding setup. During training, participants imitated PLAYMOBIL stories, while sitting at the filming setup. Their performance was monitored by the experimenter. Figure adapted from Siestrup et al. (2022).

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

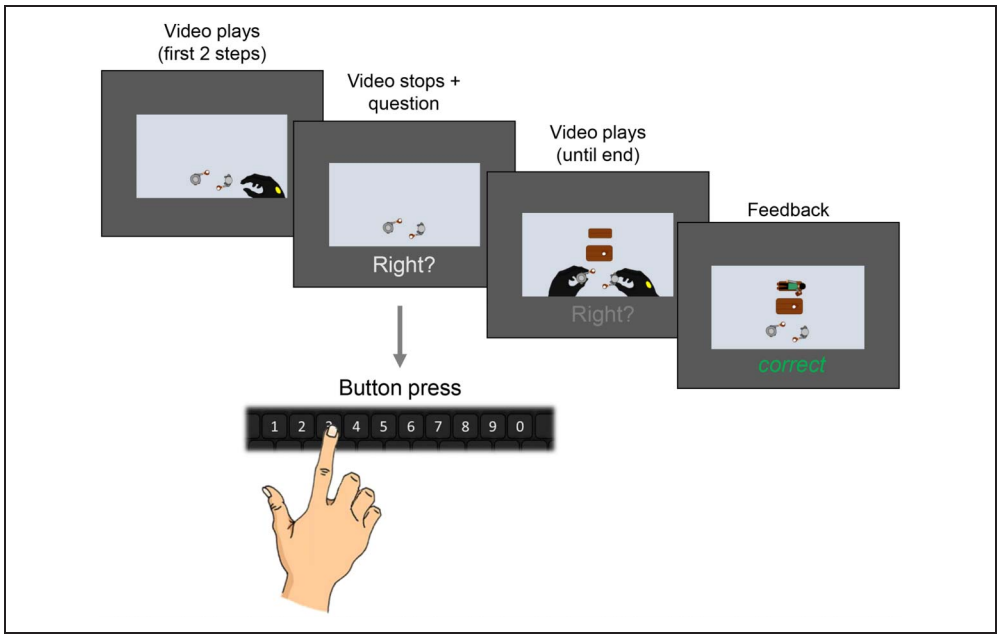

Figure 3. Schematic depiction of retrieval task. Participants were presented the first two steps of a demo video. Then, the video stopped and a question ("Left?" or "Right?") was displayed. Participants had to visualize the story from memory and answer how many steps were conducted with the left or right 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.

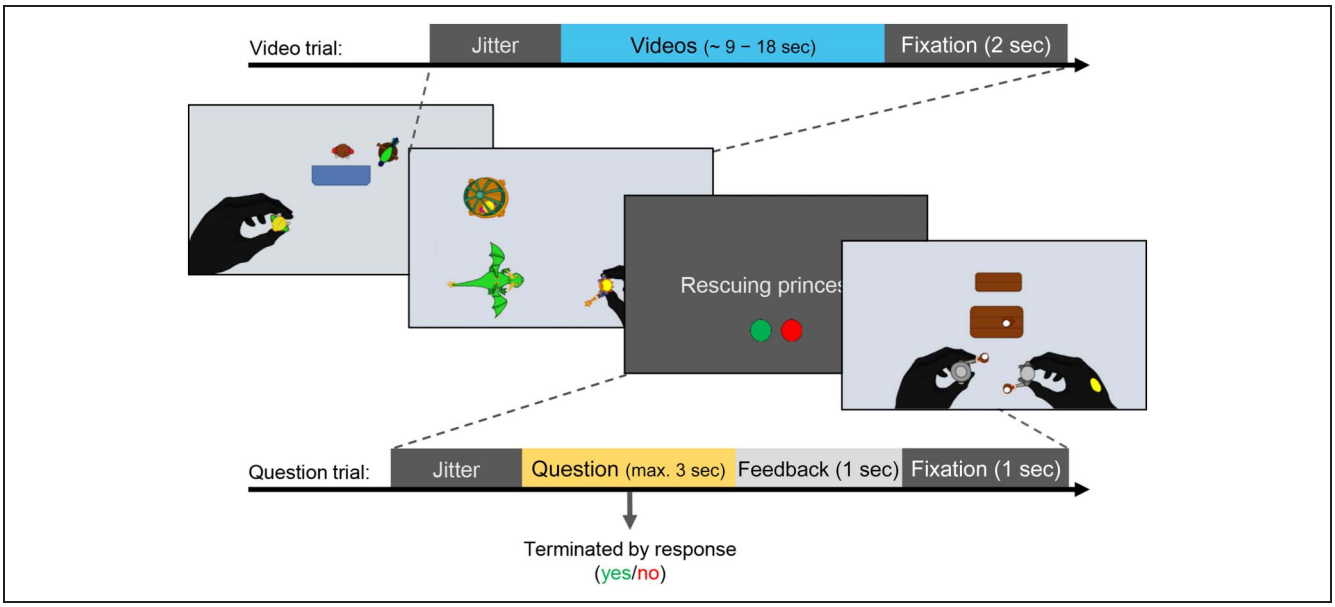

Figure 4. Schematic depiction of task during fMRI session. Video trials included a variable jitter (0, 0.5, 1, or 1.5 sec of fixation), a video showing a PLAYMOBIL story (ca. 9–18 sec), and a 2-sec ISI (fixation). Question trials consisted of a variable jitter, a question regarding the story shown in 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

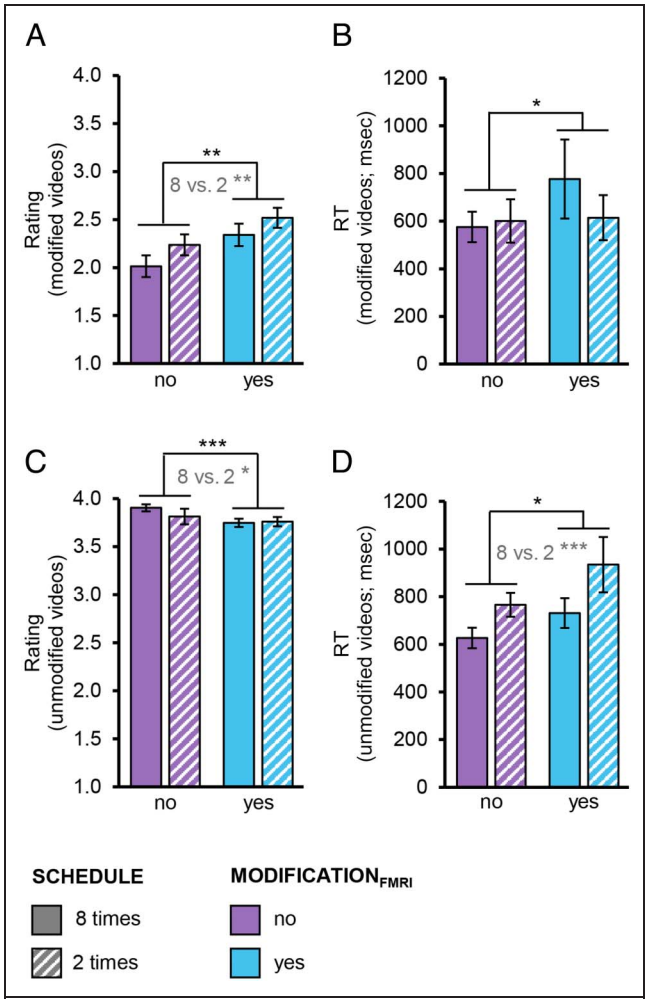

Figure 5. Behavioral results from post-fMRI memory test. Episodes had either been retrieved 8 or 2 times (factor Times) following a spaced or massed schedule (factor Schedule) before entering the fMRI session during which episodes were shown in an original or modified version (factor $\text{Modification}_{\text{FMRI}}$). 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, $F(1, 35) = 0.04$, $p = .85$, $\eta_p^2 = .001$.

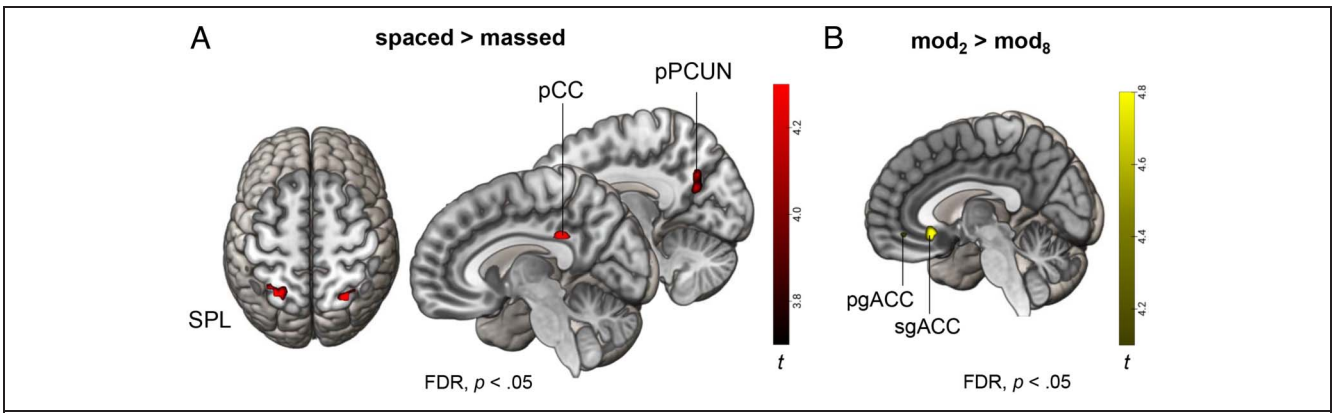

Figure 6. Whole-brain activation for retrieval Schedule and Times effects. (A) FDR-corrected *t*-map ($p < .05$) for the *spaced* > *massed* contrast in episodes (based on GLM 1). (B) FDR-corrected *t*-map ($p < .05$) for the $mod_2 > mod_8$ contrast (based on GLM 2). pCC = posterior cingulate cortex; 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
Pregenual 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.

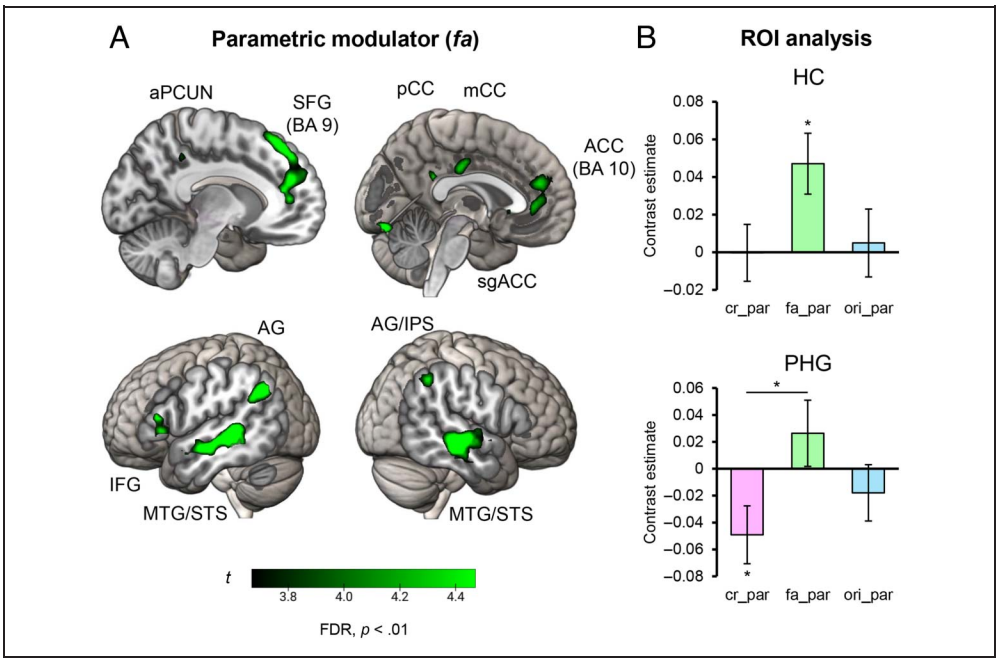

Figure 7. Results from the parametric analysis. (A) FDR-corrected *t*-map ($p < .01$) for the whole-brain contrast of the parametric modulator modeling the repeated presentation of modified episodes, which lead to false alarm responses in the post-fMRI memory test. Activation clusters indicate an increase 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 subthreshold 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

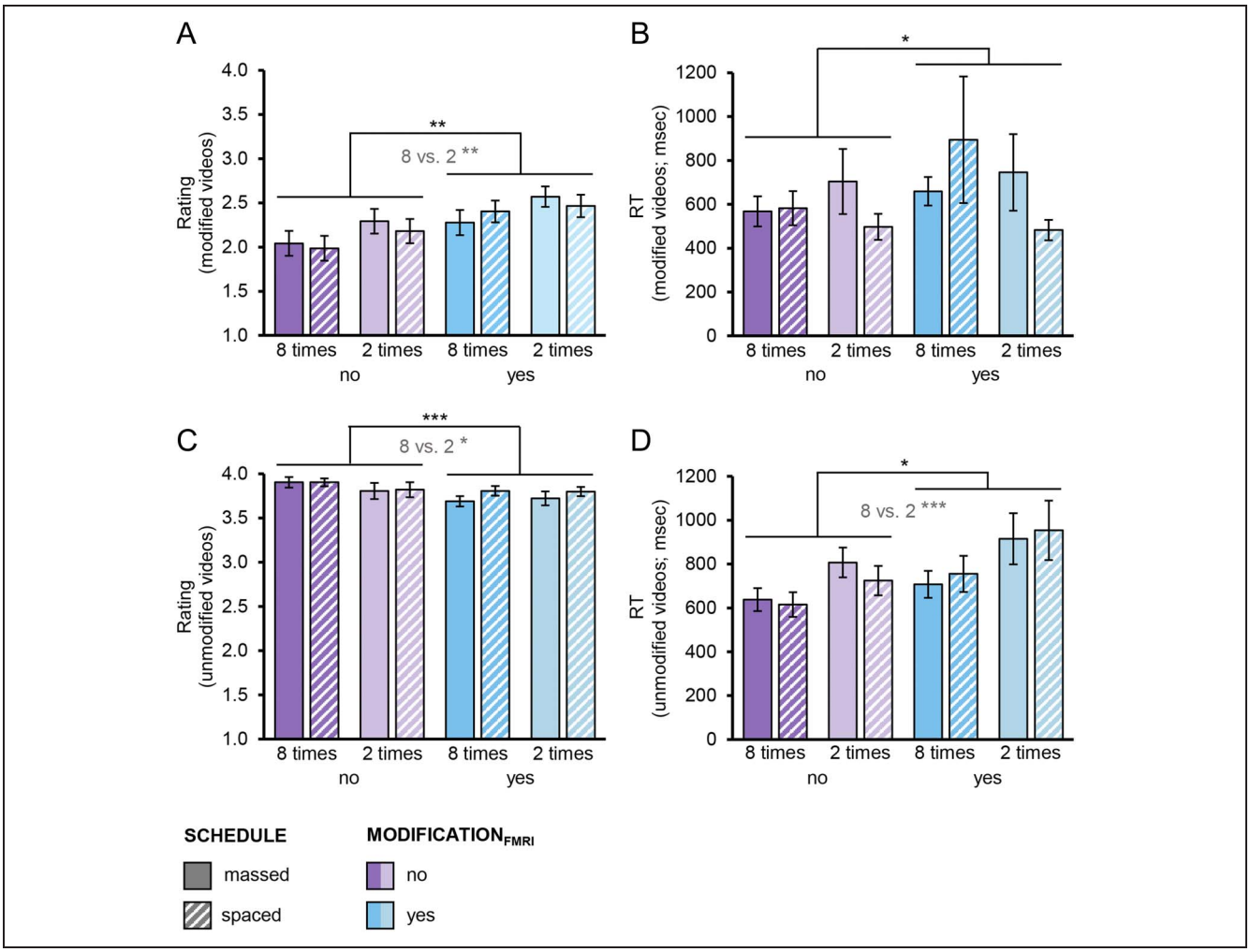

Figure A1. Behavioral results from post-fMRI memory test for all factors. Episodes had either been retrieved 8 or 2 times (factor TIMES) following a spaced or massed schedule (factor SCHEDULE) before entering the fMRI session during which episodes were shown in an original or modified version (factor $\text{MODIFICATION}_{\text{FMRI}}$). 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

Localization	H	Cluster Extent	MNI Coordinates			t Value
			x	y	z	
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

Localization	H	Cluster Extent	MNI Coordinates			t Value
			x	y	z	
(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