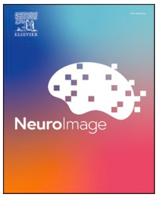
NeuroImage 311 (2025) 121194

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Finding the sweet spot of memory modification: An fMRI study on episodic prediction error strength and type

Nina Liedtke [a,b,*], Marius Boeltzig [a,b], Falko Mecklenbrauck [a,b], Sophie Siestrup [a,b], Ricarda I. Schubotz [a,b]

[a] *Department of Psychology, University of Münster, Fliednerstraße 21, 48149 Münster, Germany*
[b] *Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany*

ARTICLE INFO

Keywords:
Episodic memory
Prediction error
Memory updating
fMRI

ABSTRACT

Previous research has highlighted the critical role of prediction errors (PEs) in signaling the need to adapt memory representations in response to unexpected changes in the environment. Yet, the influence of PE type and strength on memory remains underexplored. In this study, participants encoded naturalistic dialogues prior to undergoing fMRI scanning. During the fMRI session, they listened to dialogues that had been modified in their surface or gist, to varying extents. As expected, our findings revealed robust activation in the inferior frontal gyrus for all PEs. Notably, gist modifications elicited additional activations within the episodic memory network, including the hippocampus. A post-fMRI recognition test demonstrated that surface modifications had no significant impact on memory. Conversely, weak gist changes impaired memory for the original content and hindered learning of the modification. These weak gist changes also triggered activation in the parahippocampal cortex. These results underscore the importance of both the type and strength of PEs in shaping brain activations and memory outcomes, highlighting their complex interplay in cognitive processes.

Even though we may not realize it in our day-to-day life, our episodic memories are not as stable as we think. Previous research has shown that prediction errors (PEs) are one of the driving forces in the modification of memories, as they signal the need to update and dynamically adapt memory representations to changes in the environment (Friston and Kiebel, 2009; Sinclair and Barense, 2019). A mnemonic PE arises when the input we encounter does not match the expectations which are automatically and continuously generated by the brain (Sinclair and Barense, 2019). However, not all PEs lead to the same memory outcomes (Bein et al., 2023). While in some studies the original memory was weakened (Forcato et al., 2007) or pruned (Kim et al., 2014, 2017), in others it was enhanced (Brod et al., 2013; Sinclair et al., 2021). Similarly, the unexpected input can either lead to the formation of new memories (Bein et al., 2021; Brod et al., 2018; Greve et al., 2017) or intrusions (Jainta et al., 2022; Siestrup et al., 2022; Siestrup and Schubotz, 2023; Sinclair et al., 2021; Sinclair and Barense, 2018). Where does this variability of PE effects on memory come from?

The effect that a specific PE has on memory might depend on its strength. PE strength is influenced by the size of the mismatch between expected and encountered input (Greve et al., 2017). According to the Latent Cause Theory (Gershman et al., 2017), a new and separate memory is formed when the current sensory information is so different from existing memories (strong PE) that it is inferred to have been generated by a new latent cause. If the mismatch of old and new sensory information is small (weak PE), no memory modification happens and the original old memory is kept intact, given that it still reflects the state of the environment accurately enough. In contrast, if the old and new sensory information are different, but still inferred to have the same underlying cause (medium-sized PE), the old memory becomes susceptible for modification (Gershman et al., 2017).

Determining the strength of a PE based on a given stimulus remains ambiguous. Previous research has primarily categorized PEs into binary classifications of presence versus absence (Bein et al., 2015; Brod et al., 2015; Greve et al., 2017) or utilized three categories: PE-inducing, expected, and neutral conditions (Greve et al., 2019). Studies that examine more nuanced, quantitatively graded PEs are less common. Examples include a reward learning paradigm (Rouhani et al., 2018) and a behavioral study incorporating two to three different PE strengths

---

[*] Corresponding author at: Department of Psychology, Fliednerstraße 21, 48149 Münster, Germany.
E-mail addresses: nina.liedtke@uni-muenster.de (N. Liedtke), marius.boeltzig@uni-muenster.de (M. Boeltzig), f_meck01@uni-muenster.de (F. Mecklenbrauck), s.siestrup@uni-muenster.de (S. Siestrup), rschubotz@uni-muenster.de (R.I. Schubotz).

https://doi.org/10.1016/j.neuroimage.2025.121194
Received 26 July 2024; Received in revised form 14 March 2025; Accepted 7 April 2025

(Ortiz-Tudela et al., 2023). Importantly, the strength of the PE can be influenced by several factors, including the precision of the new input, the strength of the prediction, and the degree of difference between prediction and new input (Greve et al., 2017; Henson and Gagnepain, 2010). While we have demonstrated the role of prediction strength for memory outcomes in a previous study (Boeltzig et al., 2025), the current study focuses on the degree of difference between prediction and new input.

Another set of studies explored various triggers of PEs, distinguishing between content versus structural changes (Siestrup et al., 2022) and gist versus surface modifications (Siestrup and Schubotz, 2023). A recent study discovered that surface modifications in video narratives led to a higher rate of erroneous acceptance of the modification as the original, while gist modifications did not (Siestrup and Schubotz, 2023). This study also identified distinct brain activation patterns for surface and gist modifications, indicating qualitative differences between these different types of PEs. Consequently, it is plausible that both strength and quality, which we will call type, of PEs influence brain activations and memory outcomes.

In the present fMRI study, we sought to deepen our understanding of how the brain responds to episodic PEs of varying strengths and types, and how these responses influence both the retention of original memories and the integration or learning of new, unexpected information. To this end, we used naturalistic dialogues, some of which were modified over the course of the experiment on one of two ways, namely in *what* was said (gist) or in *how* it was said (surface). In order to generate a wide range of mismatches inducing weak, moderate, and strong PEs, each dialogue was produced in five different versions. In addition to the original version, there were one weak and one strong deviation in both surface and gist. In a first experimental session, participants encoded the original versions of these dialogues. In the next session, they either were presented with the original or a modified version while undergoing fMRI scanning. A post fMRI recognition test was used to assess the stability of the original memory and new learning of the modification. In the last session, participants rated how much the modification differed from the original to obtain a subjective measure of PE size.

Depending on the subjective PE strength and PE type, we expected different outcomes for the memory test and brain activations. On the behavioral level we expected impaired original and modification recognition after moderate PEs, due to an updating process, resulting in one memory trace containing some new and old information (Gershman et al., 2017). For strong PEs we hypothesized participants to perform better at recognizing the modified version, since strong PEs have been found to promote new learning (Gershman et al., 2017). After weak PEs, we did not expect any changes in memory strength.

We thought it possible for moderate PEs to be elicited by strong surface and/or weak gist changes. In a study that used short video stories as stimuli, surface modifications led to more confusion with the original version (Siestrup and Schubotz, 2023), possibly reflecting moderate PEs. However, for verbal stimuli this might be different, since some studies found that ad-verbatim memory for verbal statements can be rather poor (Bransford and Franks, 1971; Brown-Schmidt and Benjamin, 2018; Poppenk et al., 2008). Since the ad-verbatim predictions that can be made should therefore be vague, it is possible that big surface changes do not lead to strong but rather to moderate PEs.

Several studies also found memory impairments caused by stimulus manipulations that may be considered to reflect gist changes, which the authors have argued led to moderate PEs (Kim et al., 2014, 2017) or strong PEs (Siestrup and Schubotz, 2023). Instead of just having one kind of gist manipulation, we aimed to elicit a broad range of PEs, with weaker gist changes potentially leading to moderate PEs and stronger gist changes to strong PEs. We therefore hypothesized moderate PEs to occur for high rated surface and/or low rated gist changes. We considered low rated surface changes as weak PEs and high rated gist changes as strong PEs.

Since activation in the inferior frontal gyrus (IFG) has consistently been observed in response to PEs, regardless of stimulus modality and memory outcomes, it was expected for all types of modifications (Bubic et al., 2009; El-Sourani et al., 2020; Gläscher et al., 2010; Jainta et al., 2024; Wurm and Schubotz, 2012). Apart from this general response to PEs, we hypothesized that both different PE types and strengths would trigger specific activation patterns (Long et al., 2016; Siestrup et al., 2022; Siestrup and Schubotz, 2023). Specifically for medium-sized PEs, which would allow memories to be modified, we expected additional activations. In order to adjust the internal model, the old memory first needs to be retrieved (Gershman et al., 2017) which is why we assumed to see activation in the episodic memory network. This network overlaps with the default mode network and comprises, among others, the medial temporal lobe (MTL), precuneus, posterior parietal cortex (PCC), temporo-parietal junction (TPJ), fronto-median cortex (FMC) and the hippocampus (Jeong et al., 2015; Rugg and Vilberg, 2013; Schacter et al., 2008; Schacter and Addis, 2007). Especially the hippocampal complex has been reliably found after successful memory retrieval (for a review see Rugg and Vilberg, 2013) but also in response to mnemonic PEs (Bein et al., 2020; Duncan et al., 2012; Long et al., 2016). Subsequent memory changes have been related to greater activity in the hippocampal complex following a PE (Shao et al., 2023; Sinclair et al., 2021) and an fMRI study using word-picture pairs found that hippocampal mismatch signals were sensitive to two different categories of PE strength (Long et al., 2016). We therefore expected to see effects in the hippocampus for different prediction error types and strengths, especially after moderate PEs that might lead to changes in memory.

1. Methods

1.1. Participants

A total of 50 participants was recruited for the study. All had normal or corrected-to-normal vision, were native German speakers and were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). They reported no history of neurological or psychiatric disorders or substance abuse. The data from two participants was excluded due to excessive movement in at least one fMRI session and five participants dropped out of the study before completing all sessions. The final sample for fMRI analyses consisted of $N = 43$ participants (36 female, seven male) between 18 and 31 years old ($M = 22.66$, $SD = 3.29$). A previous experiment confirmed that a sample size of $N = 40$ is sufficient to find behavioral effects with this paradigm (Boeltzig et al., 2025). Participants received course credits or money (100€) for their participation and gave written informed consent to participate in this study. The study protocol was approved by the Ethics Committee of the Faculty of Psychology and Sports Science at the University of Münster.

1.2. Stimuli

The stimulus set consisted of 36 dialogues between two speakers each, discussing topics that are relatable for a student audience. The dialogues were written by the authors of this paper and produced by 20 professional voice actors (ten male, ten female, age 31 – 58 years, $M = 40.20$, $SD = 6.86$). Each dialogue had a unique background sound matching the theme and setting of the conversation, resulting in highly naturalistic and meaningful stimuli. The duration varied between 21 and 34 s ($M = 27.31$, $SD = 3.02$). To elicit PEs of varying strength and quality, four modifications were written and recorded for each dialogue (see Fig. 1A). While the head (i.e., the beginning of the dialogue) and the end always remained unchanged across the different versions, one statement in the middle part of the dialogue was altered.

We varied two factors, the type (surface vs. gist) and the degree (low vs. high) of the modification. While the surface changes were only superficial, for example by using a synonym or a change in sentence structure, and did not affect the content of the statement, gist changes were more substantial and affected its meaning. Modifications with a

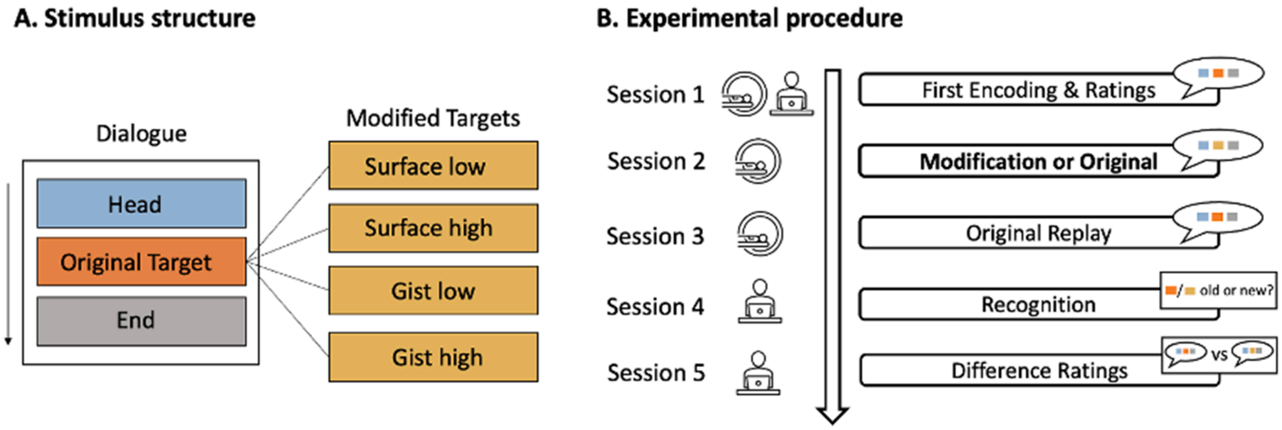


Fig. 1. Experimental stimuli and structure.

Note. Stimuli and structure of the experiment. **A.** The dialogues all followed the underlying structure of head, target, end. For each original target, four modifications were written that differed from the original either in their surface or gist to a low or high degree. Over the course of the experiment, participants first listened to the original version and then to one of the modifications. Heads and ends were never modified and were constructed so that they fit with all five different versions. Note that some dialogues were not altered over the course of the experiment. **B.** The experiment consisted of five sessions, the first three of which were conducted using fMRI. Session 1 started with a first encoding of the dialogues in the scanner. Immediately after that, participants were sat in front of a laptop and listened to the dialogues once more while rating them on different scales. In Session 2, participants heard a modified version for a majority of the dialogues, and in Session 3 they listened once more to the original versions. In Session 4, a recognition test was administered and in Session 5, participants rated the difference between the two dialogue versions as a measure of prediction error strength.

low degree had smaller deviations from the surface or gist of the original than high degree ones. This resulted in four different conditions: surface low (e.g., use of a synonym for one word), surface high (e.g., paraphrase of a sentence), gist low (low alteration of content) and gist high (strong alteration of content; see Table 1 for an example).

While the classification into weak/strong and surface/gist changes served to ensure a wide range of PE strengths during stimulus construction and presentation, the perception of modification degree may vary based on individual factors like knowledge, emotions and personal interests. Therefore, participants gave individual ratings for the modifications at the end of the experiment. In the analysis, these ratings were then used as a continuous measure of modification degree.

Table 1
Translated example dialogue.

	The dialogue is set in a car, there are car sounds in the background.
Head	A: Could you navigate us in a second? Once we turn off the main road, I'm not sure where we're going. B: Sure, just a moment. *(short pause)* Ah, there's a roadworks warning here too, maybe we'd better take a crossroads to the right. Iiiinto this one … oh, we've already passed it. A: *(angry)* Can't you let me know a bit earlier? Where do I have to go now?
Target	**O:** B: *(calmly)* Relax, we'll just take another route. Take the next exit on the right and then the third exit at the traffic circle. **SL:** B: *(calmly)* Relax, we'll just take another route. Take the next exit on the left and then the third exit at the traffic circle. **SH:** B: *(calmly)* Relax, we'll just drive past the roadworks. We have to take the next turn on the right and then the third one at the traffic circle. **GL:** B: *(stressed)* I don't know, the route is still updating. I hope we don't miss another turn! **GH:** B: *(angry)* Why are you getting on my nerves like that? Turn on the GPS from the start if you don't know the way.
End	A: Ah, never mind, the swimming pool is already marked up ahead. I'll just follow the signs.

Note. All dialogues were first played in the original version, using the original target (O). The modified dialogues were later played in one of the four alternative versions: Surface low (SL), surface high (SH), gist low (GL), or gist high (GH). More examples and the presented audios can be obtained from the corresponding author.

1.3. *Procedure*

The experiment consisted of three fMRI sessions and two behavioral sessions that took place over the course of ten days (see Fig. 1B). The experiments in the scanner were implemented in Presentation® (Version 23.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), the experiments conducted in the lab were implemented in PsychoPy (Peirce et al., 2019). The assignment of dialogues to the type of modification was counterbalanced across participants and the stimulus order was randomized. After every tenth trial, there was a ten-second break.

1.3.1. *First encoding and ratings*

In the first session, participants listened to 30 of the 36 dialogues during an fMRI scanning session. Each session started with a short soundcheck, during which participants could self-adjust the volume. They were instructed to listen attentively to the dialogues and to imagine themselves in the scene, as if overhearing a conversation between two people. As an attention check, after each dialogue, a word was presented on the screen and participants had to indicate with a button press if they had heard this word in the dialogue or not (see Fig. 2A). The word was either chosen from the head or the end of the dialogue or was an unused word, that was however semantically related to the topic of the conversation.

After the fMRI session, participants were seated in a separate room in front of a laptop and listened to the same 30 dialogues once more. Instead of an attention check, after each dialogue, they gave ratings regarding everyday typicality ("How typical is a situation like this in your everyday life?"), social consistency ("To what extent do the characters in the dialog adhere to general social norms?"), valence ("Does the dialogue have a positive or negative impression on you?"), arousal ("To what extent did the dialogue evoke emotions in you?"), and autobiographical association ("How vivid was the personal memory triggered by the dialogue?"). These ratings are not further analyzed in this paper. The duration of the first session was around 60 min in total (30 min for each task).

1.3.2. *Modifications*

Session 2 took place exactly two days after the initial encoding. The instructions were the same as in Session 1 and dialogues were followed up with an attention check. Participants were informed that they would

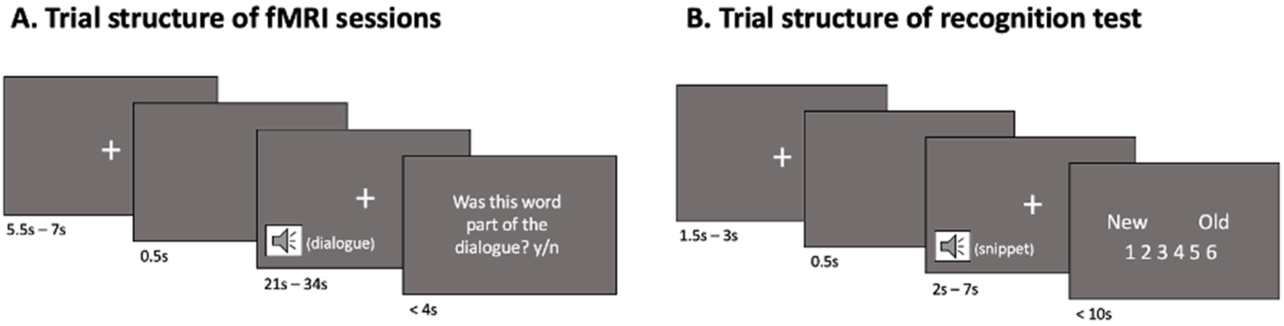


Fig. 2. Trial structure.

Note. All trials started with a jittered fixation cross that disappeared for 0.5 s before stimulus presentation to heighten participants' attention. While the audio played the fixation cross remained on the screen. **A.** In the three fMRI sessions participants then saw a word and had to decide whether this word occurred in the dialogue. **B.** In the recognition session the task was to decide whether the snippet was old or new.

listen to dialogues again, but not that 24 of the 30 original dialogues would be modified (changing dialogues; with six dialogues in each of the four modification categories). The other six dialogues from Session 1 remained unmodified (unchanging dialogues). Additionally, six new dialogues were included in this session, which had not been presented in Session 1 (referred to as novels). The two unmodified conditions serve as a memory baseline and allowed us to check for unique effects of the modifications. All dialogues in this session, namely the 24 changing dialogues, the six unchanging dialogues and the six novels were presented twice. The session took around 50 min.

1.3.3. Original replay

In Session 3, the originals from Session 1 were played once again. This served for a pre-post comparison of the original dialogues which is not analyzed in this paper.

1.3.4. Recognition

Five days after the replay session, a recognition test was used to assess the memory for the original and the modified episodes (see Fig. 2B). Participants heard short snippets (2.4 s – 6.9 s, $M = 4.55$ s) from the dialogues and were instructed to indicate on a scale from 1 = definitely new to 6 = definitely old (Brady et al., 2023) whether they had heard this exact version over the course of the experiment or not.

For each dialogue, four snippets were presented. For the changing dialogues these were the original target and the modified target that they heard during the experiment and two unheard modifications which served as similar lures. For the unmodified dialogues, namely the unchanging dialogues and the novels, participants also heard the original target as well as two unheard modifications. Additionally, a snippet of the dialogue head was presented, to balance out old and new answers and to keep the chance of guessing at 50 % within each dialogue. Snippets relating to the same dialogue were spaced out over the session, by covertly dividing the task into four blocks with only one snippet from each dialogue per block. The order of snippet types for each dialogue was counter-balanced across dialogues so that the original and the modified target were presented first an equal number of times.

In total, 144 snippets were presented. Due to possible similarities between different modifications of a dialogue, participants were instructed to pay close attention to the exact wording and to decide individually for each trial, independent of prior decisions for similar snippets. When participants found a snippet more likely to be old (4–6 on the rating scale), they saw a follow-up question on when they first had heard it, in Session 1 or 2. This response was used as a measure of source memory.

After the recognition test, a cued recall task was conducted. Participants were cued with the head of the dialogue and were instructed to continue the following phrase (i.e., the target) word by word in the exact phrasing as they remembered it from the dialogue. They gave their response orally and were recorded during the task. These responses were later transcribed and then used for analysis. In the case that participants remembered two versions, they had the possibility to report both. However, they were instructed only to report a second version if they clearly remembered it and were asked not to guess. Since the cued recall task took place after the recognition test, it cannot be ruled out that it was influenced by the preceding test, where participants were confronted with much of the material. Due to this limitation, we report this analysis in the supplement. The cued recall results are largely in line with the recognition results presented here.

1.3.5. Difference rating

In the last session, which took place one day after the memory test, participants rated the degree of difference between the original (Sessions 1 and 3) and the modified version (Session 2) of each changing dialogue. To this end, they first listened to the modified dialogue and then to its original version. Subsequently, they were asked how much the two dialogues differ from each other overall and gave their rating on a scale from 1 = very small difference to 7 = very large difference. Additionally, if participants had an autobiographical memory that this dialogue reminded them of, they were asked to report this memory. These memory reports are not analyzed in this paper. After this session, participants were debriefed on the purpose of the experiment, thanked for their effort, and compensated.

1.4. MRI data acquisition and preprocessing

Magnetic resonance imaging was conducted with a 3-Tesla Siemens Magnetom Prisma MR tomograph (Siemens) using a 20-channel head coil. Participants lay supine on the scanner bed and were lightly fixated using form-fitting cushions to minimize movement. The right index and middle finger were placed on the two appropriate buttons of a response box. To attenuate scanner noise, participants were provided with earplugs. The dialogues were presented over headphones. Instructions and attention checks were presented via a screen that participants saw through a mirror mounted on the head coil.

Before starting the experiment, high-resolution T1-weighted anatomical images were obtained using a 3-D magnetization prepared rapid gradient echo sequence (192 slices, slice thickness = 1 mm, repetition time = 2140 ms, echo time = 2.28 ms, flip angle = 8°, field of view = 256 × 256 mm^2). Functional images of the whole brain were acquired in interleaved order along the AC–PC plane using a gradient-echo EPI sequence to measure BOLD contrast (scanning parameters: 33 slices, slice thickness = 3 mm, repetition time = 2000 ms, echo time = 30 ms, flip angle = 90°, FoV = 192 × 192 mm^2). Preprocessing of the imaging data was conducted with SPM12 (Wellcome Trust) implemented in Matlab (Version R2022a, MathWorks Inc.). Data was preprocessed by slice time correction to the middle slice, movement correction and realignment to the mean image, co-registration of the individual structural scans to the mean functional image, normalization of functional and structural images into the standard MNI space (Montreal Neurological Institute, Montreal, QC, Canada) on the basis of segmentation parameters, and spatial smoothing using a Gaussian kernel of full-width at half maximum (FWHM) of 8 mm. Additionally, a 128 s high-pass temporal filter was applied.

1.5. Statistical data analysis

1.5.1. fMRI data analysis

We used a general linear model (GLM) for serially autocorrelated observations (Friston et al., 1994; Worsley and Friston, 1995) and convolved regressors with the canonical hemodynamic function. Surface and gist modifications were modeled as epochs, from modification onset until the end of the dialogue. To each of these regressors, a parametric modulator was added that accounted for the difference ratings between original and modification that were collected individually from each participant. The difference ratings were mean-centered within each participant (Mumford et al., 2015). Since we expected moderate PEs for either high surface or low gist changes, we tested these two possibilities separately: first, by examining the parametric effect of difference ratings for surface modifications (stronger vs. weaker surface differences); and second, by examining the parametric effect of difference ratings for gist modifications (weaker vs. stronger gist changes). For the latter, we inverted the difference ratings for the gist modifications to specifically investigate the effects of weaker (potentially moderate) gist PEs compared to stronger gist PEs, using a parametric regressor. Please note that we conducted a two-sample *t*-test to confirm that gist changes in the 50 % percentile ($M = 2.75$, $SD = 1.06$) were on average perceived as larger by the participants than surface changes in the 50 % percentile ($M = 1.53$, $SD = 0.50$), $t(595) = 18.31$, $p < .001$. This supports our idea that while low surface changes might elicit weak PEs, moderate PEs may be elicited by high surface or low gist changes.

Unchanging dialogues and novels were modeled as full dialogues from onset to end. Two additional regressors modeled the 72 dialogues and 72 null events as epochs, containing their full presentation time (21 s - 34 s and 5.5 s −7 s, respectively). Responding to the attention check was modeled as events with onset on the button press. Six subject-specific rigid body transformations obtained from realignment were included as regressors of nuisance. In total, the GLM comprised 15 regressors.

1.5.1.1. Whole-brain analysis. On the first level, we first calculated the unchanging > novel contrast to ensure that participants had actually encoded the dialogues. Then, we calculated *t*-contrasts for surface > unchanging, gist > unchanging and gist > surface to analyze brain activity in response to the different types of modification. Furthermore, we calculated the parametric contrasts for the surface and gist difference ratings, to further specify the effect of subjective PE strength. On the second level, group analyses were conducted with one-sample *t*-tests across participants. We applied false discovery rate (FDR) correction to the resulting *t*-maps with a threshold of $p < .005$ (voxel-wise). Only clusters with a minimum number of 5 voxels were reported. Significant clusters were visualized using MRIcroGL (Version 1.2.20200331, McCausland Center for Brain Imaging, University of South Carolina).

1.5.1.2. ROI analysis. To investigate the brain regions involved in the processing of PEs more closely, we performed ROI analyses. Two ROI analyses were conducted in the hippocampal area regarding the degree of modification (operationalized by the individual difference ratings provided by each participant for each dialogue), one for the parametric surface and one for the inverted parametric gist modulation. Additionally, we looked at effects of PE type in this area, comparing gist against surface. We used anatomical ROIs of the hippocampus (HC) and the parahippocampal cortex (PHC) and extracted the mean estimate values for the parametric modulations and the gist vs. surface contrast using the MarsBar Toolbox (Brett et al., 2002). Anatomical masks were created by parcellating the MNI brain according to the Desikan-Killiany atlas (Desikan et al., 2006) using FreeSurfer (Fischl, 2012). Contrast estimates were tested with one-sample *t*-tests for significant activation within the ROIs. *p*-values obtained from the *t*-tests were FDR corrected for multiple comparisons within each ROI/analysis.

1.5.1.3. Brain activity predictive of memory outcomes. To deepen our understanding of how the brain responds to prediction errors, we conducted an exploratory trial-by-trial analysis that examined the relationship between neural activation and memory outcomes, focusing on effects that may otherwise be obscured by averaging activation. This analysis focused on the key ROIs (HC, PHC, IFG) to assess their roles in PE detection and memory modification. We obtained the single trial estimates by modeling each dialogue as a separate regressor of a GLM. Then, we extracted the betas from the ROIs and used these single trial activations to predict the recognition performance of original and modification using one linear mixed model per ROI. The models contained the trial-by-trial activation in the corresponding ROI, the difference rating, and the PE type (surface/gist), as well as all interactions as fixed effects. Participants and dialogues were modeled as random intercepts.

1.5.2. Behavioral data analysis

The analysis of the behavioral data was conducted with RStudio (R Core Team, 2024). After confirming that overall attention and memory performance were good, we tested for the stability of the original versions and the learning of the new versions after PEs of different strengths and types. Retention of the original and new learning were operationalized as weighted accuracy for the original or modified targets in the recognition test, respectively. For the weighted accuracy measure (Boeltzig et al., 2025), we coded wrong answers as 0 and correct answers as 1–3, according to participants' confidence judgements. PE strength was operationalized as the difference rating provided by each participant individually, and PE type as surface or gist changes. We used two separate linear mixed models, which were fitted using the *lme4* package (Bates et al., 2015) and tested using the *lmerTest* package (Kuznetsova et al., 2017). Participants and dialogues were modeled as random intercepts. The difference ratings were *z*-standardized within each participant to compensate for potential inter-individual differences of scale use.

Lastly, it was investigated whether moderate PEs lead to more source confusion. To answer this question a logistic regression model was fitted in which modification type (surface, gist) and difference ratings were used to predict source memory accuracy. In this analysis, only trials in which participants correctly recognized the snippet were used.

2. Results

2.1. fMRI results

2.1.1. Whole-brain analysis

As a first step, in order to validate that participants had in fact encoded the original versions, we calculated the unchanging > novels contrast and found significant activation in the PCC, precuneus and orbitofrontal regions. This finding indicates that the dialogues were successfully encoded in Session 1 and were not perceived as novel by the brain.

To investigate which brain regions respond specifically to surface or gist PEs and to confirm that surface and gist PEs not only cause stronger but distinct activations, we computed the contrasts surface > unchanging and gist > surface. In contrast to dialogues that were not modified, surface modifications elicited stronger activations in the left

IFG, the left premotor cortex (PMC), the left pre-SMA, the right superior temporal sulcus (STS) and bilaterally in the cerebellum (Table 2, Fig. 3a). This significant activation is evidence that the dialogue surface was encoded and that the deviations on the language level were in fact detected by the brain. Gist modifications in contrast to surface modifications elicited similar, but overall stronger activations in the right and left STS, the IFG and in the cerebellum. Additionally, we found activations in the TPJ, the medial and lateral superior frontal gyrus (BA 8 and 9), precuneus and posterior cingulate cortex (PCC), the cerebellum, habenula, putamen and hippocampus (Table 2, Fig. 3b).

To examine whether the gist > surface contrast merely reflects differences in PE strength, we computed a model without distinguishing between surface and gist modifications and instead estimated the parametric modulator of the difference ratings for all modifications combined. This combined contrast did not reach significance, suggesting that the gist > surface contrast captures additional factors beyond PE strength alone.

2.1.2. *ROI analysis*

To investigate the role of the hippocampal complex in the processing of different modifications more closely, we performed ROI analyses in the hippocampus and parahippocampal cortex (Fig. 4). Gist in contrast to surface modifications elicited a significant activation in the hippocampal, $t(42) = 2.43$, $p = .020$, but not in the parahippocampal cortex, $t(42) = 1.12$, $p = .269$. Interestingly, in the parahippocampal gyrus we found significant activation for the inverted parametric gist modulation, $t(42) = 2.08$, $p = .044$, with higher parahippocampal activation for smaller gist changes. The corresponding activation in the hippocampus did not reach significance, $t(42) = 1.56$, $p = .127$. For the parametric surface modifications, we found no effect in either of the two regions.

Table 2
Whole-brain activation for surface and gist modifications at FDR $p < .005$ (voxel level).

			MNI Coordinates			
Localization	***H***	***Cluster***	***X***	***Y***	***Z***	***t***
Surface > Unchanging (FDR corrected at < 0.005)						
Cerebellum Crus2	L	63	−18	−82	−37	6.10
Cerebellum Crus2	R	47	21	−82	−34	5.57
Superior Precentral Gyrus (dPMC)	L	45	−39	−1	50	5.72
preSMA (BA6)	L	25	−6	2	56	5.25
Temporal pole	R	16	45	14	−31	5.21
IFG (BA 47)	L	8	−48	29	−4	4.96
Gist > Surface (FDR corrected at < 0.005)						
STS/MTG	L	2091	−57	−31	−4	7.55
IFG comprising BA 44 / 45 / 47	L	l.m.	−51	35	−10	6.21
TPJ	L	l.m.	−45	−58	35	6.15
Temporal pole	L	l.m.	−51	−1	−28	6.13
STS/MTG	R	2196	60	−34	−4	6.26
Temporal pole (BA21)	R	l.m.	54	5	−31	6.11
Superior Frontal Gyrus (BA9)	R	1724	9	32	56	6.99
Superior Frontal Gyrus (BA8)	R	l.m.	9	23	65	6.84
Superior Frontal Gyrus (BA9)	L	l.m.	−6	56	32	6.27
Cerebellum Crus2	L	581	−18	−85	−40	6.93
Cerebellum Crus 1	R	311	24	−79	−31	6.42
Precuneus (posterior)	L	499	−3	−61	35	6.46
PCC	L	l.m.	−3	−49	32	6.24
PCC	R	l.m.	9	−52	35	5.55
Cerebellum	L	126	−6	−58	−40	6.24
Cerebellum	R	l.m.	6	−55	−37	5.34
Putamen	R	78	30	−13	−7	4.94
Habenula	*R + L*	81	0	−25	5	4.88
Medial Front-Orbital Gyrus	R	52	6	47	−16	4.68
Medial Front-Orbital Gyrus	L	l.m.	−6	50	−16	3.91
Hippocampus	L	5	−24	−28	−7	3.46

Note. Only clusters with a minimum extent of 5 voxels are reported. *H* = hemisphere; MNI = Montreal Neurological Institute; *L* = left; *R* = right; BA = Brodmann's area; l.m. = local maximum.

2.1.3. *Brain activity predictive of memory outcomes*

To better understand how neural activation relates to memory outcomes and to explore whether the activations in the key ROIs signal involvement in PE detection or memory modification, we related trial-by-trial activation in the hippocampus, parahippocampal gyrus and IFG to the recognition results. Single trial activation in the hippocampus did not have a significant effect on the recognition of the original, $\beta = 0.07$, 95 % CI [−0.02, 0.16], $SE = 0.05$, $t(961) = 1.39$, $p = .164$, nor the modification, $\beta = 0.02$, 95 % CI [−0.09, 0.13], $SE = 0.06$, $t(976) = 0.36$, $p = .718$. The interaction between hippocampal activation and modification type did not reach significance, $\beta = -0.13$, 95 % CI [−0.27, 0.02], $SE = 0.07$, $t(957) = -1.75$, $p = .081$. Only descriptively, more hippocampal activation led to worse memory for the original version after gist modifications, while on the surface level there was no influence of hippocampal activation. Similarly, there was no effect of single trial activation in the parahippocampal gyrus on the retention of the original, $\beta = 0.04$, 95 % CI [−0.07, 0.14], $SE = 0.06$, $t(960) = 0.64$, $p = .521$, or modification learning, $\beta = -0.02$, 95 % CI [−0.14, 0.10], $SE = 0.06$, $t(977) = -0.35$, $p = .728$. In the IFG, on the other hand, trial-by-trial activation predicted significantly the recognition of the original, $\beta = 0.11$, 95 % CI [0.01, 0.21], $SE = 0.05$, $t(967) = 2.14$, $p = .033$, and marginally of the modification, $\beta = 0.09$, 95 % CI [0.02, 0.17], $SE = 0.04$, $t(961) = 2.43$, $p = .015$.

The association between activation in the IFG and memory outcomes was confirmed in another exploratory, univariate model, which used the weighted accuracy measure for the originals as a parametric modulator. The model revealed significant activations in the right IFG and the bilateral pSTS.

2.2. *Behavioral analysis*

2.2.1. *Cover task*

Overall, performance for the cover task in the fMRI sessions was good, with participants responding correctly on 78.03 % to 97.72 % ($M = 90.15$ %, $SD = 4.74$ %) of trials across all sessions. Therefore, it can be assumed that participants listened to the dialogues attentively.

2.2.2. *Memory for unchanging vs. changing dialogues*

To assess whether the modifications had an effect on memory, recognition of changing dialogue targets after a surface or gist modification was compared to unchanging dialogue targets. This was done using a linear mixed model with dialogue type (unchanging, changing – original target after surface modification, changing – original target after gist modification) as the independent variable and subject and dialogue as random intercepts on the second level. The model showed a significant lower recognition accuracy of original targets after hearing a gist modification ($M = 0.81$, $SD = 0.15$) compared to no modification ($M = 0.88$, $SD = 0.16$), $\beta = 0.07$, 95 % CI [0.01, 0.13], $SE = 0.03$, $t(1206) = 2.73$, $p = .018$. Targets after surface modifications ($M = 0.85$, $SD = 0.14$) did not differ significantly from unchanging dialogues, $\beta = 0.04$, 95 % CI [−0.02, 0.10], $SE = 0.03$, $t(1206) = 1.44$, $p = .320$, or from targets after gist modifications, $\beta = 0.03$, 95 % CI [−0.02, 0.08], $SE = 0.02$, $t(1206) = 1.58$, $p = .254$. This result is a first indication that PEs induced by gist modifications had a weakening effect on original memories.

2.2.3. *Memory for the original version*

In a next step, we tested whether modification type (surface or gist) and the perceived degree of modification (assessed through the z-standardized difference ratings) had an effect on recognition of the original targets. The linear mixed model showed a significant effect of modification type, $\beta = 0.17$, 95 % CI [0.03, 0.31], $SE = 0.07$, $t(960) = 2.32$, $p = .021$, and modification degree, $\beta = 0.11$, 95 % CI [0.01, 0.20], $SE = 0.05$, $t(978) = 2.23$, $p = .026$, on weighted accuracy. As shown in the first panel of Fig. 5A, memory for original targets on average was worse after gist modifications ($M = 1.93$, $SD = 1.16$) than after surface modifications ($M = 2.04$, $SD = 1.09$) and worse after weaker than after stronger

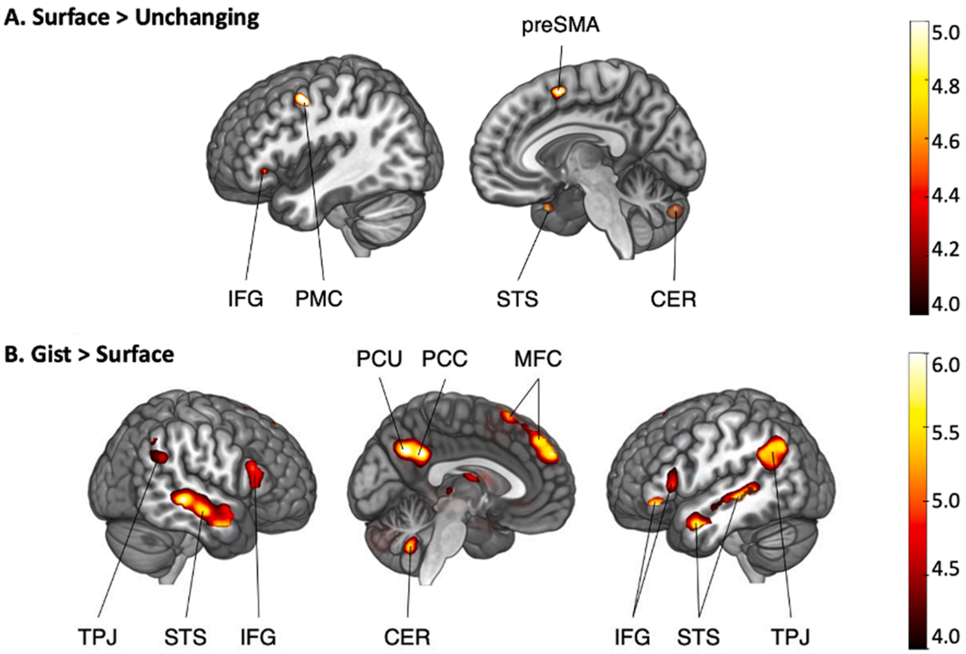


Fig. 3. Whole-brain activation for surface and gist modifications.

Note. Whole-brain activations for surface and gist modifications. **A.** FDR-corrected ($p < .005$) t-map for the surface > unchanging contrast. Activations were found in the left inferior frontal gyrus (IFG), left dorsal premotor cortex (PMC), left pre-supplementary motor area (preSMA), right anterior superior temporal sulcus (STS) and in the cerebellum. **B.** For the gist > surface contrast activations were found in the ventral precuneus (PCU), posterior cingulate cortex (PCC), medial frontal cortex (MFC), temporo-parietal junction (TPJ), superior temporal sulcus (STS), inferior frontal gyrus (IFG) and in the cerebellum (CER). Sagittal cuts were made at $x = -50$ (left), $x = -4$ (middle) and $x = 57$ (right). Please note that for a more conclusive attribution of activation to anatomical regions, we show the gist > surface contrast at FDR $p < .001$ (voxel level).

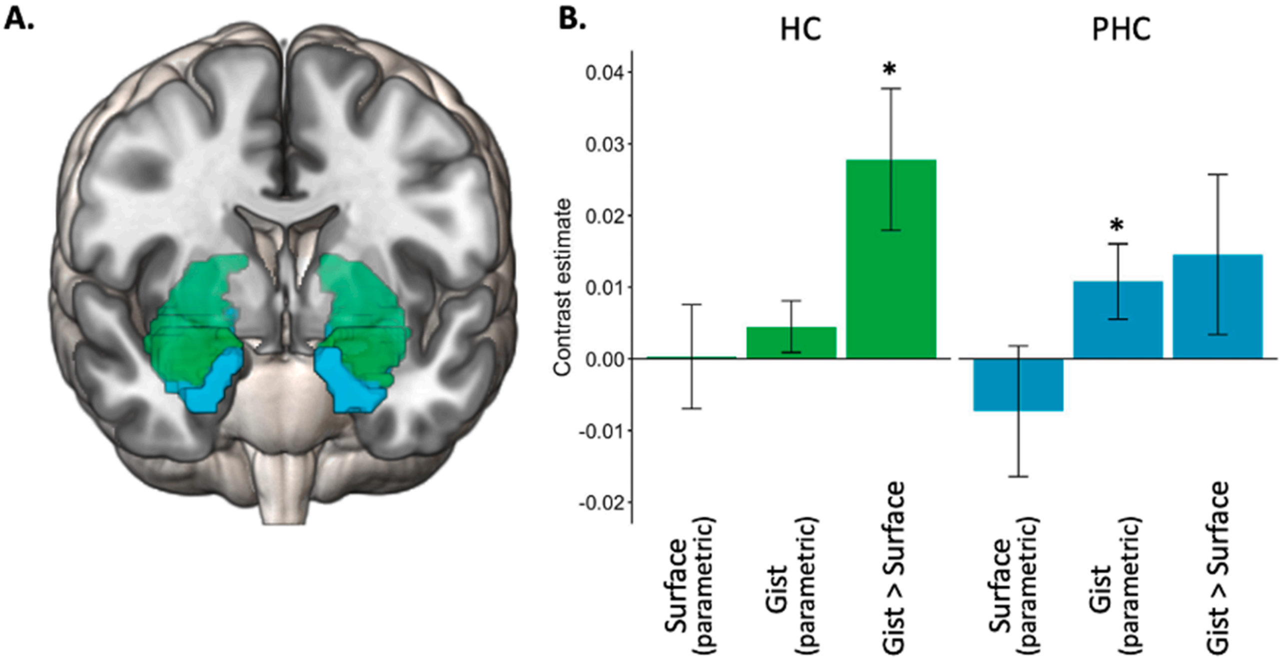


Fig. 4. ROI analysis localization and results.

Note. ROI analysis of the hippocampus (HC) and parahippocampal cortex (PHC). The error bars represent the standard error. **A.** Anatomical ROIs of the HC and PHC. **B.** Contrast estimates were extracted from the parametric surface, the inverted parametric gist modulations and the gist vs. surface contrast. One-sample t-tests revealed significant activation in the PHC for weaker gist changes and in the HC for gist vs. surface modifications (*$p < .05$).

modifications. The interaction between modification type and degree was not significant, $\beta = -0.10$, 95 % CI [−0.26, 0.05], $SE = 0.08$, $t(985) = -1.37$, $p = .170$.

2.2.4. *Memory for the modification*

Similarly, we investigated the impact of modification type and degree on the recognition of modifications, reflecting new learning processes (Fig. 5A, second panel). All predictors, modification type, $\beta =$

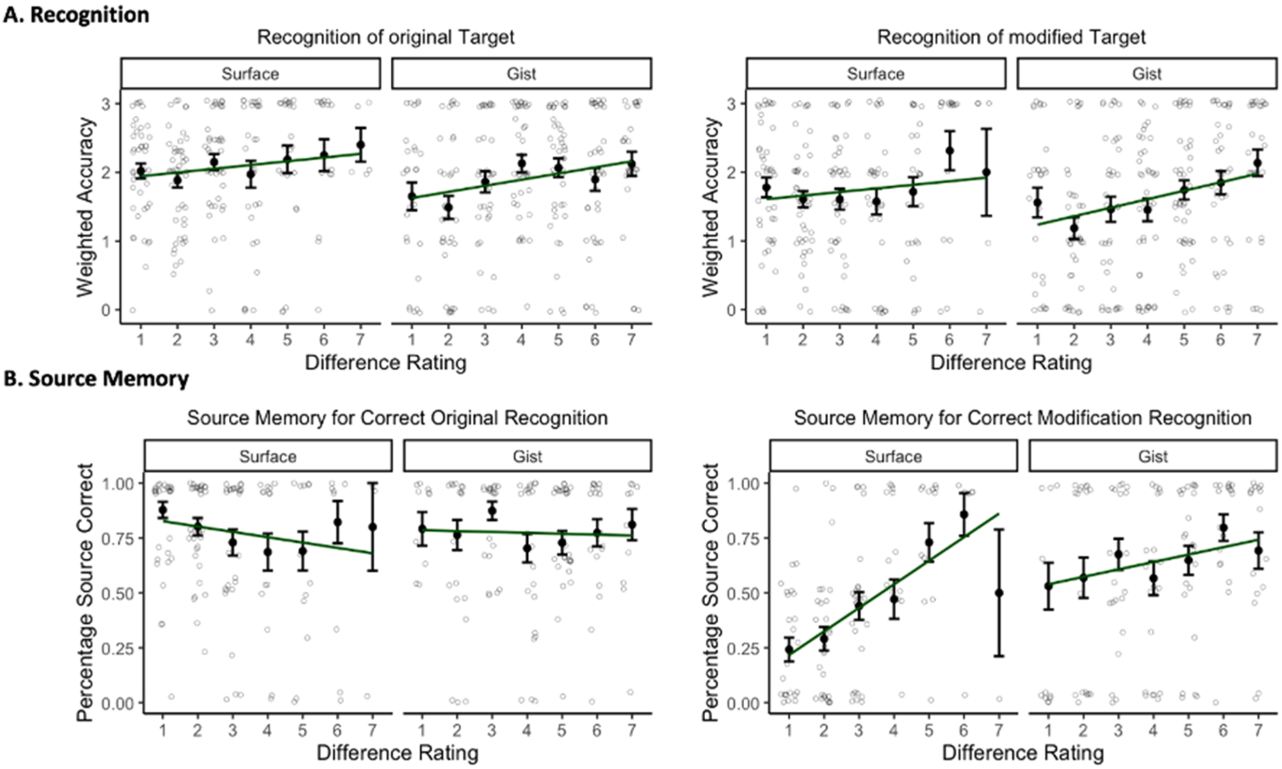


Fig. 5. Results recognition & source memory tests.

Note. The influence of modification type (surface, gist) and modification degree (assessed by the difference rating) on recognition and source memory results. The error bars represent the standard error. **A.** Weighted accuracy for the original targets (left) and modified targets (right). **B.** Percentage of correct source attribution (Session 1 or Session 2) of the targets.

0.17, 95 % CI [0.01, 0.33], $SE = 0.08$, $t(975) = 2.09$, $p = .037$, modification degree, $\beta = 0.23$, 95 % CI [0.13, 0.34], $SE = 0.05$, $t(992) = 4.29$, $p < .001$, and the interaction, $\beta = -0.18$, 95 % CI [−0.34, −0.01], $SE = 0.09$, $t(1000) = -2.07$, $p = .038$, exhibited significant effects on weighted accuracy. In parallel with the findings above, gist modifications and weaker modifications were associated with reduced learning of the new versions. Notably, new learning was particularly low for weak gist modifications, compared to strong gist modifications and surface modifications.

2.2.5. Source memory

Lastly, we tested the effect of the difference rating on source memory. Source memory was tested by asking participants on which day they had first heard a given statement if they said they recognized the statement. We only used trials in which participants correctly recognized the snippet as heard during the experiment. To that end, a logistic regression model was used that predicted correct source attribution based on modification type and degree. For the original version, we found no significant main effects for modification type, $\beta = -0.04$, 95 % CI [−0.43, 0.35], $SE = 0.20$, $z = -0.22$, $p = .825$, or modification degree, $\beta = -0.00$, 95 % CI [−0.28, 0.28], $SE = 0.14$, $z = -0.00$, $p = .999$, but a significant interaction of the two, $\beta = -0.43$, 95 % CI [−0.85, −0.01], $SE = 0.21$, $z = -2.02$, $p = 0.043$ (Fig. 5B, first panel). For the modification, the modification type, $\beta = -0.96$, 95 % CI [−1.36, −0.57], $SE = 0.20$, $z = -4.84$, $p < .001$, and the interaction, $\beta = 0.70$, 95 % CI [0.29, 1.12], $SE = 0.21$, $z = 3.33$, $p < .001$, reached significance, but not the modification degree, $\beta = 0.15$, 95 % CI [−0.11, 0.41], $SE = 0.13$, $z = 1.14$, $p = .254$. Here, source attribution was generally better for gist than for surface targets (Fig. 5B, second panel). Especially for surface changes, the difference rating had a strong effect on correct source recognition, with small surface differences being most likely to be wrongly attributed to the first experimental session, when the dialogue actually stemmed from the second one.

3. Discussion

The goal of the present study was to investigate how different types and strengths of mnemonic prediction errors (PEs), particularly moderate PEs, impact brain activation and memory outcomes. To this end, naturalistic dialogues were first encoded and later modified to varying degrees either in their gist or in their surface form. As expected, we found a robust activation in the IFG to all PEs. Gist modifications produced additional activations in the episodic memory network, including the hippocampus. While the degree of modification did not modulate the brain responses to surface changes, for gist changes it modulated both brain activations and memory outcomes, with parametrically weaker gist changes leading to worse recognition as well as significant parahippocampal activation. This crucial finding reveals that moderate PEs are particularly influential in altering existing memories while impeding the encoding of new, distinct episodes, aligning with the assumptions of the Latent Cause Theory (Gershman et al., 2017).

Our most important finding is that episodic PEs differed both in the regions that they activated and in the extent to which they did so as well as behaviorally in retention of the original stimuli and learning of the new input. We thereby extend previous literature that investigated episodic PEs mainly in binary categories (Siestrup and Schubotz, 2023; Sinclair et al., 2021) by showing that PE strength is an important factor that can explain different memory outcomes in previous studies. Dialogues that we engage in or witness in our daily life are a central example of episodic memories (Atir et al., 2022). These interactions are a fundamental part of our daily experiences, encapsulating the context, emotions, and sequences that define episodic memory (Fischer et al., 2015; Kumar et al., 2023). Importantly, our study represents the first approach to quantify PEs using naturalistic, verbal stimuli in an fMRI paradigm. This is in contrast to previous methods that relied on more artificial constructs such as word-picture pairs (Long et al., 2016) or scene-object pairs (Ortiz-Tudela et al., 2023). By utilizing dialogues, our research captures the complexity and richness of real-life memory processing, providing a more accurate reflection of how PEs manifest in everyday cognitive functions.

In accordance with our hypotheses, we found a significant activation in the left IFG for surface modifications in comparisons to unmodified dialogues. Several studies have found activation in the IFG in response to unexpected information across different modalities, which indicates a modality-independent process (Bubic et al., 2009; El-Sourani et al., 2020; Gläscher et al., 2010; Jainta et al., 2024; Wurm and Schubotz, 2012). In the acoustic domain, activation of the left IFG has been shown in an auditory oddball paradigm during the detection of acoustically deviant features (Hofmann-Shen et al., 2020), but no study to date has investigated the response of the IFG to PEs using complex, verbal stimuli. Our results therefore extend these findings by showing that the IFG also responds to episodic prediction violations on a verbatim level for naturalistic speech. Whether the subjects were aware of this process or not, and whether this was decisive for IFG activation, we cannot decide on the basis of the data and the task. In previous studies, however, we also used cover tasks and often found that IFG activations in response to PEs were independent of the conscious detection of expectation deviations (Jainta et al., 2022; Siestrup et al., 2022, 2023).

Interestingly, for surface modifications we found additional involvement of the preSMA, dorsal premotor cortex (dPMC), and the cerebellum. It has long been understood that the higher motor system, beyond its role in motor control, is also involved in processing sequentially structured stimuli and events (Schubotz, 2007), and relatedly, the motor system might play a role in speech perception (Liberman and Whalen, 2000; Whalen, 2019). Specifically, the cerebellum and its projections to the pre-SMA have been found to be important for top-down predictions of upcoming verbal information (for a review see Mariën et al., 2013) while the pre-SMA has been linked to ambiguity resolution, attentional switching and context integration in the context of language processing (Hertrich et al., 2016). It is therefore possible that the activation in this network reflects the processing of the unexpected speech input on a verbatim level in the present study. However, future studies will have to further investigate the role of the higher motor system in verbal predictions. Another interesting finding was that the right anterior temporal pole was susceptible to surface modifications. Previously, the anterior STS has been found to be activated in response to surprisal in naturalistic, spoken narratives (Willems et al., 2016).

These results indicate the high sensitivity of several brain areas to the minor deviation in previously encoded episodes, whose change consisted only in the protagonists' choice of words, without affecting the meaning of the statement. In contrast to gist modifications, however, the degree of such surface changes had no effect on memory performance. Based on the Latent Cause Theory (Gershman et al., 2017), we aimed to generate weak, moderate, and strong PEs. Our findings show that -even maximal- surface modification had no effect on the retention of the original, which is in line with the view that surface modification of the dialogues only generated weak PEs. A potential interpretation of these results is that surface modifications were still inferred to have the same underlying cause and the deviations were not significant enough to cause memory changes according to the framework by Gershman and colleagues (2017).

This looked quite different for modifications that not only changed the phrasing, but also the content of one or several statements. As expected, for gist modifications, the IFG activation was more pronounced than for surface modifications, extending over BA 44, 45, and 47. Since the surface activations showed a similar expansion subthreshold, we interpreted this as a quantitative increase in activation instead of functionally different processes. According to our hypothesis, we therefore saw that the IFG not only responds to modifications in phrasing but also in content. In previous studies, the pars triangularis of IFG, BA 45, has been suggested to have a role in the selection of competing memories (Badre et al., 2005; Martin, 2007; Moss et al., 2005; Schlichting et al., 2015), while the anterior part, BA 47, has been related to updating of predictions following unexpected information (El-Sourani et al., 2020). Hence, the extensive activation in the IFG might reflect different sub-processes, but more research is needed to disentangle the roles of different IFG subfields in response to prediction violations.

Consistent with our hypotheses, we did not only find a quantitative increase in activation of certain brain areas but also distinct activation patterns when comparing gist with surface modifications. Specifically, we saw involvement of different parts of the episodic memory network (Rugg and Vilberg, 2013). This network is known to be active during the encoding and retrieval of episodic memories, which is why we hypothesized to see activation in associated regions for PEs that would trigger an updating at the highest level of memory representations.

Especially activity in the hippocampus has repeatedly been found in response to PEs (Long et al., 2016; Sinclair et al., 2021). In contrast to a previous study (Long et al., 2016), however, we did not see a response to modification degree in the hippocampus proper, but instead in the parahippocampal gyrus. Notably, weaker gist modifications led to a stronger parahippocampal activation than did strong gist modifications, as was shown by the significant effect of the parametric modulator. Correspondingly, weak gist modifications led to a less accurate recognition of the original and the modified version than strong gist modifications. A possible interpretation according to the Latent Cause Theory (Gershman et al., 2017) is that for weak gist modifications the same latent cause was inferred, but it was updated with the new information from the modification leading to an integrated memory representation. Due to the specificity of the parahippocampal activation only for weak gist changes we assume that this region may be involved in this specific updating process. Previous research has found parahippocampal activation for content in comparison to structure changes (Siestrup et al., 2022) or in response to false alarms (Siestrup et al., 2023), which further supports this interpretation.

Furthermore, the parahippocampal cortex has been proposed to play a central role in the representation of contextual information when recalling a memory (Diana et al., 2007). Additionally to spatial information (Davachi, 2006; Epstein and Kanwisher, 1998; Suzuki et al., 2005), the parahippocampal gyrus has been found to provide a more global representation of the encoded target, including semantic and gist information (Diana, 2016). In our study, the presented dialogues not only consisted of the interaction between two protagonists, but also contained a background sound that implied a certain setting of the encounter. Furthermore, participants were instructed to imagine the scene as if they were witnessing it directly, which might have helped to establish a specific situational context for the dialogue (e.g., “in the car”, “at the supermarket”). The speakers, the background sound and therefore the implied setting of the dialogue always remained unchanged across modifications, such that this activation might actually reflect the updating of local information (i.e., what was said) in the global context of the scenario instead of just a simple updating of information in the conversation. However, more research is needed to evaluate the role of the parahippocampal region in memory updating.

As a first step towards a better understanding of the role of the parahippocampal gyrus, but also of the hippocampus and IFG in PE detection and memory updating, we conducted an exploratory analysis, relating the single trial activation in our ROIs to memory outcomes. The results showed that neither activation in the hippocampus nor in the parahippocampal cortex significantly predicted recognition performance, while activation in the IFG predicted recognition of both the original and the modified versions. In a previous study that used verbal stimuli to investigate the sudden comprehension of novel meaningful relationships, the hippocampus responded to the surprising event of sudden comprehension (Kizilirmak et al., 2019). However, learning of the suddenly comprehended solution was only associated with increased cortical but not hippocampal activation. Hence, it is possible that the activation in the hippocampal area serves as a teaching signal to the cortex, in our study the IFG. Accordingly, activity in the IFG would reflect learning from PEs rather than the PE signal itself. This assumption is also supported by a previous study investigating deviations from expectations in action observation. The study found that the IFG responds less to objects that either fit very well or not at all within an action context, compared to moderately fitting objects, which were particularly informative for the upcoming action. The authors interpreted this finding to suggest that PE-related activity in the IFG indicates the learning-from-PE process rather than the PE signal itself (El-Sourani et al., 2020).

Strong gist changes, which were accompanied by episodic memory network activity, caused no memory changes for the original but instead led to accurate recognition of the old and the new version as well as good source memory. In accordance with the Latent Cause Theory (Gershman et al., 2017), we take strong gist changes to have elicited strong PEs, which led to a new latent cause being inferred and a new and separate memory being encoded. Our results therefore provide further evidence that the episodic memory network is relevant for both updating of existing memories and new learning through episodic PEs and extend insights into the mechanisms of memory changes.

Another notable point that emerges from the data is a substantial activation in the STS. The middle part of the left STS is thought to be specialized for speech perception and language processing (Liebenthal et al., 2014; Specht and Wigglesworth, 2018), but more anterior parts have also been related to the theory of mind network (Carrington and Bailey, 2009; Fedorenko et al., 2024). Together with the activation in the TPJ, another important part of the theory of mind network (Carrington and Bailey, 2009; Fedorenko et al., 2024), this activation might reflect the processing of the changed goals and beliefs of the speakers. However, due to the extensive expansion of the activation and considerable overlap between episodic memory network and theory of mind network, future studies will have to further investigate the role of the STS in the context of prediction violations in the context of remembered social encounters.

All in all, our findings are consistent with the Latent Cause Theory (Gershman et al., 2017) and provide an important extension of previous research by testing its assumptions on a more fine-grained, quantitative scale for PE strength. Our results also seem to be in line with another, related account, the so-called Non-monotonic Plasticity Hypothesis (Ritvo et al., 2019). According to this framework, especially a medium reactivation of the original memory leads to its destabilization, while weak and strong reactivations do not. In the present study, we saw a weakening of the original memory specifically after weak gist changes, which had enough similarity with the original version to reactivate parts, but not all of it. Our findings are therefore consistent with both interpretations and highlight the important role of PE strength for understanding brain activations and memory outcomes.

4. Limitations & future research

To maintain high external validity not only in the material, but also in the experimental manipulation, PEs were induced by modifying stimuli, and thus directly replacing expected input with new statements. This approach has been taken previously (e.g., Greve et al., 2017; Kim et al., 2014), also with highly naturalistic temporally unfolding stimuli (e.g., Siestrup and Schubotz, 2023). Other studies, however, have first induced a PE by stopping the stimulus or action, and only later providing new information (e.g., Forcato et al., 2013; Sinclair and Barense, 2018). These studies have argued that PEs induce lability to a memory trace, which can then lead to vulnerability to interference from new and similar input, while PEs may even strengthen memories when no new information is added. It is important to stress that we had no condition of PE without subsequent new information, given that the PE was induced by new information. Future studies should therefore further focus on the effects of different PE inductions on memory outcomes and the role of reconsolidation in this process.

An effect of our naturalistic stimuli was a limit of the possible number of trials within a single fMRI session. Applying the findings from previous studies to naturalistic material is crucial (Maguire and Harris, 2022), especially in the field of predictions, which are dynamically updated and adjusted continuously and therefore benefit from stimuli that are extended in time such as narratives (e.g., Siestrup and Schubotz, 2023) or dialogues. However, as a consequence the unchanging dialogues and the novels were required to be modeled in their entirety, including the dialogues' heads, to achieve optimal power from the limited data. This low trial number reflects a necessary compromise to prioritize the power of the modification trials, which were central to our research question, over the unchanging and novel trials, which primarily served as baseline conditions. While a control analysis balancing modeling of unchanging and modified dialogues confirmed our findings, the reduced number of time points in the unchanging dialogues impacted power. In future studies, it would be interesting to use a more balanced number of unchanging and modified trials to address this limitation. Similarly, to keep the encoding situations as naturalistic as possible, we refrained from collecting a measure of PE size immediately after presenting the modified version in Session 2 (Siestrup and Schubotz, 2023; Sinclair and Barense, 2018). During presentation of the modification, a PE emerged based on the (possibly implicit) comparison of a previously encoded episode with the new input while during the rating both versions were immediately presented back to back, and had to be consciously compared. As these are distinct mechanisms, future studies should systematically assess the effect of immediate and delayed difference ratings.

Another limitation of the study could be a potential influence of the assignment of stimuli to conditions. We tried to minimize a potentially confounding influence of the material by carefully counterbalancing the assignment of dialogues to conditions and additionally modeling dialogue-specific influences as a random factor in all analyses.

In the present study, the original dialogues were replayed once more after the modification session, which might have impacted memory outcomes. For example, participants might have consciously compared the two versions, especially after a big change in the modification session, which might have led to the better memory performance for strong PEs. In a previous behavioral study (Boeltzig et al., 2025) with the same material and similar design, there was no such replay of originals, and the test happened one day after the encoding of the modification. In both studies we saw better encoding of the new information after strong PEs and the behavioral study showed a comparable pattern for well-encoded originals with better recognition after strong than moderate PEs. Therefore, while the replay might have enhanced overall recognition of the original, it is improbable that it caused these effects.

5. Conclusion

Our results show that both the type and the strength of PEs, individually and interactively, influence brain activations and related memory outcomes. In our design, changes in the phrasing of a statement elicited weak PEs, which did not have a detrimental effect on memory. Minor changes in the meaning of the utterance led to less accurate memory for the original and modification, while strong changes promoted the encoding of a new, separate episode.

Funding information

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft) under grant number SCHU1439_10-2, project number 397530566. The funder had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

CRediT authorship contribution statement

Nina Liedtke: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Marius Boeltzig:** Writing – review & editing, Software, Project administration, Methodology, Investigation, Conceptualization. **Falko Mecklenbrauck:** Writing – review & editing, Software, Formal analysis. **Sophie Siestrup:** Writing – review & editing, Software, Conceptualization. **Ricarda I. Schubotz:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors report there are no competing interests to declare.

Acknowledgements

We thank Monika Mertens, Lena Puder, Britt Hasslöver, Leon Exeler, Kinan Elachkar and Michelle Bellstedt for their help during data collection and Britt Hasslöver, Tabea Krause, Lana Steuernagel and Michelle Bellstedt for their help during stimulus construction. We also want to thank Michael Borgard for the production of the dialogues. Last, we thank Anoushiravan Zahedi and the rest of the biopsychology group for their helpful comments and for the valuable discussions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2025.121194.

Data availability

The behavioral data is available on OSF (https://osf.io/g54bt/?view_only=640c58331dc74a6eba134b66e2f4d735). Unthresholded statistical maps of all reported fMRI contrasts in the manuscript have been deposited on NeuroVault (https://neurovault.org/collections/KGYTHXXN/). Samples of the stimuli can be provided upon request to the corresponding author.

References

Atir, S., Wald, K.A., Epley, N., 2022. Talking with strangers is surprisingly informative. Proc. Natl. Acad. Sci. 119 (34). https://doi.org/10.1073/pnas.2206992119. Article e2206992119.

Badre, D., Poldrack, R.A., Paré-Blagoev, E.J., Insler, R.Z., Wagner, A.D., 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. Neuron 47 (6), 907–918. https://doi.org/10.1016/j.neuron.2005.07.023.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Soft. 67 (1). https://doi.org/10.18637/jss.v067.i01.

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

Bein, O., Gasser, C., Amer, T., Maril, A., Davachi, L., 2023. Predictions transform memories: How expected versus unexpected events are integrated or separated in memory. Neurosci. Biobehav. Rev. 153. https://doi.org/10.1016/j.neubiorev.2023.105368. Article 105368.

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., Maril, A., 2015. Delineating the effect of semantic congruency on episodic memory: the role of integration and relatedness. PLoS One 10 (2). https://doi.org/10.1371/journal.pone.0115624. Article e0115624.

Bein, O., Plotkin, N.A., Davachi, L., 2021. Mnemonic prediction errors promote detailed memories. Learn. Mem. 28 (11), 422–434. https://doi.org/10.1101/lm.053410.121.

Boeltzig, M., Liedtke, N., Schubotz, R.I., 2025. Prediction errors lead to updating of memories for conversations. Memory 33 (1), 73–83. https://doi.org/10.1080/09658211.2024.2404498.

Brady, T.F., Robinson, M.M., Williams, J.R., Wixted, J.T., 2023. Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. Psychon. Bull. Rev. 30 (2), 421–449. https://doi.org/10.3758/s13423-022-02179-w.

Bransford, J.D., Franks, J.J., 1971. The abstraction of linguistic ideas. Cogn. Psychol. 2 (4), 331–350. https://doi.org/10.1016/0010-0285(71)90019-3.

Brett, M., Anton, J.L., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using an SPM toolbox [abstract]. In: Presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan, 16. *Available on CD-ROM in Neuroimage*Article S497.

Brod, G., Hasselhorn, M., Bunge, S.A., 2018. When generating a prediction boosts learning: the element of surprise. Learn. Instr. 55, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013.

Brod, G., Lindenberger, U., Werkle-Bergner, M., Shing, Y.L., 2015. Differences in the neural signature of remembering schema-congruent and schema-incongruent events. Neuroimage 117, 358–366. https://doi.org/10.1016/j.neuroimage.2015.05.086.

Brod, G., Werkle-Bergner, M., Shing, Y.L., 2013. The influence of prior knowledge on memory: a developmental cognitive neuroscience perspective. Front. Behav. Neurosci. 7. https://doi.org/10.3389/fnbeh.2013.00139. Article 139.

Brown-Schmidt, S., Benjamin, A.S., 2018. How we remember conversation: implications in legal settings. Policy Insights Behav. Brain Sci. 5 (2), 187–194. https://doi.org/10.1177/2372732218786975.

Bubic, A., Schroger, E., Schubotz, R.I., 2009. Violation of expectation: neural correlates reflect bases of prediction. J. Cogn. Neurosci. 21 (1), 155–168. https://doi.org/10.1162/jocn.2009.21013.

Carrington, S.J., Bailey, A.J., 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. Hum. Brain Mapp. 30 (8), 2313–2335. https://doi.org/10.1002/hbm.20671.

Davachi, L., 2006. Item, context and relational episodic encoding in humans. Curr. Opin. Neurobiol. 16 (6), 693–700. https://doi.org/10.1016/j.conb.2006.10.012.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31 (3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021.

Diana, R.A., 2016. Parahippocampal cortex processes the nonspatial context of an event. Cereb. Cort. 27 (3), 1808–1816. https://doi.org/10.1093/cercor/bhw014.

Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn. Sci. 11 (9), 379–386. https://doi.org/10.1016/j.tics.2007.08.001.

Duncan, K., Ketz, N., Inati, S.J., Davachi, L., 2012. Evidence for area CA1 as a match/mismatch detector: a high-resolution fMRI study of the human hippocampus. Hippocampus 22 (3), 389–398. https://doi.org/10.1002/hipo.20933.

El-Sourani, N., Trempler, I., Wurm, M.F., Fink, G.R., Schubotz, R.I., 2020. Predictive impact of contextual objects during action observation: evidence from functional magnetic resonance imaging. J. Cogn. Neurosci. 32 (2), 326–337. https://doi.org/10.1162/jocn_a_01480.

Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392 (6676), 598–601. https://doi.org/10.1038/33402.

Fedorenko, E., Ivanova, A.A., Regev, T.I., 2024. The language network as a natural kind within the broader landscape of the human brain. Nat. Rev. Neurosci. 25 (5), 289–312. https://doi.org/10.1038/s41583-024-00802-4.

Fischer, N.M., Schult, J.C., Steffens, M.C., 2015. Source and destination memory in face-to-face interaction: a multinomial modeling approach. J. Exp. Psychol.: Appl. 21 (2), 195–204. https://doi.org/10.1037/xap0000046.

Fischl, B., 2012. FreeSurfer. Neuroimage 62 (2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021.

Forcato, C., Burgos, V.L., Argibay, P.F., Molina, V.A., Pedreira, M.E., Maldonado, H., 2007. Reconsolidation of declarative memory in humans. Learn. Mem. 14 (4), 295–303. https://doi.org/10.1101/lm.486107.

Forcato, C., Fernandez, R.S., Pedreira, M.E., 2013. The role and dynamic of strengthening in the reconsolidation process in a human declarative memory: what decides the fate of recent and older memories? PLoS One 8 (4), e61688. https://doi.org/10.1371/journal.pone.0061688.

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. Hum. Brain Mapp. 2 (4), 189–210. https://doi.org/10.1002/hbm.460020402.

Friston, K.J., Kiebel, S., 2009. Predictive coding under the free-energy principle. Philos. Trans. R. Soc. B: Biol. Sci. 364 (1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300.

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

Gläscher, J., Daw, N., Dayan, P., O'Doherty, J.P., 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66 (4), 585–595. https://doi.org/10.1016/j.neuron.2010.04.016.

Greve, A., Cooper, E., Kaula, A., Anderson, M.C., Henson, R., 2017. Does prediction error drive one-shot declarative learning? J. Mem. Lang. 94, 149–165. https://doi.org/10.1016/j.jml.2016.11.001.

Greve, A., Cooper, E., Tibon, R., Henson, R.N., 2019. Knowledge is power: prior knowledge aids memory for both congruent and incongruent events, but in different ways. J. Exp. Psychol.: Gen. 148 (2), 325–341. https://doi.org/10.1037/xge0000498.

Henson, R.N., Gagnepain, P., 2010. Predictive, interactive multiple memory systems. Hippocampus 20 (11), 1315–1326. https://doi.org/10.1002/hipo.20857.

Hertrich, I., Dietrich, S., Ackermann, H., 2016. The role of the supplementary motor area for speech and language processing. Neurosci. Biobehav. Rev. 68, 602–610. https://doi.org/10.1016/j.neubiorev.2016.06.030.

Hofmann-Shen, C., Vogel, B.O., Kaffes, M., Rudolph, A., Brown, E.C., Tas, C., Brüne, M., Neuhaus, A.H., 2020. Mapping adaptation, deviance detection, and prediction error in auditory processing. Neuroimage 207. https://doi.org/10.1016/j.neuroimage.2019.116432. Article 116432.

Jainta, B., Siestrup, S., El-Sourani, N., Trempler, I., Wurm, M.F., Werning, M., Cheng, S., Schubotz, R.I., 2022. Seeing what I did (not): cerebral and behavioral effects of agency and perspective on episodic memory re-activation. Front. Behav. Neurosci. 15. https://doi.org/10.3389/fnbeh.2021.793115. Article 793115.

Jainta, B., Zahedi, A., Schubotz, R.I., 2024. Same same, but different: brain areas underlying the learning from repetitive episodic prediction errors. J. Cogn. Neurosci. 36 (9), 1847–1863. https://doi.org/10.1162/jocn_a_02204.

Jeong, W., Chung, C.K., Kim, J.S., 2015. Episodic memory in aspects of large-scale brain networks. Front. Hum. Neurosci. 9. https://doi.org/10.3389/fnhum.2015.00454. Article 454.

Kim, G., Lewis-Peacock, J.A., Norman, K.A., Turk-Browne, N.B., 2014. Pruning of memories by context-based prediction error. Proc. Natl. Acad. Sci. 111 (24), 8997–9002. https://doi.org/10.1073/pnas.1319438111.

Kim, G., Norman, K.A., Turk-Browne, N.B., 2017. Neural differentiation of incorrectly predicted memories. J. Neurosci. 37 (8), 2022–2031. https://doi.org/10.1523/JNEUROSCI.3272-16.2017.

Kizilirmak, J.M., Schott, B.H., Thuerich, H., Sweeney-Reed, C.M., Richter, A., Folta-Schoofs, K., Richardson-Klavehn, A., 2019. Learning of novel semantic relationships via sudden comprehension is associated with a hippocampus-independent network. Conscious. Cogn. 69, 113–132. https://doi.org/10.1016/j.concog.2019.01.005.

Kumar, M., Goldstein, A., Michelmann, S., Zacks, J.M., Hasson, U., Norman, K.A., 2023. Bayesian surprise predicts human event segmentation in story listening. Cogn. Sci. 47 (10). https://doi.org/10.1111/cogs.13343. Article e13343.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. (13), 82. https://doi.org/10.18637/jss.v082.i13.

Liberman, A.M., Whalen, D.H., 2000. On the relation of speech to language. Trends Cogn. Sci. 4 (5), 187–196. https://doi.org/10.1016/S1364-6613(00)01471-6.

Liebenthal, E., Desai, R.H., Humphries, C., Sabri, M., Desai, A., 2014. The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. Front. Neurosci. 8. https://doi.org/10.3389/fnins.2014.00289. Article 289.

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. J. Neurosci. 36 (50), 12677–12687. https://doi.org/10.1523/JNEUROSCI.1850-16.2016.

Maguire, R.R., Harris, J.M., 2022. Turning over a new leaf: differences in search ability across naturalistic leaf litter textures. J. Vis. 22 (14), 4352. https://doi.org/10.1167/jov.22.14.4352.

Mariën, P., Ackermann, H., Adamaszek, M., Barwood, C.H.S., Beaton, A., Desmond, J., De Witte, E., Fawcett, A.J., Hertrich, I., Küper, M., Leggio, M., Marvel, C., Molinari, M., Murdoch, B.E., Nicolson, R.I., Schmahmann, J.D., Stoodley, C.J., Thürling, M., Timmann …, D., Ziegler, W., 2013. Consensus paper: language and the cerebellum: an ongoing enigma. Cerebellum 13, 386–410. https://doi.org/10.1007/s12311-013-0540-5.

Martin, A., 2007. The representation of object concepts in the brain. Annu. Rev. Psychol. 58 (1), 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143.

Moss, H.E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., Tyler, L.K., 2005. Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. Cereb. Cort. 15 (11), 1723–1735. https://doi.org/10.1093/cercor/bhi049.

Mumford, J.A., Poline, J.-B., Poldrack, R.A., 2015. Orthogonalization of regressors in fMRI models. PLoS One 10 (4). https://doi.org/10.1371/journal.pone.0126255. Article e0126255.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Ortiz-Tudela, J., Nolden, S., Pupillo, F., Ehrlich, I., Schommartz, I., Turan, G., Shing, Y.L., 2023. Not what u expect: effects of prediction errors on item memory. J. Exp. Psychol.: Gen. 152 (8), 2160–2176. https://doi.org/10.1037/xge0001367.

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

Poppenk, J., Walia, G., McIntosh, A.R., Joanisse, M.F., Klein, D., Köhler, S., 2008. Why is the meaning of a sentence better remembered than its form? An fMRI study on the role of novelty-encoding processes. Hippocampus 18 (9), 909–918. https://doi.org/10.1002/hipo.20453.

R Core Team. (2024). *R: a language and environment for statistical computing*. https://www.R-project.org/.

Ritvo, V.J.H., Turk-Browne, N.B., Norman, K.A., 2019. Nonmonotonic plasticity: How memory retrieval drives learning. Trends Cogn. Sci. 23 (9), 726–742. https://doi.org/10.1016/j.tics.2019.06.007.

Rouhani, N., Norman, K.A., Niv, Y., 2018. Dissociable effects of surprising rewards on learning and memory. J. Exp. Psychol.: Learn. Mem. Cogn. 44 (9), 1430–1443. https://doi.org/10.1037/xlm0000518.

Rugg, M.D., Vilberg, K.L., 2013. Brain networks underlying episodic memory retrieval. Curr. Opin. Neurobiol. 23 (2), 255–260. https://doi.org/10.1016/j.conb.2012.11.005.

Schacter, D.L., Addis, D.R., 2007. The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philos. Trans. R. Soc. B: Biol. Sci. 362 (1481), 773–786. https://doi.org/10.1098/rstb.2007.2087.

Schacter, D.L., Addis, D.R., Buckner, R.L., 2008. Episodic simulation of future events: concepts, data, and applications. Ann. N. Y. Acad. Sci. 1124 (1), 39–60. https://doi.org/10.1196/annals.1440.001.

Schlichting, M.L., Mumford, J.A., Preston, A.R., 2015. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat. Commun. 6 (1). https://doi.org/10.1038/ncomms9151. Article 8151.

Schubotz, R.I., 2007. Prediction of external events with our motor system: towards a new framework. Trends Cogn. Sci. 11 (5), 211–218. https://doi.org/10.1016/j.tics.2007.02.006.

Shao, X., Li, A., Chen, C., Loftus, E.F., Zhu, B., 2023. Cross-stage neural pattern similarity in the hippocampus predicts false memory derived from post-event inaccurate information. Nat. Commun. 14 (1). https://doi.org/10.1038/s41467-023-38046-y. Article 2299.

Siestrup, S., Jainta, B., Cheng, S., Schubotz, R.I., 2023. Solidity meets surprise: cerebral and behavioral effects of learning from episodic prediction errors. J. Cogn. Neurosci. 35 (2), 291–313. https://doi.org/10.1162/jocn_a_01948.

Siestrup, S., Jainta, B., El-Sourani, N., Trempler, I., Wurm, M.F., Wolf, O.T., Cheng, S., Schubotz, R.I., 2022. What happened when? Cerebral processing of modified structure and content in episodic cueing. J. Cogn. Neurosci. 34 (7), 1287–1305. https://doi.org/10.1162/jocn_a_01862.

Siestrup, S., Schubotz, R.I., 2023. Minor changes change memories: functional magnetic resonance imaging and behavioral reflections of episodic prediction errors. J. Cogn. Neurosci. 35 (11), 1823–1845. https://doi.org/10.1162/jocn_a_02047.

Sinclair, A.H., Barense, M.D., 2018. Surprise and destabilize: prediction error influences episodic memory reconsolidation. Learn. Mem. 25 (8), 369–381. https://doi.org/10.1101/lm.046912.117.

Sinclair, A.H., Barense, M.D., 2019. Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. Trends Neurosci. 42 (10), 727–739. https://doi.org/10.1016/j.tins.2019.08.007.

Sinclair, A.H., Manalili, G.M., Brunec, I.K., Adcock, R.A., Barense, M.D., 2021. Prediction errors disrupt hippocampal representations and update episodic memories. Proc. Natl. Acad. Sci. 118 (51). https://doi.org/10.1073/pnas.2117625118. Article e2117625118.

Specht, K., Wigglesworth, P., 2018. The functional and structural asymmetries of the superior temporal sulcus. Scand. J. Psychol. 59 (1), 74–82. https://doi.org/10.1111/sjop.12410.

Suzuki, M., Tsukiura, T., Matsue, Y., Yamadori, A., Fujii, T., 2005. Dissociable brain activations during the retrieval of different kinds of spatial context memory. Neuroimage 25 (3), 993–1001. https://doi.org/10.1016/j.neuroimage.2004.12.021.

Whalen, D.H., 2019. The motor theory of speech perception. In: Whalen, D.H. (Ed.), Oxford Research Encyclopedia of Linguistics. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.404.

Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., van den Bosch, A., 2016. Prediction during natural language comprehension. Cereb. Cort. 26 (6), 2506–2516. https://doi.org/10.1093/cercor/bhv075.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—again. Neuroimage 2 (3), 173–181.

Wurm, M.F., Schubotz, R.I., 2012. Squeezing lemons in the bathroom: contextual information modulates action recognition. Neuroimage 59 (2), 1551–1559. https://doi.org/10.1016/j.neuroimage.2011.08.038.