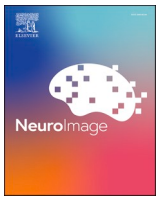
NeuroImage 325 (2026) 121660

Contents lists available at ScienceDirect

NeuroImage

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



Shaping Memory from the Start: Initial Prediction Errors during First Encoding

Nina Liedtke [a,b,*], Marius Boeltzig [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 stability
fMRI

ABSTRACT

The brain constantly makes predictions about upcoming input, and prediction errors (PEs) have been shown to promote encoding of the unexpected information. So far, previous experimental designs have left it unclear if PEs that may be evoked by the first exposure to a coherent novel stimulus, based on individual knowledge, experiences, and beliefs, can affect subsequent memory processes. In the current study, we aimed to test the neural and mnemonic consequences of these initial PEs and how they influence such outcomes together with later induced, experimental PEs. To this end, participants (N = 42) listened to naturalistic dialogues, which induced an initial PE, while undergoing fMRI scanning. Later, the dialogues were modified to induce a second, experimental PE, and memory for the original and modified versions was assessed using a recognition test. The results showed that initial PEs, like experimentally induced PEs, shifted the balance from top-down predictions to bottom-up processing, as reflected in reduced predictive reinstatement and stronger activation in the auditory cortex upon re-exposure. Moreover, semantic components of both initial and experimental PEs enhanced learning, while IFG activation biased memory towards the currently activated representation rather than the novel input. Taken together, these findings provide first evidence for the existence and relevance of initial PEs that are evoked during the encoding of coherent episodes not obviously violating world knowledge based on individual experiences and beliefs, indicating that they should be taken into consideration in paradigms investigating episodic PEs.

1. Introduction

A substantial body of research suggests that episodic memories not only serve to recall past experiences, but also support the generation of predictions about future events (Bubic et al., 2010; Friston & Kiebel, 2009). When these predictions deviate from the actual experience, a prediction error (PE) arises, potentially updating the currently active predictive model (Bubic et al., 2010; Friston & Kiebel, 2009; Sinclair & Barense, 2019). Previous research has shown that episodic PEs, mismatches that occur after a prediction based on episodic memories, can influence memory in different ways (Bein et al., 2023; Nolden et al., 2024). While they can promote encoding of the unexpected information as a new memory (Bein et al., 2021; Brod et al., 2018; Greve et al., 2017), they can also lead to updating of the old memory with the new information (Jainta et al., 2022; Siestrup et al., 2022; Siestrup & Schubotz, 2023; Sinclair et al., 2021; Sinclair & Barense, 2018) or even a weakening (Forcato et al., 2007) or pruning (Kim et al., 2014, 2017) of the original memory. Important factors shaping these outcomes include the content of prediction (Liedtke et al., 2025; Siestrup et al., 2022; Siestrup & Schubotz, 2023; Varga et al., 2025), the correctness of prediction (Boeltzig, Liedtke, & Schubotz, 2025; Greve et al., 2017; Liedtke et al., 2025), and prediction strength (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025; Greve et al., 2017; Kim et al., 2014).

However, one potential source of variability in previous studies that has received little attention is the unpredictability inherent in the first presentation of stimuli. Most studies have experimentally induced PEs by creating and then violating expectations, for example by presenting an episode that is later modified or interrupted (e.g., Jainta et al., 2022; Siestrup et al., 2023; Sinclair et al., 2021). However, predictive coding theory implies that a PE should also arise at the very first encounter with any novel event – as the default response (Friston & Kiebel, 2009). The

* Corresponding author. Nina Liedtke, 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), s.siestrup@uni-muenster.de (S. Siestrup), rschubotz@uni-muenster.de (R.I. Schubotz).

https://doi.org/10.1016/j.neuroimage.2025.121660
Received 10 September 2025; Received in revised form 16 December 2025; Accepted 16 December 2025
Available online 17 December 2025

brain continuously generates predictions based on prior knowledge, experiences, and beliefs, thereby shaping individual predictive models for everyday situations (Brown & Brüne, 2012). Thus, even before any experimental manipulation, the same material should elicit different PEs across individuals (Maguire et al., 1999; Raykov et al., 2022). For instance, a pianist may anticipate a different conversation in a piano lesson than a drummer or a non-musician. We refer to this as the *initial prediction error* (initial PE).

Prior research on PEs at first encounter with the stimulus material has largely focused on semantic violations, such as presenting material that conflicts with general world knowledge or schemas (Varga et al., 2025; Zöllner et al., 2021). These studies show that semantic PEs at first encounter with a stimulus can influence memory, for instance by increasing the likelihood of erroneously recalled details (Varga et al., 2025). However, no study to date has examined whether initial PEs occur in naturalistic, everyday-like situations without overt violations of world knowledge or schemas, or whether they leave measurable neural or behavioral traces. Here, we examine the neural and mnemonic consequences of the initial PE to establish its presence and functional relevance.

To this end, we also compared initial PEs to explicit violations of expectation that are purposefully induced during an experiment, which we refer to as *experimentally induced prediction errors* (experimental PE), for clarity. Although these experimental PEs are based on specific episodic predictions, derived from episodes encoded in the process of the experiment, while initial PEs are potentially based on semantic schema-related and/or episodic predictions, we are interested in whether they are processed similarly despite their differences. If so, this would provide first evidence for the universal role of PEs in stimulus processing.

In the current study, participants underwent fMRI scanning during their first encounter with naturalistic dialogues covering diverse scenarios. Each dialogue could elicit a varying degree of initial PE, depending on individual world knowledge and experiences. In a later fMRI session, we modified the dialogues to introduce an experimental PE. Finally, we tested recognition memory for both original and modified dialogues.

To investigate initial PEs and their consequences, we had two aims. First, we tested whether the initial PE could affect processing at the next exposure, and if it did so similarly to experimental PEs. Second, we asked whether this potential shift in processing could modulate the influence that the subsequently induced experimental PE has on memory outcomes.

Two brain regions that have consistently been found to be involved in processing PEs across modalities are the inferior frontal gyrus (IFG; Bubic et al., 2009; El-Sourani et al., 2018, 2020; Gläscher et al., 2010; Jainta et al., 2024; Varga et al., 2025; Wurm & Schubotz, 2012) and the hippocampus (HC; Bein et al., 2020; Duncan et al., 2012; Long et al., 2016). The IFG, which scales with how informative a PE is in a given scenario but not with PE size, reflects updating of the currently operating predictive model (El-Sourani et al., 2020). In a similar vein, Cope and colleagues (2023) argued that the IFG plays a primary role in supporting the reconciliation of top-down predictions with linguistic stimuli. However, in a recent study on episodic prediction, we found that only the IFG and not the hippocampus or the whole brain robustly reinstated, and therefore predicted, previously encoded episodes at re-exposure (Boeltzig, Liedtke, Siestrup, et al., 2025). Furthermore, recent research suggests that the HC is sensitive only to PEs arising from episodic prediction, not to those driven by semantic- or schema-based predictions, whereas the IFG responds to both (Varga et al., 2025). Because predictions in our task can stem from specific memories as well as from general world or schema knowledge, the resulting PEs can have both episodic and semantic features. We therefore used single-trial IFG activation at first presentation as an index of processing the initial PE. Likewise, single-trial IFG activation when the familiar dialogue was unexpectedly modified was used to index processing the experimental PE.

Using this PE measure, we first investigated the neural consequences of the initial PE and then compared them to those of experimental PEs. Predictive coding theory suggests that PEs can shift the balance from top-down predictions to bottom-up processing when the input conflicts with the current predictive model, prompting model updating (Friston & Kiebel, 2009). Previous research showed that this shift can modulate top-down predictions (Summerfield et al., 2008) as well as lower sensory processing (Richter et al., 2024), leading to enhanced sensory processing for unexpected compared with expected input (Summerfield et al., 2008; Todorovic et al., 2011).

To test whether initial and experimental PEs triggered a shift toward more bottom-up processing, we conducted two analyses. First, using representational similarity analysis (RSA), we measured how strongly the continuation of a dialogue was predicted when it was cued at the next exposure (Boeltzig, Liedtke, Siestrup, et al., 2025; Kim et al., 2014, 2017), since stronger prediction would signify more top-down and less bottom-up processing. We therefore hypothesized that stronger IFG activation during the initial PE in Session 1, indicating the processing of more informative initial PEs, would lead to less predictive reinstatement at the next stimulus exposure in Session 2 (H1.1a). Correspondingly, we hypothesized that stronger IFG activation during the experimental PE in Session 2 would lead to less reinstatement of a dialogue in Session 3 (H1.1b). Second, we assessed bottom-up processing with a whole-brain univariate analysis. We hypothesized that stronger IFG activation during the initial PE would be followed by stronger activation in the primary auditory cortex at the next exposure to the stimulus during Session 2, indicating more bottom-up processing (H1.2a). Likewise, we expected that stronger IFG activation during the experimental PE in Session 2 would cause stronger primary auditory cortex activation at re-exposure in Session 3 (H1.2b).

After establishing the neural consequences of the initial PE, we analyzed whether it influenced memory outcomes in interaction with the experimental PE. To capture further aspects of the initial and experimental PE, we included five further potentially relevant regions of interest in the analysis of memory outcomes, namely the hippocampus (HC) and parahippocampal gyrus (PHC), both associated with episodic memory (Rugg et al., 2015; Rugg & Vilberg, 2013) and episodic PEs (Liedtke et al., 2025; Sinclair et al., 2021; Varga et al., 2025), and the superior temporal gyrus (STG) and dorsomedial prefrontal cortex (dmPFC, BA8 + 9), associated with semantic memory and control (Jackson, 2021) and semantic PEs (Liu et al., 2023; Wang et al., 2023). We used linear mixed-effects models to test whether single-trial activity in these five ROIs predicted subsequent recognition of original and modified dialogues. We hypothesized that the initial PE would affect memory outcomes both directly and through its interaction with the experimental PE, thereby modulating the effect of the experimental PE on memory for the original (H2.1) and modified dialogue version (H2.2).

2. Methods

The current study has two companion papers (Boeltzig, Liedtke, Siestrup, et al., 2025; Liedtke et al., 2025) that are based on the same data collection but focus on the experimentally induced PE.

2.1. Participants

In total, 50 participants were recruited for the study, which all had normal or corrected-to-normal vision, were native German speakers, were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) and reported no history of neurological or psychiatric disorders or substance abuse. The data of two participants had to be excluded due to extensive movement in the scanner and the data of one participant due to technical problems. Five participants failed to complete all sessions. The final sample size was $N = 42$ (35 female, seven

male, age: $M = 21.98$, $SD = 3.30$, range: 18-31), which was based on the sample size of a behavioral study with a similar paradigm (Boeltzig, Liedtke, & Schubotz, 2025). Participants gave written informed consent to participate in the study and were compensated with course credit or money. The study protocol was approved by the Ethics Committee of the Faculty of Psychology and Sports Science at the University of Münster.

2.2. Stimuli

The stimulus set consisted of 36 naturalistic dialogues (described in more detail in Liedtke et al., 2025) that were written by the authors of this paper in the German language and recorded by 20 professional voice actors (ten male, ten female, age 31 – 58 years, $M = 40.20$, $SD = 6.86$). The dialogues covered a wide range of situations that could occur more or less commonly for someone from a student audience. They were designed so that the head (i.e., the beginning of the dialogue) would create an expectation about its continuation depending on individual experiences and world knowledge that could either be fulfilled or violated (see Table 1 for an example). They all had a unique background sound that matched the setting of the conversation. Their length varied between 21 and 34 seconds ($M = 27.31$, $SD = 3.02$).

In addition to the original version, participants were presented with slightly changed versions of the dialogues. To that end, each dialogue was prepared in four alternative versions where either the surface (i.e., the phrasing) or the gist (i.e., the content) was changed to a low or high degree. This change always occurred in the target, while the head and end remained unaltered across the different versions. Each participant was only exposed to one modification per dialogue during encoding. Further analyses on the strength and type of the experimental PE can be found in Liedtke et al. (2025).

2.3. Procedure

The five experimental sessions took place over the course of ten days. The fMRI experiments were presented using Presentation® (Version 23.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), and the recognition test was implemented in PsychoPy (Peirce et al., 2019). The experimental procedure is illustrated in Figure 1A.

In Session 1, 30 of the 36 dialogues were presented while participants underwent fMRI scanning. The instructions were to listen to the dialogues attentively as if overhearing a conversation between two people in real life and to visually imagine the scene as if witnessing it directly. Before the experiment started, participants could self-adjust the volume and practice the cover task. The cover task, used to ensure constant attention throughout the session, was to decide whether a word that was presented on the screen had appeared in the previous dialogue (yes/no) with a button press (see Figure 1B). The word either stemmed from the head or the end of the dialogue or was a new but semantically related word. After every tenth trial, there was a ten-second break. After the fMRI scan, participants listened to the same 30 dialogues once more and rated them on five different scales (everyday typicality, social consistency, valence, arousal and autobiographical association) of which only the everyday typicality rating was used for the current analysis. This task was performed in a separate room in front of a laptop. The session took around 60 minutes.

Two days later, Session 2 took place with the same procedure, instructions, and cover task as Session 1. This time, however, 24 of the 30 dialogues were presented in a modified version (surface low, surface high, gist low, gist high) in order to induce an experimental PE. The assignment of dialogues to the type of modification was counterbalanced across participants with six dialogues in each modification category. The remaining six dialogues from Session 1 remained unchanged. Also, six novel dialogues were presented, which served as a manipulation check that participants had in fact encoded the dialogues in Session 1 (see Liedtke et al., 2025). All dialogues in this session were presented twice, taking around 50 minutes in total.

In Session 3, one day later, the original dialogue versions from Session 1 were played once again to enable the comparison of the neural consequences of the experimentally induced PE at next stimulus exposure with those of the initial PE.

Five days later, in Session 4, participants completed a recognition test assessing their memory for both the original and changed dialogue versions. To that end, they listened to a total of 144 short excerpts from the dialogues (2.4 s – 6.9 s, $M = 4.55$ s), hereafter referred to as probes, and were asked to indicate their confidence whether they had heard this exact statement before on a scale from 1 = definitely new to 6 = definitely old (see Figure 1C) in a one-step procedure (Brady et al., 2023). To minimize strategic responding, participants were instructed to make a decision individually for each trial, independently of previous decisions. From each of the 36 dialogues, four excerpts were tested. Two of the probes had actually been presented in the experiment and two were unheard dialogue versions that served as similar lures. For the changing dialogues, participants heard the original target and the changed target that they had heard during the experiment. For the unchanging dialogues, where only the original target had been presented in the experiment, an additional probe from the dialogue head was used. The selection of lures was counterbalanced. For each modification category (e.g., surface low) each of the three possible lure combinations (e.g., surface high + gist low, surface high + gist high, gist low + gist high) was assigned to two dialogues (two x three combinations = six dialogues per modification category). The order of probes was counterbalanced and probes pertaining to one dialogue were distributed over the session by covertly organizing the experiment in four blocks, each containing one probe per dialogue.

In Session 5 on the next day, participants gave individual difference ratings about how differently they perceived the two presented dialogue versions on a scale of 1 = very small difference to 7 = very large difference ($M = 3.30$, $SD = 1.79$, $min = 1$, $max = 7$). These ratings were used as a control variable for the amount of induced change in Session 2. For analyses regarding the PE strength with the difference rating as the variable of interest, please refer to Liedtke et al. (2025).

Table 1
Translated Example Dialogue.

	The dialogue takes place during a lecture, the lecturer is speaking in the background.
Head	A: Today's lecture is pretty complicated. Are you still following? B: I was okay at the beginning, but since the break I've completely lost track… A: Could you maybe send me your notes afterwards? I haven't understood anything so far…
Target	B: To be honest, I'd rather not. If there is something wrong with it, I don't want to be responsible for it.
End	A: Maybe we can ask Anton later, he always pays attention. B: Right, I'll see him later at the movies, I'll ask him then.

Note. More examples as well as the audio files can be obtained from the corresponding author.

2.4. MRI data acquisition and preprocessing

Magnetic resonance imaging was conducted with a 3-Tesla Siemens Magnetom Prisma MR tomograph using a 20-channel head coil. Participants lay supine on the scanner bed and were lightly fixated using form-fitting cushions to minimize movement. Their right index and middle finger rested on the two appropriate buttons of a response box. Participants wore earplugs to reduce scanner noise and headphones, through which the dialogues were presented. Instructions and the cover task were presented via a screen that participants saw through a mirror mounted on the head coil.

Before the start of the experimental tasks, high-resolution T1-weighted anatomical images were obtained using a 3-D magnetization prepared rapid gradient echo sequence (192 slices, slice thickness =

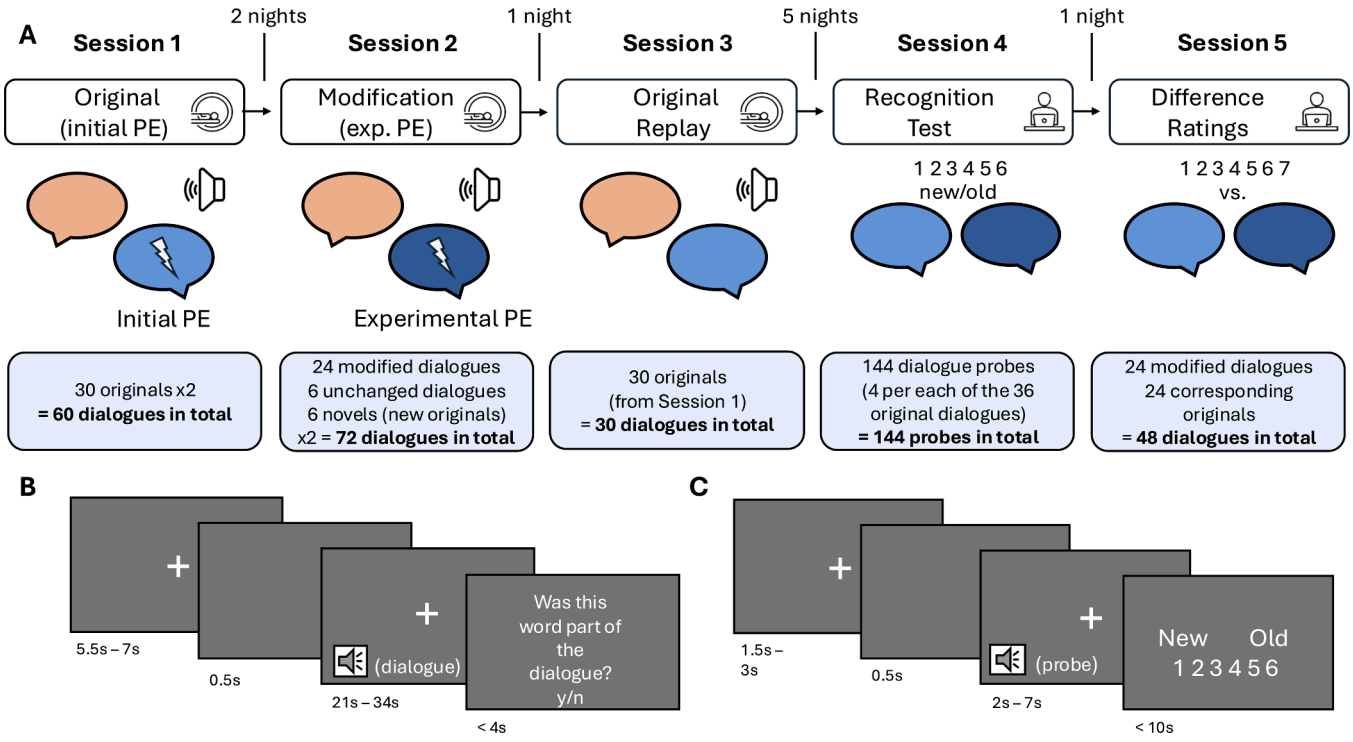


Figure 1. Experimental procedure and trial structure.
Note. Procedure and trial structure of the experiment. **A** The experiment consisted of five sessions that took place over the course of ten days. The first three sessions took place in the fMRI scanner. In Session 1, participants encoded the dialogues for the first time while in the scanner. In Session 2, some of the dialogues were played in an altered version and in Session 3, participants listened to the original version once again. In Session 4, participants completed a recognition test. Session 5 consisted of a difference rating between the original and modification. **B** The trials started with a jittered fixation cross that shortly disappeared right before stimulus presentation. After the dialogue finished, a word appeared on the screen and participants had to decide whether it had occurred in the dialogue or not. **C** For the recognition test, a probe (an excerpt from a dialogue or a similar lure) was played and participants had to indicate their confidence whether it was old or new on one item.

1mm, 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.). The preprocessing consisted of 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. Lastly, a 128s high-pass temporal filter was applied.

2.5. *Data preparation*

2.5.1. *Single-trial activation*

As a measure of the initial and experimental PE, we used single-trial IFG activation from the target (i.e., the middle part) of a dialogue (see Figure 2). Since individuals continuously generate predictions during unfolding narratives based on their individual predictive models (Baldassano, 2023), the beginning of a given dialogue was assumed to elicit expectations about its continuation. If these predictions were

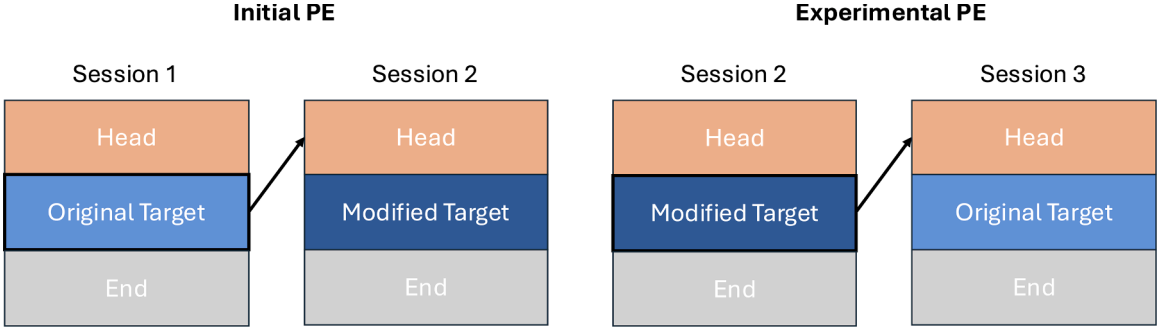


Figure 2. The single-trial activations and representational similarities of interest.
Note. For the initial PE, we used the single-trial activation in the IFG during the target of the original dialogue in Session 1. In line with this, the experimental PE was operationalized as the single-trial activation during the modified target in Session 2. For the RSA, we used the representational similarity between the Session 1 target and the Session 2 head to test the consequences of the initial PE and the Session 2 target and the Session 3 head to do the same for the experimental PE.

violated, increased single-trial activation was expected during the target in the IFG.

To further investigate how the episodic and semantic properties of the initial PE influence memory, we exploratorily included four additional ROIs closely associated with either episodic or semantic PEs in the memory analysis. These were the hippocampus (HC), parahippocampal gyrus (PHC), superior temporal gyrus (STG) and the dorsomedial prefrontal cortex (dmPFC). For the ROIs we used bilateral anatomical masks which were generated by parcellating the MNI brain according to the Desikan-Killiany atlas (Desikan et al., 2006) using FreeSurfer (Fischl, 2012). To calculate the trial-by-trial estimates, we modeled each original and modified dialogue target as a separate regressor of a GLM and then extracted the betas in each ROI. The obtained beta estimates were used as predictors of representational and mnemonic consequences.

2.5.2. Representational Similarity Analysis (RSA)

Our goal was to investigate whether top-down predictions are modulated by the initial PE and whether this modulation is similar to those of established, experimental PEs. As a measure of prediction strength we calculated how strongly the dialogue target would be reinstated (i.e., predicted) in the IFG at a subsequent presentation of the dialogue head in the next session. To this end, we conducted a representational similarity analysis (RSA; Kriegeskorte, 2008). We used the representational similarity between the Session 1 target and the Session 2 head as a measure of reinstatement upon cueing after the initial PE (Boeltzig, Liedtke, Siestrup, et al., 2025; Kim et al., 2014, 2017). Congruently, the similarity between the Session 2 target and the Session 3 head served as a measure of reinstatement after the experimental PE (see Figure 2).

To compute representational similarities, a generalized linear model was fitted with the normalized fMRI data, with one regressor per each dialogue head and target, while the two identical presentations in Session 2 were collapsed. Additionally, the model contained three parameters denoting the session, and six movement parameters per session. Then, using the CosmoMVPA toolbox (Oosterhof et al., 2016), we calculated a 192×192 similarity matrix containing the Pearson correlation coefficient between all heads and targets in the IFG. From this matrix, our measure of interest was extracted and baseline-corrected to account for overall similarity in the material. The baseline correction was performed by subtracting from the measure of interest the mean of the similarities between a given cue and all other targets (Boeltzig, Liedtke, Siestrup, et al., 2025; Shao et al., 2023).

2.5.3. Behavioral Data

To measure memory performance, we used a weighted accuracy measure (Boeltzig, Liedtke, & Schubotz, 2025; Liedtke et al., 2025) to avoid using confidence ratings of incorrect answers. Hence, incorrect responses on the recognition test ("new" answers for known items, which were indicated by confidence ratings ranging from 1-3) were coded as 0, while correct responses ("old" answers for old items, confidence ratings from 4-6) were assigned values from 1 to 3 based on participants' confidence ratings.

2.6. Statistical data analysis

The analysis of the RSA and behavioral data was conducted with RStudio (R Core Team, 2025).

2.6.1. Cover task

To test whether participants paid attention to the dialogues across all three fMRI sessions and were able to maintain an appropriate level of attention over the course of the experiment, we fit a linear mixed model for binomial distributions and performed pairwise comparisons between the sessions using Bonferroni correction.

2.6.2. Neural consequences of the initial and experimental PE

2.6.2.1. Reinstatement at re-exposure. Previous research showed that PEs can shift the balance from top-down prediction to bottom-up processing. To test whether this may result from initial PEs similarly to experimental PEs and whether the initial PE could thereby potentially affect the processing of the experimental PE, we looked at how strongly the dialogue target was predicted when participants listened to the head of a given dialogue once again, right before the experimental PE was induced in the following session. First, we fitted a linear mixed model to investigate the effects of the initial PE on target prediction at following stimulus presentation. The dependent variable was the representational similarity measure between the Session 1 target and the Session 2 head as a measure of prediction strength. The fixed effect was the trial-by-trial activation in the IFG from Session 1. Participants and dialogues were modeled as random intercepts. Then, we fitted the same model for the experimental PE. Here, the dependent variable was the representational similarity between the Session 2 target and the Session 3 head. Correspondingly, the fixed effect was the IFG single-trial activation from Session 2.

Exploratorily, we tested whether both PEs could also affect reinstatement of the original target when it was presented again in Session 3. To this end, we estimated an additional model that contained the Session 1 and Session 2 single-trial IFG activation and their interaction as fixed effects and participants and dialogues as random intercepts. The dependent variable was the representational similarity between the Session 1 target and the Session 3 head.

2.6.2.2. Brain activations at re-exposure. To investigate the consequences of the initial PE on brain activation during the next exposure to the dialogue more closely, we calculated a general linear model (GLM) for serially autocorrelated observations (Friston et al., 1994; Worsley & Friston, 1995) with the Session 2 data. Regressors were convolved with the canonical hemodynamic response function. All dialogues were modeled as epochs. Since we were especially interested in the prediction effects at the beginning of the dialogue in order to see the effects of the initial PE before the experimental PE was induced, the epoch was modeled from dialogue onset until the beginning of the target. To this dialogue regressor, a parametric modulator was added that contained the single-trial IFG activation during the original target of this dialogue (initial PE) from Session 1. The parametric modulator was mean-centered within each participant (Mumford et al., 2015). Responses to the cover task were modeled as events with onset on the button press. An additional regressor modeled the 72 null events (fixation crosses before dialogue onsets) as epochs with their full duration (5.5s-7s). Twelve novel dialogues that were only presented in Session 2 and not further analyzed in this experiment were modeled separately as epochs from dialogue onset to end. Six subject-specific rigid body transformations obtained from realignment were included as regressors of nuisance. In total, the GLM comprised ten regressors.

For the experimental PE, we calculated a parallel model with the Session 3 data. Dialogues were again modeled as epochs spanning only the beginning of the dialogue and the parametric modulator contained the IFG single-trial activation during the modified dialogue target of Session 2 (experimental PE). The only difference was that Session 3 did not contain any novels, which is why this model comprised nine regressors in total.

For each model, on the first level, we then calculated the parametric contrast of the single-trial IFG activation during the experimental PE and during the initial PE, respectively. On the second level, group analyses were performed using one-sample *t*-tests across participants. We applied false discovery rate (FDR) correction to the resulting *t*-maps with a threshold of $p < .001$ (voxel-wise). Reported results are restricted to clusters with an extent of at least ten voxels. Significant clusters were visualized using MRIcroGL (Version 1.2.20200331, McCausland Center

for Brain Imaging, University of South Carolina).

2.6.3. Influence of the initial PE and experimental PE on memory outcomes

As a next step, we investigated the effects of the initial PE in interaction with the experimental PE on recognition memory for the original and the modified dialogue versions. To this end, we calculated two linear mixed models. The dependent variables in these models were the weighted accuracy measures for the original (which induced the initial PE) or the modified version (which induced the experimental PE), respectively. As fixed effects, the models contained the trial-by-trial activation from Session 1 and Session 2 in all ROIs, the interactions between the IFG in Session 1 and 2 with all ROIs from the respective other session and the difference rating as a control. Again, participants and dialogues were modeled as random intercepts.

To account for potential individual differences in scale usage, the difference ratings were *z*-standardized within each participant. The single-trial data was mean centered within participants to facilitate interpretation. To assess multicollinearity in the models, variance inflation factors (VIFs) were calculated for all predictors. All VIFs were < 10, suggesting that multicollinearity was not a concern (Shrestha, 2020). The models were fitted using the *lme4* package (Bates et al., 2015) and tested using the *lmerTest* package (Kuznetsova et al., 2017).

2.6.4. Explorative Analysis: Role of the hippocampus

In the foregoing analyses, we focused on the IFG as mediating the updating of currently active predictions in response to prediction errors; however, the potential role of the HC in this process has not yet been addressed. As a recent study suggests, the HC may only respond to episodic PEs but not PEs based on schemas and general knowledge (Varga et al., 2025). To further explore the nature of the initial PE and whether the HC is activated by initial PEs that are based more on episodic predictions than on general world knowledge and schemas, we conducted a ROI analysis in the HC. To this end, we used the everyday typicality rating ("How typical is a situation like this in your everyday life?", 1 = very atypical to 7 = very typical, $M = 3.61$, $SD = 1.77$) that participants gave in Session 1 to model the brain activation during first encoding. The idea was that experiencing similar situations in their own lives might help participants to form predictions based on episodic memories instead of just general knowledge. For example, a person that plays football may form predictions about what happens in a football practice dialogue based on their own episodic memories in contrast to a person that has never played and can only rely on their general knowledge about football from external sources.

The GLM contained all dialogues modeled as epochs of their full length, together with a parametric modulator consisting of the everyday typicality ratings participants gave in Session 1. The ratings were *z*-standardized within participants. Again, responses to the cover task were modeled as events time-locked to the button press. A separate regressor captured the 72 null events, modeled as epochs spanning their full duration (5.5–7 seconds) and the six subject-specific motion parameters from realignment were included as nuisance regressors. For this analysis we used the same anatomical mask of the hippocampus as for the single-trial analysis. The mean estimate values of the parametric regressor of everyday typicality were extracted using the MarsBar Toolbox (Brett et al., 2002). Lastly, we performed a one-sample *t*-tests to check for significant activation within the HC.

3. Results

3.1. Cover task

Performance on the cover task was good in all three fMRI sessions with participants responding correctly in 78.03% to 97.72% of trials across all sessions (Session 1 $M = 86.74\%$, $SD = 7.85\%$, Session 2 $M = 89.11\%$, $SD = 5.90\%$, Session 3 $M = 91.47\%$, $SD = 5.79\%$). From Session 1 to Session 2, there was an increase in performance, $\beta = 0.23$, $SE = 0.10$, $Z = 2.43$, $p = 0.045$, which might be due to increasing familiarity with the dialogues. Between Session 2 and 3 there were no significant differences, $\beta = 0.22$, $SE = 0.11$, $Z = 2.01$, $p = 0.132$. Therefore, we can assume that participants listened to the dialogues attentively and that there was no decline in performance across sessions.

3.2. Neural consequences of the initial and experimental PE

As a first step, we wanted to establish the neural consequences of the initial PE and whether they could affect the processing of the later induced, experimental PE. Also, we were interested if these neural consequences were similar to those of experimental PEs. To that end we tested whether both PEs affect top-down prediction and activations in the whole brain in a similar manner when encountering the same stimulus again in the following session.

3.2.1. Reinstatement at re-exposure

Our hypothesis was that an initial PE could influence the processing of a following, experimental PE by shifting the balance from top-down predictions to bottom-up processing right before the experimental PE is induced. We therefore expected to see less reinstatement upon cueing of a dialogue after a more informative initial PE (H1.1a). Consistent with this hypothesis, reinstatement in the IFG was reduced as a function of single-trial IFG activation related to the initial PE at first encoding, $\beta =$ -0.02, 95% CI [-0.03, -0.01], $SE = 0.01$, $t(957) = -3.06$, $p = .002$ (see Figure 3A). A comparable effect emerged for the experimental PE (H1.1b): more single-trial IFG activation during the modification (i.e., the experimental PE in Session 2) also predicted less reinstatement of the target in the IFG at re-exposure in Session 3, $\beta =$ -0.03, 95% CI [-0.04, -0.01], $SE = 0.01$, $t(957) = -3.37$, $p < .001$ (see Figure 3B).

Exploratorily, we tested if initial PE and experimental PE affected reinstatement of the original target in Session 3. IFG activation related to the initial PE did not predict Session 3 original reinstatement, $\beta =$ -0.00, 95% CI [-0.02, 0.01], $SE = 0.01$, $t(953) = -0.57$, $p = .570$. Neither did the Session 2 IFG activation related to the experimental PE, $\beta =$ -0.00, 95% CI [-0.02, 0.01], $SE = 0.01$, $t(945) = -0.72$, $p = .472$. The interaction between the two showed a non-significant trend, $\beta =$ -0.02, 95% CI [-0.03, 0.00], $SE = 0.01$, $t(953) = -1.91$, $p = .057$.

3.2.2. Brain activations at re-exposure

In line with our hypothesis of reduced top-down modulation, we expected enhanced sensory-driven processing, reflected in auditory cortex activation, after more informative initial PEs (H1.2a). To test this, we examined brain responses during the second encounter with a dialogue (Session 2) that had elicited a more informative initial PE in Session 1. This hypothesis was confirmed by a parametric modulation analysis, which showed that stronger trial-by-trial IFG activation during the first presentation was associated with increased bilateral auditory cortex activation when the same dialogue was repeated in Session 2 (Table 2). In a corresponding analysis, we tested whether experimental PEs produced a comparable effect at re-exposure in Session 3 by using single-trial IFG activation from Session 2 as the parametric modulator. As expected, this analysis revealed a highly similar pattern (H1.2b), with robust bilateral auditory cortex activation (Figure 3C, Table 2).

3.3. Influence of the initial PE on memory outcomes

After establishing the neural consequences of the initial PE, we were interested whether it influenced memory outcomes, even if the memories were later challenged by the experimentally induced PE. To this end, we calculated two linear mixed models, one for the memory of the original version that participants encoded in Session 1 and one for the modified version that induced the experimental PE in Session 2.

3.3.1. Original Recognition

For the original version, STG activation during first encoding had a

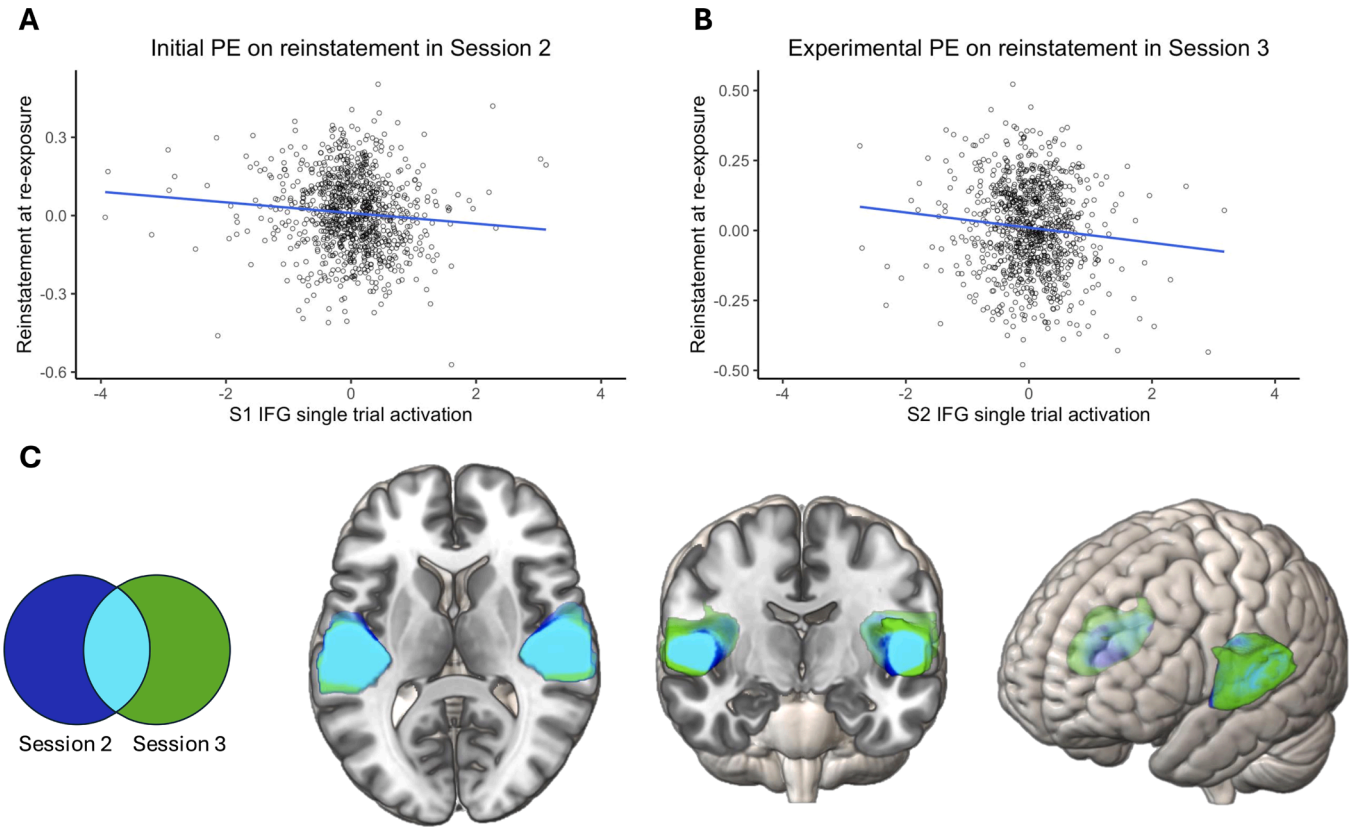


Figure 3. Influence of the initial and experimental PEs at re-exposure.
Note. **A** More single-trial IFG activation in Session 1 led to less reinstatement of a dialogue upon cueing in Session 2. **B** Similarly, more IFG activation in Session 2 led to less reinstatement of the dialogue in Session 3. **C** The results showed increased activations in the auditory cortex at the beginning of a dialogue, after parametrically higher IFG activation during that same dialogue in the previous session. Results are shown at FDR-corrected *p*-values < .001 and cluster-size threshold of k=10, to have the same $t > 4.5$ threshold on both contrasts. Blue: Session 2 data modulated by the trial-by-trial IFG activation in Session 1. Green: Session 3 data modulated by the trial-by-trial IFG activation in Session 2. Turquoise: Overlap of the two contrasts.

Table 2
Whole-brain activation modulated by IFG activation during previous encoding.

			MNI Coordinates			
Localization	***H***	***Cluster Size***	***X***	***Y***	***Z***	***t Value***
Parametric contrast: Session 1 IFG activation on Session 2 whole brain data (FDR p < .001)						
Auditory cortex	L	626	-48	-19	8	13.16
	R	531	63	-19	8	12.36
Parametric contrast: Session 2 IFG activation on Session 3 whole brain data (FDR p < .001)						
Auditory cortex	L	838	-48	-19	8	14.06
	R	730	63	-19	8	14.12

Note. Only clusters with a minimum extent of 10 voxels are reported. H = hemisphere; MNI = Montreal Neurological Institute; L = left; R = right; l.m. = local maximum. Both contrasts are corrected at FDR $p < .001$ (voxel level).

beneficial effect on recognition, $\beta = 0.27$, 95% CI [0.07, 0.47], $SE = 0.10$, $t(931) = 2.68$, $p = .008$. When there was more STG activation in a trial during first encoding of the dialogue, the original dialogue version was remembered better (Figure 4A). For Session 1 IFG single-trial activation we saw a negative non-significant trend, $\beta = -0.18$, 95% CI [-0.38, 0.01], $SE = 0.10$, $t(918) = -1.78$, $p = .075$.

In Session 2, IFG activation affected the original memory, $\beta = 0.33$, 95% CI [0.08, 0.58], $SE = 0.13$, $t(929) = 2.57$, $p = .010$ (Figure 4B). Contrary to Session 1, however, more IFG activation in Session 2 led to better recognition of the original dialogue version. Additionally, there was a significant interaction between the Session 2 IFG activation and Session 1 PHC activation, $\beta = -0.48$, 95% CI [-0.92, -0.04], $SE = 0.23$, $t(896) = -2.13$, $p = .034$. The beneficial effect of better memory for the original after more Session 2 IFG activation was stronger after lower PHC activation in Session 1 than after stronger PHC activation in Session 1 (Figure 4C). The other predictors had no significant effect (*ps* > .120).

3.3.2. Modification Recognition

Recognition of the modification was affected by single-trial activation in the dmPFC in Session 1, $\beta = 0.23$, 95% CI [0.07, 0.38], $SE = 0.08$, $t(929) = 2.76$, $p = .006$, with higher dmPFC activation leading to better modification memory (Figure 4D). As for the original, there was a significant positive effect of single-trial STG activation in Session 2 (at first encoding of the respective version) on modification recognition, $\beta = 0.39$, 95% CI [0.12, 0.65], $SE = 0.14$, $t(835) = 2.84$, $p = .005$ (Figure 4E). Additionally, we saw a negative effect of Session 2 PHC activation, $\beta = -0.43$, 95% CI [-0.84, -0.01], $SE = 0.22$, $t(929) = -1.97$, $p = .049$. Stronger activation in the PHC during the experimental PE led to worse recognition in this trial (Figure 4F). The other predictors showed no significant effect (*ps* > .058).

3.4. Explorative Analysis: Role of the hippocampus

Our previous analyses confirmed the proposed role of the IFG in PE detection but so far, we could not show involvement of the hippocampus as a mismatch detector in the initial PE. To investigate the role of the hippocampus in spontaneous PEs more closely, we performed a ROI analysis. The results showed that with increasing exposure to similar situations in their everyday life, there was more hippocampal activation during first dialogue presentation, $t(40) = 3.21$, $p = .003$.

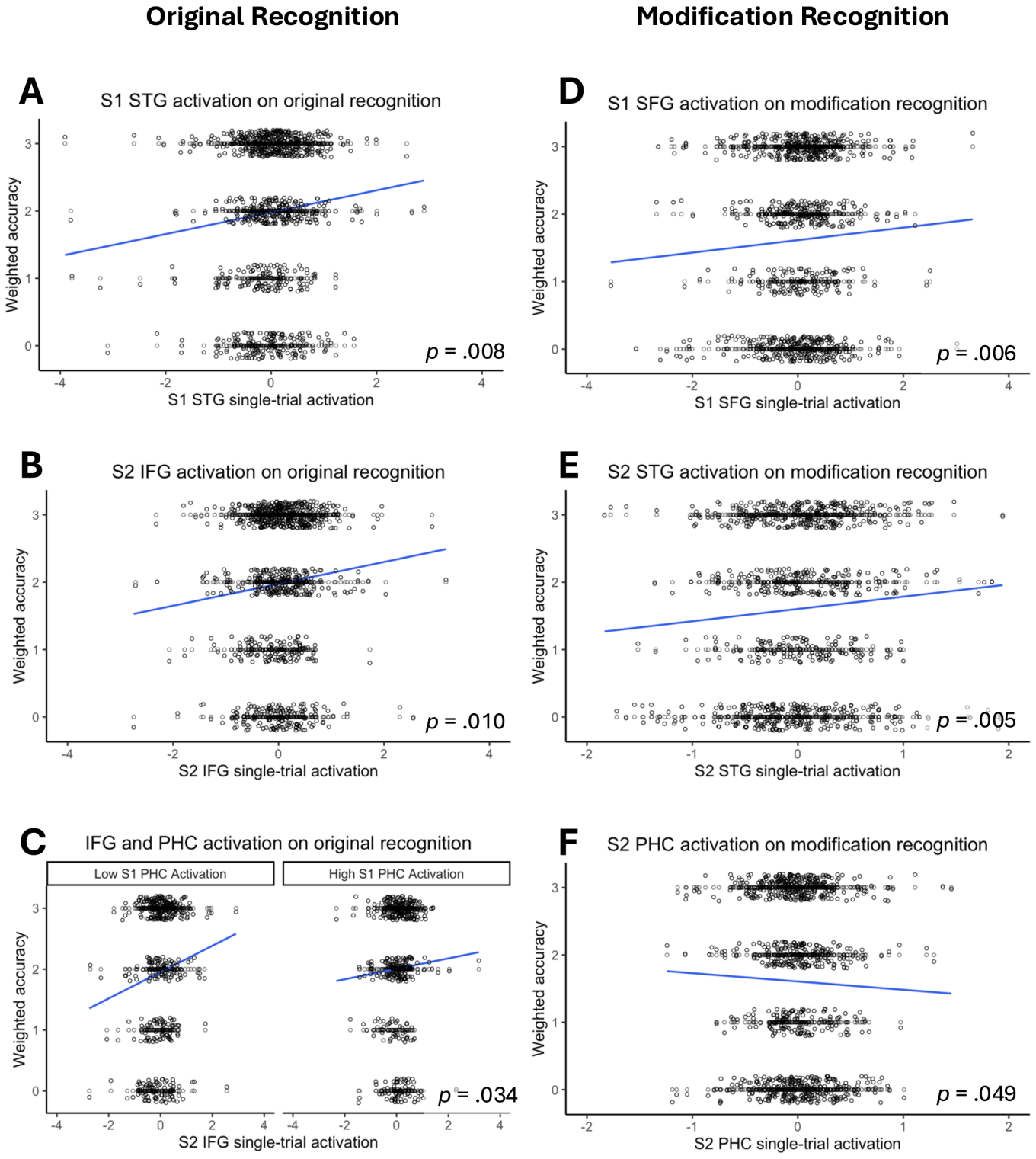


Figure 4. Effects of the initial and induced PE on memory outcomes.

Note. Effects of the initial (Session 1) and experimental (Session 2) PEs on memory. Vertical jitter was added for visualization. **A** Stronger single-trial activation in the STG in Session 1 led to better recognition of the original. **B** Enhanced single-trial activation in the IFG in Session 2, i.e., during the presentation of the modification, led to better recognition of the original dialogue version. **C** The positive effect of Session 2 IFG activation on original memory was stronger after weaker PHC activation in Session 1. **D** Session 1 dmPFC activation had a positive effect on recognition of the modification. **E** Stronger single-trial activation in the STG in Session 2 led to better recognition of the modification. **F** Stronger single-trial PHC activation in Session 2 led to worse recognition of the modification.

4. Discussion

Research on episodic prediction errors (PEs) has so far overlooked errors that occur spontaneously without an obvious violation of world knowledge or schemas at first encoding. Yet if the brain continuously generates predictions, such errors - varying in magnitude - are an inherent feature of every new experience. Recognizing these *initial PEs* is therefore essential, both for understanding everyday memory formation

and for designing experimental paradigms that probe episodic PEs. Here, we set out to demonstrate their occurrence at encoding and to examine their impact on neural activity and subsequent memory.

Our results show that such initial PEs shifted the balance between predictive and sensory-driven processing and left measurable neural signatures at re-exposure. Specifically, they reduced reinstatement (i.e., prediction) of expected dialogue continuations and increased activation in the auditory cortex. Importantly, both effects covaried with trial-by-trial IFG activity during the first encounter, our index of initial PE processing. In parallel, initial PEs also shaped memory: activation in brain regions associated with semantic processing reliably enhanced recognition of both original and modified dialogues, while IFG engagement biased memory toward the currently active variant rather than the novel input. Taken together, these findings establish that PEs at initial encoding are not only detectable but also leave lasting consequences for neural processing and memory.

4.1. Neural consequences of initial prediction errors

Within the predictive coding framework, initial PEs can be expected to transiently boost processing of sensory input while reducing the impact of top-down predictions (Friston & Kiebel, 2009), consistent with evidence that sensory cortices are modulated by surprise (Richter et al., 2024) and show enhanced responses to unexpected compared with expected stimuli (Summerfield et al., 2008; Todorovic et al., 2011). In contrast to previous studies that examined PE effects at the moment of induced mismatch, we focused on the re-exposure in Session 2, immediately before the experimental PE, because this is the point where the original scenario is reinstated from memory, and any traces of the initial PE should manifest. In line with our hypotheses, dialogues that elicited stronger IFG responses at first presentation (indexing the initial PE) showed both reduced reinstatement of the expected continuation (H1.1a) and enhanced auditory cortex activation (H1.2a) when re-exposed in Session 2.

These findings indicate that initial PEs leave traces that shape how subsequent encounters are processed and may modulate the impact of later, experimentally induced PEs. While most previous studies have focused on a single, experimentally induced PE (Greve et al., 2017; Jainta et al., 2022; Liedtke et al., 2025; Siestrup & Schubotz, 2023), few have investigated conditions involving multiple PEs. One study, for instance, compared variable and repetitive episodic PEs and found that both types of violations recruited IFG and hippocampus during the mismatch, but variable PEs additionally engaged the caudate and amygdala, attributed to a failure to adapt top-down predictions under recurrent violations (Jainta et al., 2024). Another study with multiple PEs presented short video clips with either typical or atypical target actions, thereby inducing purely semantic PEs, purely episodic PEs, or both (Varga et al., 2025). Yet, neither of these studies examined whether one PE influences the impact of another. Our findings therefore suggest that initial PEs can leave lasting traces that shape how new experiences are encoded and how stable memories remain when challenged by later changes.

4.2. Parallels between initial and experimental PEs

To assess whether the observed effects truly reflect PE processing rather than general encoding-related activity, we asked whether the neural consequences of initial PEs resemble those of experimental PEs. We therefore repeated the two analyses described above for the experimental PE, using the IFG single-trial activation during the modified target of each dialogue in Session 2. The results showed that experimental PEs, like initial PEs, predicted reduced reinstatement and increased auditory cortex activation when the same dialogues were repeated in Session 3, provided they had elicited a more informative episodic PE during the previous exposure (H1.1b & H1.2b). Thus, the neural signature of the experimental PE closely resembled the pattern observed for the initial PE, highlighting their similar role in shaping subsequent processing. This finding is especially interesting given the potentially quite different predictions that both PEs are based on. While the experimental PE is most likely based on a specific, episodic prediction, the initial PE could be based on a mix of episodic and semantic, schema-related predictions. Nevertheless, their processing appears to be highly similar.

Building on these parallels between initial and experimental PEs, our findings converge with prior evidence that PEs can reconfigure the balance between top-down predictions and bottom-up processing. Previous research demonstrated that this reconfiguration can also be observed in hippocampal connectivity, with experimental PEs increasing the connectivity of hippocampal subfield CA1 and the entorhinal cortex, supporting bottom-up processing, and decreasing CA1/CA3 connectivity associated with top-down retrieval (Bein et al., 2020). Similarly, beta-band power, which is associated with top-down predictions, decreases linearly with unpredictability, while gamma-band power, reflecting bottom-up prediction errors, increases (Van Pelt et al., 2016). However, both studies focused on the time point of the mismatch itself and did not investigate longer-lasting effects on subsequent processing. To our knowledge, this study is the first to demonstrate that experimental PEs can tip this balance not only during the mismatch but also at later re-exposure to the same material.

4.3. Memory effects of prediction errors

Next, we examined how the initial PE affected memory outcomes. We investigated the individual and combined effects of the initial and the experimental PE on recognition memory. Although the initial PE was elicited by the original version and the experimental PE by the modified version, both shaped recognition of each dialogue version.

The IFG, our central region of interest, showed context-dependent effects. During the initial PE, stronger IFG engagement tended to reduce memory for the presented original dialogue (note that this was a trend short of significance), whereas IFG activation during the experimental PE enhanced memory for the original version. Within a predictive brain perspective, these patterns are consistent with the idea that the IFG does not directly update predictive models, but rather modulates whether the currently active model is carried forward when confronted with conflicting input. At first exposure, robust prior knowledge may still have been brought to bear against unexpected sensory evidence, weakening memory for the presented dialogue. At the experimental mismatch, by contrast, the trace from the initial encounter was reinstated, strengthening memory for the original version. Taken together, the IFG appears to support a reconciliation process that temporarily sustains the active model under conflict, ensuring stability until sufficient evidence for an update accumulates.

In contrast, the STG showed a consistent positive influence on memory across both PEs. For the original version, which elicited the initial PE (H2.1), STG activation during first encoding predicted better recognition, and for the modification (H2.2), STG activation during its first presentation likewise enhanced memory. This aligns with prior work showing that the STG supports the encoding of speech and the extraction of meaningful linguistic features (Bhaya-Grossman & Chang, 2022; Mesgarani et al., 2014; Yi et al., 2019), as well as narrative comprehension (Babajani-Feremi, 2017). Thus, semantic components of the PE reliably promoted encoding, regardless of whether the episode was part of an initial or an experimental PE.

The PHC, in turn, generally had a weakening effect on memory. Stronger parahippocampal activation during the presentation of the modification led to worse recognition thereof. Furthermore, the PHC moderated the effects of IFG activation: the positive effect that IFG activity during the experimental PE exerted on later recognition of the original version was particularly pronounced when the PHC activation was low during the initial PE (Session 1). Given that the PHC is known to support contextual integration (Aminoff et al., 2013) of spatial (Davachi,

2006; Suzuki et al., 2005) as well as non-spatial information (Diana, 2016), these results suggest that the PHC may contribute a contextual component to prediction errors. One possible interpretation is that strong PHC engagement at first encoding already embedded the original dialogue in a broad contextual network, which could have made later predictions less specific and thereby limited the influence of IFG activity. Conversely, when initial PHC engagement was weaker, the original trace may have been less well integrated, allowing IFG activity during the experimental PE to exert a stronger effect on memory for the original. This interpretation remains tentative but points to a potential division of labor, with the PHC contributing contextual integration and the IFG modulating memory under conflict. Thus, it is possible that when the original dialogues were more integrated before, predictions were less specific, buffering the effect that IFG-related PEs have on original memory.

Finally, the dmPFC contributed specifically to memory for the modified version. Stronger dmPFC activation during Session 1 predicted better recognition memory of the modification. The dmPFC has been implicated in the encoding of narratives (Yarkoni et al., 2008) and specifically in narrative speech (Babajani-Feremi, 2017). In our data, however, this effect emerged only in the context of the initial PE, and not the experimental PE. One possible interpretation is that strong narrative encoding at the very first encounter provided a stable reference frame, making later modifications easier to detect and remember as deviations from the original story. This interpretation remains tentative and requires direct testing in future work.

4.4. The nature of initial prediction errors

While the current analyses reveal the consequences of initial PEs, they do not clarify the nature of the predictions underlying them. Since expectations during the first encounter can be based on specific episodes, general world knowledge and/or schemas (Brown & Brüne, 2012), it is plausible that the initial PE contains both episodic and semantic components. Our main analyses therefore focused on the IFG, a region known to respond to both episodic and semantic PEs (Varga et al., 2025). Yet this approach cannot determine whether the initial PE in our paradigm contained episodic elements or was purely semantic, which motivated an exploratory ROI analysis targeting the hippocampus.

The results showed that when participants reported higher everyday typicality for a certain dialogue, that is, when they were more familiar with similar situations in their own daily life, the hippocampus showed increased activation during first dialogue presentation. Importantly, the use of everyday typicality ratings, collected for each stimulus, provides a novel approach to capture individual variability in initial PEs and to explore when predictions could possibly also draw on episodic experiences rather than just semantic knowledge. The finding that hippocampal activation scaled with typicality could be interpreted in line with its specificity to episodic PEs, consistent with prior reports of hippocampal involvement when expectations derived from previously encoded episodes are violated (Liedtke et al., 2025; Sinclair et al., 2021; Varga et al., 2025). Correspondingly, in our data, hippocampal involvement was stronger when participants regularly experienced similar situations, possibly because they could map a dialogue onto experiences from their own everyday life (e.g., recalling previous football practices). In the absence of personal experiences, expectations were likely guided by general semantic knowledge (e.g., schemas and stereotypes about footballers and football practice).

As Varga et al. (2025) previously showed that the hippocampus does not respond to semantic violations, it seems possible that the activation observed here stems from PEs based on episodic memories, even though participants with more experience with a given situation may also have more semantic knowledge about it. At the same time, an alternative interpretation cannot be ruled out: stronger hippocampal activation under high typicality may also reflect increased personal relevance and associated shifts in attentional states (Aly & Turk-Browne, 2016). Thus, the precise circumstances under which individual predictions rely on episodic versus semantic memory cannot be resolved here and remain an important target for future research.

Nevertheless, the conclusion that predictions may be based on pre-existing episodic memories is intriguing. When facing a new situation, relevant memories may be activated to serve as predictive model. The initial PEs that result from this process may then impact those memories, by weakening them (Kim et al., 2014, 2017), adding new details (Jainta et al., 2022; Siestrup et al., 2022; Siestrup & Schubotz, 2023; Sinclair et al., 2021; Sinclair & Barense, 2018), or changing representational formats (Bein et al., 2020; Boeltzig, Liedtke, Siestrup, et al., 2025). Such PE-based modifications therefore provide a potential mechanism through which memories can undergo change, even in the absence of deliberate retrieval.

4.5. Limitations

The goal of the current study was to create natural communicative scenarios as close as possible to everyday life. We therefore refrained from asking participants to overtly state their expectations about what the speakers would say next or to rate their surprise during encoding. Because of this, we do not know the explicit, consciously accessible predictions that participants generated during their first exposure to each dialogue. Knowing these expectations could have been helpful for assessing the strength of the initial PE and for understanding interactions with the experimental PE. For instance, if the modification aligned more closely with a participant's original expectation of the dialogue, the experimental PE could have been weaker than the initial PE, and vice versa. Future studies could address this by obtaining a subjective PE strength score, for example through an expectedness rating directly after each dialogue, or by pausing dialogues after the initial segment and letting participants explicitly predict the continuation (as done for example in Stawarczyk et al., 2023), which could then be evaluated using large language models.

Furthermore, it was necessary in our study design to replay the original dialogues once more in Session 3 to examine the neural consequences of the episodic PE on reinstatement. However, this additional replay might have affected memory outcomes, since it provided a further encoding opportunity for the original version.

4.6. Conclusion

The current study investigated the consequences of initial PEs, which arise spontaneously upon the first encounter with a novel episode. Our results show that the initial PE, just as experimentally induced PEs, can shift the balance from top-down predictions to bottom-up processing and shape memory outcomes. This demonstrates that initial PEs are not a mere by-product of novelty but a systematic factor in shaping neural processing and memory formation. Future studies on episodic PEs should therefore consider that not only the experimental manipulation but also the material itself evokes PEs that affect memory.

Data availability statement

The behavioral data is available on OSF (https://osf.io/2ykva/?view_only=169d99b2a70a42868c2791ab68b5097c). Unthresholded statistical maps of all reported fMRI contrasts in the manuscript have been deposited on NeuroVault (https://identifiers.org/neurovault.collection:21778). Samples of the stimuli can be provided upon request to the corresponding author.

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, Data curation, Conceptualization. **Marius Boeltzig:** Writing – review & editing, Software, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Sophie Siestrup:** Writing – review & editing, Methodology, 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.

Acknowledgments

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 the Biological Psychology Group of the University of Münster for their helpful comments and for the valuable discussions.

References

Aly, M., Turk-Browne, N.B., 2016. Attention promotes episodic encoding by stabilizing hippocampal representations. Proceedings of the National Academy of Sciences 113 (4), E420–E429. https://doi.org/10.1073/pnas.1518931113.

Aminoff, E.M., Kveraga, K., Bar, M., 2013. The role of the parahippocampal cortex in cognition. Trends in Cognitive Sciences 17 (8), 379–390. https://doi.org/10.1016/j.tics.2013.06.009.

Babajani-Feremi, A., 2017. Neural mechanism underling comprehension of narrative speech and its heritability: Study in a large population. Brain Topography 30 (5), 592–609. https://doi.org/10.1007/s10548-017-0550-6.

Baldassano, C., 2023. Studying waves of prediction in the brain using narratives. Neuropsychologia 189, 108664. https://doi.org/10.1016/j.neuropsychologia.2023.108664. Article.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

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

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. Neuroscience & Biobehavioral Reviews 153, 105368. https://doi.org/10.1016/j.neubiorev.2023.105368. Article.

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

Bhaya-Grossman, I., Chang, E.F., 2022. Speech computations of the human superior temporal gyrus. Annual Review of Psychology 73 (1), 79–102. https://doi.org/10.1146/annurev-psych-022321-035256.

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.

Boeltzig, M., Liedtke, N., Siestrup, S., Mecklenbrauck, F., Wurm, M.F., Bramão, I., Schubotz, R.I., 2025. The benefit of being wrong: How prediction error size guides the reshaping of episodic memories. NeuroImage 317, 121375. https://doi.org/10.1016/j.neuroimage.2025.121375. Article.

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. Psychonomic Bulletin & Review 30 (2), 421–449. https://doi.org/10.3758/s13423-022-02179-w.

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, p. S497. Article.

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

Brown, E.C., Brüne, M., 2012. The role of prediction in social neuroscience. Frontiers in Human Neuroscience 6, 147. https://doi.org/10.3389/fnhum.2012.00147.

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

Bubic, A., Von Cramon, D.Y., Schubotz, R.I., 2010. Prediction, cognition and the brain. Frontiers in Human Neuroscience 4, 1094. https://doi.org/10.3389/fnhum.2010.00025.

Cope, T.E., Sohoglu, E., Peterson, K.A., Jones, P.S., Rua, C., Passamonti, L., Sedley, W., Post, B., Coebergh, J., Butler, C.R., Garrard, P., Abdel-Aziz, K., Husain, M., Griffiths, T.D., Patterson, K., Davis, M.H., Rowe, J.B., 2023. Temporal lobe perceptual predictions for speech are instantiated in motor cortex and reconciled by inferior frontal cortex. Cell Reports 42 (5), 112422. https://doi.org/10.1016/j.celrep.2023.112422.

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

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. Cerebral Cortex 27 (3), 1808–1816. https://doi.org/10.1093/cercor/bhw014.

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. Journal of Cognitive Neuroscience 32 (2), 326–337. https://doi.org/10.1162/jocn_a_01480.

El-Sourani, N., Wurm, M.F., Trempler, I., Fink, G.R., Schubotz, R.I., 2018. Making sense of objects lying around: How contextual objects shape brain activity during action observation. NeuroImage 167, 429–437. https://doi.org/10.1016/j.neuroimage.2017.11.047.

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. Learning & Memory 14 (4), 295–303. https://doi.org/10.1101/lm.486107.

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

Friston, K.J., Kiebel, S., 2009. Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B: Biological Sciences 364 (1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300.

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? Journal of Memory and Language 94, 149–165. https://doi.org/10.1016/j.jml.2016.11.001.

Jackson, R.L., 2021. The neural correlates of semantic control revisited. NeuroImage 224, 117444. https://doi.org/10.1016/j.neuroimage.2020.117444. Article.

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. Frontiers in Behavioral Neuroscience 15, 793115. https://doi.org/10.3389/fnbeh.2021.793115. Article.

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

Kim, G., Lewis-Peacock, J.A., Norman, K.A., Turk-Browne, N.B., 2014. Pruning of memories by context-based prediction error. Proceedings of the National Academy of Sciences 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. The Journal of Neuroscience 37 (8), 2022–2031. https://doi.org/10.1523/JNEUROSCI.3272-16.2017.

Kriegeskorte, N., 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience 2, 249. https://doi.org/10.3389/neuro.06.004.2008.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software 82 (13). https://doi.org/10.18637/jss.v082.i13.

Liedtke, N., Boeltzig, M., Mecklenbrauck, F., Siestrup, S., Schubotz, R.I., 2025. Finding the sweet spot of memory modification: An fMRI study on episodic prediction error strength and type. NeuroImage 311, 121194. https://doi.org/10.1016/j.neuroimage.2025.121194. Article.

Liu, L., Liu, D., Guo, T., Schwieter, J.W., Liu, H., 2023. The right superior temporal gyrus plays a role in semantic-rule learning: Evidence supporting a reinforcement learning model. NeuroImage 282, 120393. https://doi.org/10.1016/j.neuroimage.2023.120393.

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. The Journal of Neuroscience 36 (50), 12677–12687. https://doi.org/10.1523/JNEUROSCI.1850-16.2016.

Maguire, E.A., Frith, C.D., Morris, R.G.M., 1999. The functional neuroanatomy of comprehension and memory: The importance of prior knowledge. Brain 122 (10), 1839–1850. https://doi.org/10.1093/brain/122.10.1839.

Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343 (6174), 1006–1010. https://doi.org/10.1126/science.1245994.

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

Nolden, S., Turan, G., Güler, B., Günseli, E., 2024. Prediction error and event segmentation in episodic memory. Neuroscience & Biobehavioral Reviews 157, 105533. https://doi.org/10.1016/j.neubiorev.2024.105533.

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.

Oosterhof, N.N., Connolly, A.C., Haxby, J.V., 2016. CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Frontiers in Neuroinformatics 10. https://doi.org/10.3389/fninf.2016.00027.

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

R Core Team, 2025. R: A language and environment for statistical computing. https://www.R-project.org/.

Raykov, P.P., Keidel, J.L., Oakhill, J., Bird, C.M., 2022. The importance of semantic network brain regions in integrating prior knowledge with an ongoing dialogue. Eneuro 9 (5). https://doi.org/10.1523/ENEURO.0116-22.2022. Article ENEURO.0116-22.2022.

Richter, D., Kietzmann, T.C., De Lange, F.P., 2024. High-level visual prediction errors in early visual cortex. PLOS Biology 22 (11), e3002829. https://doi.org/10.1371/journal.pbio.3002829. Article.

Rugg, M.D., Johnson, J.D., Uncapher, M.R., 2015. Encoding and retrieval in episodic memory: Insights from fMRI. In: Addis, D.R., Barense, M., Duarte, A. (Eds.), The Wiley Handbook on the Cognitive Neuroscience of Memory, 1st ed. Wiley, pp. 84–107. https://doi.org/10.1002/9781118332634.ch5.

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

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. Nature Communications 14 (1), 2299. https://doi.org/10.1038/s41467-023-38046-y. Article.

Shrestha, N., 2020. Detecting multicollinearity in regression analysis. American Journal of Applied Mathematics and Statistics 8 (2), 39–42. https://doi.org/10.12691/ajams-8-2-1.

Siestrup, S., Jainta, B., El-Sourani, N., Trempler, I., Wurm, M.F., Wolf, O.T., Cheng, S., Schubotz, R.I., 2022. What happened when? Cerebral processing of modified structure and content in episodic cueing. Journal of Cognitive Neuroscience 34 (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. Journal of Cognitive Neuroscience 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. Learning & Memory 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 in Neurosciences 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. Proceedings of the National Academy of Sciences 118 (51), e2117625118. https://doi.org/10.1073/pnas.2117625118. Article.

Stawarczyk, D., Wahlheim, C.N., Zacks, J.M., 2023. Adult age differences in event memory updating: The roles of prior-event retrieval and prediction. Psychology and Aging 38 (6), 519–533. https://doi.org/10.1037/pag0000767.

Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.-M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. Nature Neuroscience 11 (9), 1004–1006. https://doi.org/10.1038/nn.2163.

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.

Todorovic, A., Van Ede, F., Maris, E., De Lange, F.P., 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. Journal of Neuroscience 31 (25), 9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011.

Van Pelt, S., Heil, L., Kwisthout, J., Ondobaka, S., Van Rooij, I., Bekkering, H., 2016. Beta- and gamma-band activity reflect predictive coding in the processing of causal events. Social Cognitive and Affective Neuroscience 11 (6), 973–980. https://doi.org/10.1093/scan/nsw017.

Varga, D.K., Raykov, P.P., Jefferies, E., Ben-Yakov, A., Bird, C.M., 2025. Hippocampal mismatch signals are based on episodic memories and not schematic knowledge. Proceedings of the National Academy of Sciences 122 (34), e2503535122. https://doi.org/10.1073/pnas.2503535122. Article.

Wang, L., Schoot, L., Brothers, T., Alexander, E., Warnke, L., Kim, M., Khan, S., Hämäläinen, M., Kuperberg, G.R., 2023. Predictive coding across the left fronto-temporal hierarchy during language comprehension. Cerebral Cortex 33 (8), 4478–4497. https://doi.org/10.1093/cercor/bhac356.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—Again. Neuroimage 2 (3), 173–181. https://doi.org/10.1006/nimg.1995.1023.

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.

Yarkoni, T., Speer, N.K., Zacks, J.M., 2008. Neural substrates of narrative comprehension and memory. NeuroImage 41 (4), 1408–1425. https://doi.org/10.1016/j.neuroimage.2008.03.062.

Yi, H.G., Leonard, M.K., Chang, E.F., 2019. The encoding of speech sounds in the superior temporal gyrus. Neuron 102 (6), 1096–1110. https://doi.org/10.1016/j.neuron.2019.04.023.

Zöllner, C., Klein, N., Cheng, S., Schubotz, R.I., Wolf, O.T., 2021. Where was the toaster? Interplay of episodic memory traces and semantic knowledge during scenario construction. PsyArXiv. v. https://psyarxiv.com/2kmwy.