NeuroImage 332 (2026) 121898

Contents lists available at ScienceDirect

NeuroImage

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

To update or to separate: Neural signatures and consequences of latent cause inference in episodic memory

Marius Boeltzig [a,b,*], Sophie Siestrup [a,b], Inês Bramão [c], Ricarda I. Schubotz [a,b]

[a] *Department of Psychology, University of Münster, Fliednerstraße 21, 48149 Münster, Germany*
[b] *Otto Creutzfeldt Center for Cognitive and Behavioural Neuroscience, University of Münster, Germany*
[c] *Department of Psychology, Lund University, Allhelgona Kyrkogata 16a, 223 62 Lund, Sweden*

ARTICLE INFO

Keywords:
prediction error
episodic memory
latent cause theory
memory updating
memory modification

ABSTRACT

Episodic memories are leveraged to predict upcoming input. Latent Cause Theory (Gershman et al., 2017) predicts that after moderate prediction errors, existing memories will be updated by integrating new information, while after large prediction errors, a new latent cause is inferred, and the unpredicted event is encoded as a separate trace. To date, no study has directly compared memory updating with new latent cause inference in neural signatures and representational consequences. This study used naturalistic dialogues, being either coherent within the same latent cause or involving an unpredicted change of topic, causing new latent cause inference. Additionally, it was varied whether dialogues were familiar or novel, with only the former allowing the formation of high-precision predictions. While all prediction errors engaged a core network of inferior frontal and superior temporal regions, as well as Crus 2, updating existing memories with new information additionally relied on frontal midline regions (BA9 and vmPFC). Neural substrates of new latent cause inference depended on whether a precise prediction had been possible: Low-precision predictions evoked broad processing and information gathering, while high-precision predictions instead lead to a top-down adherence to the prediction. These findings suggest that updating memories with new information necessitates higher-order frontal engagement and highlight how prediction precision shapes prediction error processing. More broadly, the findings demonstrate how episodic memory representations are reorganized in response to unpredicted events, distinguishing between updating an existing trace or by forming a new, separate one.

1. Introduction

Episodic memory is not only a means to recall the past, but also a substantial resource for predicting the future (Addis et al., 2007). Beyond deliberate future thinking, where we intentionally imagine specific events in the future (Cole & Kvavilashvili, 2021), predictive use of memory occurs automatically and without conscious awareness, as the brain predicts upcoming sensory input on a short time scale according to the predictive coding framework (Bubic et al., 2010). When such predictions fail, so-called prediction errors (PEs) are elicited, signalling the need to explore the environment and update existing models or acquire new ones (Brown & Brüne, 2012; Friston & Kiebel, 2009). This can lead to the encoding of the unexpected event (Brod et al., 2018; Frank et al., 2020; Greve et al., 2017; Kalbe & Schwabe, 2020), the weakening of existing memories (Kim et al., 2014, 2019), and combinations between the two (Siestrup et al., 2022; Sinclair & Barense, 2018), accompanied by different representational changes (Bein et al., 2023; Boeltzig, Liedtke, Siestrup, et al., 2025). The fact that prediction usually occurs without conscious awareness (Rowe et al., 2020) makes PEs a powerful mechanism for the modification of episodic memories.

According to Latent Cause Theory (Gershman et al., 2017), the exact impact of PEs on memory depends on PE size. Specifically, moderate PEs lead to integration of the unpredicted event into the current model (which can be fed by episodic memories). Larger PEs, in contrast, lead to the inference of a new latent cause: Instead of integrating new information into the existing model, they prompt a switch to a different model capable of explaining the unexpected event better. If no such alternative model is available, a new one must be formed. Thereby, the brain structures experience into latent causes by weighing when to integrate versus separate events (Bein et al., 2023), producing an

---

* Corresponding author at: Department of Psychology, Fliednerstraße 21, 48149 Münster, Germany.
E-mail addresses: marius.boeltzig@uni-muenster.de (M. Boeltzig), s.siestrup@uni-muenster.de (S. Siestrup), ines.bramao@psy.lu.se (I. Bramão), rschubotz@uni-muenster.de (R.I. Schubotz).

https://doi.org/10.1016/j.neuroimage.2026.121898
Received 9 December 2025; Received in revised form 13 February 2026; Accepted 30 March 2026

optimally structured model of the environment. In the current study, we aimed to investigate how updating via integration after moderate episodic PEs and latent cause separation after large episodic PEs differ from each other in neural signatures as well as representational and mnemonic consequences.

Latent Cause Theory (Gershman et al., 2017) is based on modelling work of Pavlovian conditioning paradigms. In these cases, the models used for predictions are stimulus-outcome mappings, such as a conditioned relation between a tone and a shock. However, internal models are supposed to allow reliable predictions for any sensory input and thus need to reflect more complex regularities and experiences as well (Brown & Brüne, 2012; Schubotz, 2007). Consequently, prediction has been suggested to be strongly based on memory (Reichardt et al., 2020), and both semantic (Frank et al., 2018; Greve et al., 2019; van Kesteren et al., 2012) and episodic memory (Kim et al., 2014, 2017; Long et al., 2016; Varga et al., 2025; Wahlheim et al., 2021) has been demonstrated to be used for predicting. Basing predictions on episodic memories may be the most effective solution in highly repetitive situations, in unstable contexts where no regularities can be extracted (Kim et al., 2019), or in novel situations where no schema has been formed yet.

Latent Cause Theory may provide a useful framework for thinking about consequences of PEs, even when they are based not on conditioned associations, but on episodic memories. Specifically, two parallels can be drawn between what is already known about episodic PEs and the principles specified in Latent Cause Theory (Gershman et al., 2017). First, despite the higher complexity of episodic memory, PEs that signal mismatches compared to previous experiences are being generated (Jainta et al., 2022; Siestrup & Schubotz, 2023; Sinclair et al., 2021). In fact, the hippocampus, which is often described as the prime mismatch detector (Barron et al., 2020; Chen et al., 2015), selectively responded to changes to an encoded episode, but not to violations of schema in a recent study (Varga et al., 2025). Second, different studies have argued either for integration (Greve et al., 2018; Wahlheim & Zacks, 2025) of predicted and unexpected episodes, or for their separation (Bein et al., 2021; Kim et al., 2017). As integration and separation are stipulated within Latent Cause Theories after PEs of different sizes, it is intriguing that episodic PE size could have similar effects.

Based on these parallels, recent research has found support for the principles of Latent Cause Theory (Gershman et al., 2017) even when predictions are based on episodic memories. Consistent with an updating process, where new information is integrated into an existing memory, moderate PEs have been shown to lead to decreased source and recognition memory performance for both predicted original and unpredicted new versions of an episode (Boeltzig, Liedtke, & Schubotz, 2025), accompanied by parahippocampal activity (Liedtke et al., 2025). After large PEs however, the original memory trace has been shown to be left more representationally unchanged (Boeltzig, Liedtke, Siestrup, et al., 2025), and both source and recognition memory of the predicted original and unpredicted new version are enhanced (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025). This indicates that the unpredicted new version is encoded distinctly from the original one, which is consistent with the inference of a new underlying latent cause.

These neural and behavioural results support the ideas of PE size-dependent integration or separation when predictions are based on episodic memories. However, these studies using continuous manipulations of PE size have not allowed directly comparing neural processes of model updating and new latent cause inference. To fully assess consequences of moderate, compared to large PEs, it is crucial to understand these mechanisms within episodic prediction.

Latent Cause Theory focuses on the *accuracy* of the prediction as the key determinant of PE size, and thereby of whether the current model is updated or a new latent cause is inferred. However, predictive coding accounts emphasize that both accuracy and the *precision* of the prediction jointly determine the effective influence of a PE: accuracy reflects how close the prediction is to the actual input, whereas precision denotes the confidence or sharpness of the prediction itself (Henson & Gagnepain, 2010). Consequently, PEs are strongest when accuracy is low and precision of the prediction is high (Greve et al., 2017; Henson & Gagnepain, 2010). In real-life cognition, predictions vary widely not only in accuracy but also in precision: they can stem from detailed and specific episodic memories (Haque et al., 2020; Liedtke et al., 2026; Nolden et al., 2025; Stawarczyk et al., 2020) or from more abstract and general rules and schemas (Frank et al., 2020; Kalbe & Schwabe, 2020; Pupillo et al., 2023; Zöllner et al., 2022). While it has been shown that the interplay between prediction precision and accuracy shapes memory outcomes (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025), how they influence model updating and latent cause inference is hitherto unclear.

To directly compare neural signatures and consequences of memory updating and new latent cause inference and examine the influence of prediction precision in the process, the present study therefore orthogonally manipulated both prediction precision (low-precision vs. high-precision) and accuracy (PEs occurring within- vs. across-cause, i.e. leading to the inference of a new latent cause or not).

Imagine that you are at a dinner party and witness a conversation between two people you do not know (Fig. 1). Once one guest asks another about their profession, your brain will automatically form low-precision and unspecific predictions, clustered around possible jobs. The response that the person is a teacher will therefore confirm a vague prediction, situated within the same latent cause, thus eliciting a *low-precision*

within-cause PE. However, if the response was that the finger food is excellent, this would not have been even vaguely predicted: the person might have changed topics or talked to somebody else. This *low-precision across-cause PE* would trigger new latent cause inference, and thus shift models (e.g., from professions to food). To examine new latent cause inference after low-precision predictions, we contrasted these smaller within-cause (stemming from the same predictive model) and larger across-cause (inducing a change in models) PEs with each other in the analysis.

Low-precision predictions give way to high-precision ones if you have already encoded an episodic memory where the person talks about being a teacher. If they however get asked the same question again by another person and now respond that they are a musician, this would be unpredicted. Nevertheless, this *high-precision within-cause PE* should trigger model updating instead of new latent cause inference as the new information can be integrated into the previous memory: the person could well be a teacher *and* a musician. If the strong prediction that the person is a teacher clashes with the statement about the finger food, this *high-precision across-cause PE* should again trigger new latent cause inference. We contrasted both high-precision PEs with a situation where high-precision predictions are exactly confirmed (no PE) and also compared them to each other.

In terms of neural signatures, we predicted that low-precision across-cause, high-precision across-cause, and high-precision within-cause PEs would elicit activity in the inferior frontal gyrus, which is robustly implicated in processing of PEs across domains (El-Sourani et al., 2020; Fujitani et al., 2024; Jainta et al., 2022; Schliephake et al., 2021), and the superior temporal gyrus and sulcus, with their role in high-level semantic processing and responding to structural PEs in complex narratives (Liebenthal et al., 2014; Liedtke et al., 2025; Siestrup & Schubotz, 2023). Model updating after high-precision within-cause PEs, compared to new latent cause inference, is a more cognitively demanding process where novel input needs to be integrated into existing models, which we expected to be reflected in more frontal activation. In particular, we expected BA9 to be more active, which has previously been implicated in integrating novel and unexpected statements into a coherent causal situation model (Ferstl et al., 2005; Ferstl & Cramon, 2001).

Turning to the consequences of the PEs, we predicted better memory for novel information, as well as less original forgetting after larger PEs,

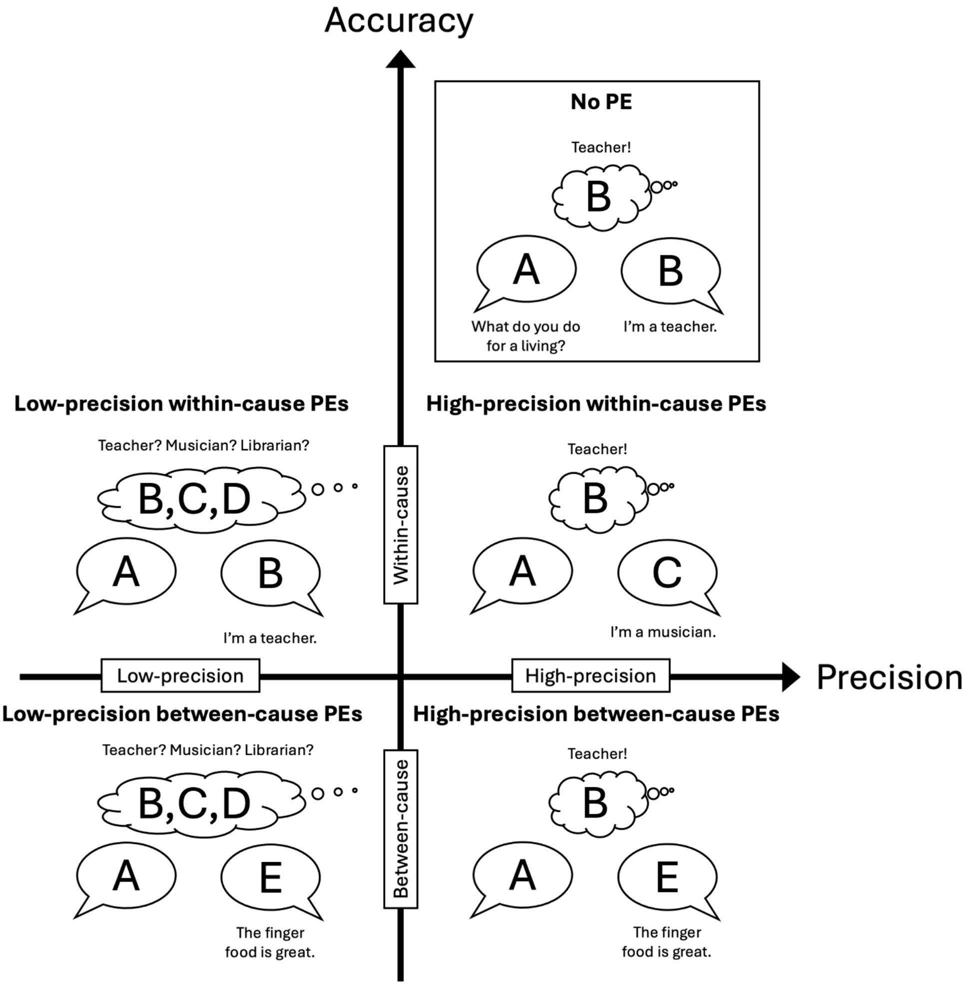

Figure 1. Schematic depiction of the PEs under study.
Note. By manipulating the two dimensions prediction precision and accuracy, we evoked four PEs, and a no PE condition. This figure illustrates them in a conversation where one person asks about the job of another (left speech bubble; A), allowing the formation of predictions (in the thought bubble), which is then compared to the answer (right speech bubble). Low-precision predictions (B, C, D) are made at the first exposure to a novel stimulus, based on several relevant episodic memories and semantic knowledge. Predictions are high-precision (B) when there is already an episodic memory that can be used for prediction. Within-cause PEs arise when the response of the person is within the predictable, so pertaining to the currently active latent cause (B or C). Across-cause PEs arise when a new latent cause has to be inferred, as the response is highly unpredicted within the situation (E).

so after across-cause compared to within-cause PEs (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025; Liedtke et al., 2025). Additionally, the latent cause separation triggered by low-precision across-cause PEs should lead to the parts of these stimuli being represented more separately, as they are deemed less connected and therefore less relevant to each other, compared to low-precision within-cause PEs. This is akin to an event segmentation process, where continuous experience is parsed into meaningful units (Nolden et al., 2024; Zacks & Swallow, 2007), which can be triggered by new latent cause inference (Gershman et al., 2017). Lastly, we assessed reactive reinstatement, reflecting the retrieval of previously encoded original versions of a stimulus while a new mismatching one is processed. We predicted that during high-precision across-cause PEs, such reinstatement is stronger than during high-precision within-cause PEs, reflecting more top-down activity.

2. Methods

2.1. Participants

Participants were recruited in the student community of the University of Münster. Only right-handed native German-speaking adults with normal or corrected-to-normal vision and hearing, no current psychological illness, and meeting all requirements to enter an MRI scanner could participate. They were compensated with course credit or money.

After exclusion of two participants with excessive movement in the scanner and one participant failing to reach above-chance levels on the recognition test, the final sample size was $N = 44$. In a previous similar study (Boeltzig, Liedtke, Siestrup, et al., 2025; Liedtke et al., 2025), we had included 42 participants, which was therefore the minimal requirement for the present study as well. Mean age in the final sample was 22.61 ($SD = 3.40$, range: 18 – 32), and there were 33 women, eight men, and three diverse participants. All but one participant were students. All participants gave written informed consent after being briefed on their rights. The Ethics Committee of the Faculty of Psychology and Sports Science at the University of Münster approved the study procedure.

2.2. Design

We manipulated two orthogonal dimensions to produce the four situations described in the Introduction and depicted in Fig. 1. *PE precision* was varied through familiarity to the material: When a given stimulus was presented for the first time, participants could make only low-precision predictions, and when the stimulus was already familiar (at third exposure) predictions were high-precision. To induce truly high-precision predictions, dialogues were played in their original form twice to allow solid encoding. The second exposure to the dialogues was not investigated in the comparison of neural signatures of the PEs in the manuscript, but is reported in the Supplement. *PE accuracy* was manipulated within the stimulus material, with dialogues being presented either in a coherent version, where two statements were logically related, or an incoherent version where they were not (see 2.3 Material). Additionally, there was a no PE condition with stimuli that were never altered.

2.3. *Material*

For this study, we adapted 60 previously used naturalistic dialogues between two speakers, discussing topics and events likely to resonate with a student sample (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025; Liedtke et al., 2025). One example is provided in Fig. 2A, four additional examples can be found in the Supplementary Materials S1, and the full transcribed material is available on OSF, along with one sample of the auditory stimuli that participants were exposed to (https://osf.io/mgu49/overview?view_only=1af278283f49433d8de09f566a2dff97). The dialogues were recorded by 20 professional speakers and paired with a unique background sound. For this study, the dialogues were shortened so that they consisted of two parts. The cue consisted of one or more sentences by one speaker,

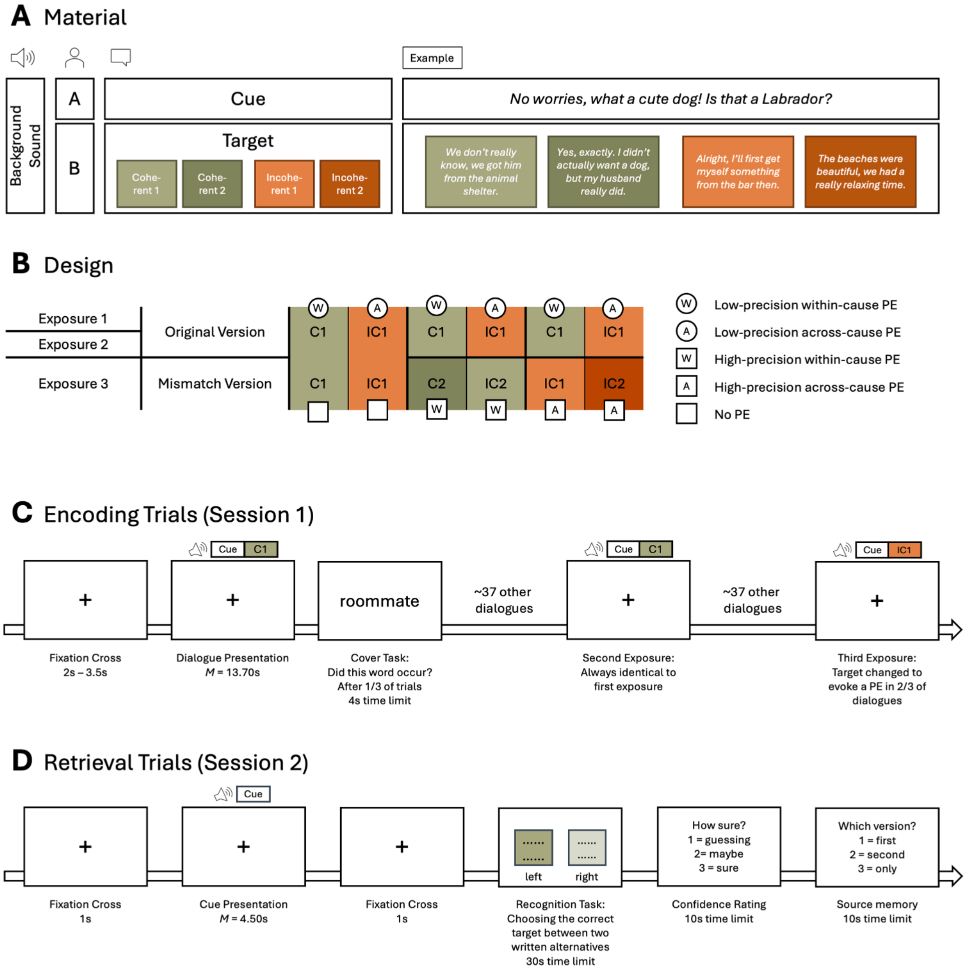

Figure 2. Experimental Procedure.

Note. **A.** Dialogues were comprised of two statements, one by each speaker. The statement of the first speaker, the cue, was always the same, but there were four alternative targets. Two of these targets were coherent with the cue, directly responding to it. Two targets were incoherent, not directly responding to the cue, inducing new latent cause inference. In addition to the provided translated sample dialogue, there are more examples in the Supplementary Materials S1 and on OSF. **B.** There were six types of dialogue histories throughout the experiment. Dialogues were always played twice in an original version that could be either coherent or incoherent. Originally coherent dialogues were believed to induce low-precision within-cause PEs, and originally incoherent dialogues induced low-precision across-cause PEs, and therefore new latent cause inference. A third of all dialogues (two leftmost columns) were not changed, but instead played again in their original form at the third exposure, so that the original and mismatch versions were identical. The other two thirds of dialogues were changed, again either being coherent or incoherent. Thus, at third exposure, high-precision within-cause PEs were evoked by a change to a novel but coherent dialogue version, and high-precision across-cause PEs were induced by a novel incoherent version. **C.** Encoding trials started with a jittered fixation cross, followed by the presentation of a dialogue, during which a fixation cross stayed on the screen. After one third of trials (once for each dialogue), a word was displayed and participants had to respond whether the word was said in the dialogue. Each dialogue was played three times, interleaved with a variable number of other dialogues, and two thirds of dialogues were modified upon third exposure. **D.** The memory test comprised one trial for the unchanging dialogues, and two trials for the two versions of the changing dialogues. After a fixation cross, the cue was presented auditorily. After another fixation cross, two written options were presented, and participants had to choose which one was actually said in response to the cue. The other option was a similar lure. Participants also rated confidence in their choice and indicated whether the just-probed version was the first, second, or only one they had heard of that specific dialogue.

and the target comprised one or more sentences by another. Crucially, while the cue always stayed the same, the targets could be exchanged, producing four dialogue versions. All four targets matched to a given cue involved the same speaker so that a mismatch would not be produced based on a different voice.

Coherence of the dialogues was manipulated in order to evoke across-cause PEs. Two versions of the dialogue were coherent: Here, the target was a direct and logical response to the cue. The two coherent versions however differed in meaning; for instance, the speaker would agree in one, but disagree in the other coherent target. Additionally, there were two incoherent versions, where the target was not a logical response to the cue – prompting new latent cause inference. We attempted to avoid any connection between cues and incoherent targets as well as between the two incoherent targets for the same dialogue. The mean duration of the cues was 4.50 seconds (range: 2s – 8s), and of the targets 9.20 seconds (range: 5s – 15s).

One third of the dialogues was never changed during the experiment and thus played in the same version three times. The other two thirds of dialogues were presented in an original form twice, and then changed upon the third exposure, meaning that a different target was played in response to the same cue. Half of the dialogues initially had a coherent original, the other half had an incoherent original. Within the changing dialogues, half of each group then had a coherent mismatch (i.e., new version), and the other half an incoherent mismatch. There were therefore six dialogue types (Fig. 2B): (1) unchanging coherent, (2) unchanging incoherent, (3) coherent original, coherent (but new) mismatch, (4) incoherent original, coherent mismatch, (5) coherent original, incoherent mismatch, (6) incoherent original, incoherent (but new) mismatch. There were ten dialogues in each of these conditions. Note that which of the two coherent or incoherent targets would be the original and which one the mismatch version was counterbalanced across participants. Assignment of dialogues to conditions was also counterbalanced.

For the memory test in Session 2, similar lures were produced to be presented in written form alongside the actual statement from the recorded dialogue. To that end, one word or short phrase in the original version was substituted by a synonymous word or short phrase. High similarity between real target and lure was ensured using Sentence BERT (Reimers & Gurevych, 2019) and all statement pairs had semantic similarity scores lower than .99, but higher than .90. These values had no statistically significant impact on recognition performance.

2.4. Procedure

The study was conducted in two sessions, with Session 2 taking place two days after Session 1. Session 1 was conducted in the MRI scanner and was dedicated to encoding and modifying the material, inducing PEs both at first encoding and when a familiar dialogue was changed. Session 2 was purely behavioural and comprised the memory test. The sessions lasted around 75 and 45 minutes, respectively.

In Session 1 (Fig. 2C), participants were told that we were interested in how people perceive conversations and were asked to listen to the dialogues carefully, and to imagine the scene in which they happened. There was no explicit encoding instruction. After self-adjustment of the volume during an EPI sequence and two practice trials, participants listened to the dialogues in the MRI scanner. The task was divided into six blocks, each of them comprising 30 dialogues. After one third of the dialogues (once for each dialogue across its three exposures), there was a cover task in which participants were asked whether a certain word had occurred in the dialogue, to which they responded with a button press (4s time limit). After each block, feedback about the percentage of correct responses was provided to increase motivation.

Each dialogue was played three times, while the changing dialogues were altered upon the third exposure. The material was arranged in an auditory stream so that new dialogues were continuously introduced, the crucial third exposure was spaced out across the whole experiment, and a variable number ($M = 37.44$, $SD = 18.36$, range: 7-94) of other dialogues was placed between any two exposures to the same one.

The trials themselves consisted of a jittered fixation cross (2 – 3.5s), the exposure to the dialogue while a fixation cross remained on the screen, and then either the cover task or the next trial.

Two days later, participants returned to complete a surprise memory test, seated at a computer (Fig. 2D). For each of the changing dialogues, both the original version (from the first two exposures), and the mismatch version (from the third exposure) were tested, while for the unchanging dialogues, only the one original version was tested, leading to a total of 100 trials. Whether original or mismatch version of a dialogue was tested first was counterbalanced across dialogues and the two related trials of the changing dialogues were spread out across different blocks. Participants were informed that some dialogues would be tested twice, reflecting the different versions that they had heard. The task was divided into four blocks, with a short break after each one.

In each trial, after a fixation cross (2s), participants listened to the cue of the dialogue, followed by a fixation cross (1s). They then saw two written statements on the screen, one was the original target, and the other was a similar lure that had not been said. Participants chose the correct version using the arrows on the keyboard (30s time limit) and location of the correct answer was pseudo-randomized. After their response, they indicated confidence using numbers (1 = guessing, 2 = maybe, 3 = sure; 10s time limit). Finally, as a measure of source memory, they were asked to indicate whether this dialogue was the original version which they had heard first, the second version which they only heard later, or the only version of that specific dialogue (10s time limit). This was illustrated with an example and the experimenter made sure that participants understood this task.

2.5. MRI data acquisition and preprocessing

Session 1 was conducted in a 3-Tesla Siemens Magnetom Prisma MR tomograph with a 20-channel head coil. Participants were placed supine on the scanner bed, wearing ear plugs and over-ear headphones, with cushions around the head to avoid movement. Responses to the cover task were given using the right index and middle finger on a button box.

First, we acquired a high-resolution T1-weighted anatomical image using a 3-D magnetization rapid gradient echo sequence (192 slices of 1mm, repetition time = 2140ms, echo time = 2.28 ms, flip angle = 8°, field of view = 256×256 mm^2). Afterwards, during the task, functional images were obtained along the AC-PC plane in interleaved order using a gradient-echo planar imaging sequence (33 slices of 3mm, repetition time = 2000ms, echo time = 30 ms, flip angle = 90°, field of view = 192×192 mm^2).

The resulting data was pre-processed with SPM 12 (Wellcome Trust) in MatLab. We applied slice time correction to the middle slice, movement correction and realignment to the mean function image, co-registration of structural and functional images, normalization of all images to Montreal Neurological Institute (MNI) space, and a temporal high-pass filter of 128s to remove low-frequency noise from the time series. For the univariate analysis, we also applied spatial smoothing with a Gaussian kernel of 8mm. For the representational similarity analysis (RSA), we followed the recommendation not to apply smoothing if regions like the hippocampus are among the ROIs (Dimsdale-Zucker & Ranganath, 2018; Weaverdyck et al., 2020).

2.6. Behavioural data analysis

First, we tested whether any covariates arising from the procedure influenced recognition performance. Age, gender, sleep duration, guessing the intention of the experiment, serial position of dialogues, cover task performance, test order (original vs mismatch first), semantic similarity between target and lure, and duration of the cue did not significantly predict recognition performance (see Supplementary Materials S2) and where therefore not controlled for in further analyses.

For the behavioural data analysis, we first tested how coherence generally affected memory, comparing the unchanging coherent and incoherent dialogues. Finally, using only changing dialogues, we tested the effect of the PEs of interest on memory. In the models for memory for the original and mismatch versions, coherence of both original and mismatch were allowed to interact. All of these analyses were conducted separately for recognition memory, and confidence and source memory for only correct recognition responses. Linear mixed models were used, including participants and dialogues as random intercepts. All analyses were carried out in R (R Core Team, 2025) with *tidyverse* (Wickham et al., 2019), using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages for linear mixed models.

2.7. Univariate MRI analysis

The univariate analyses were set up using generalized linear models (GLMs) in SPM12 and always included the six rigid-body transformation from spatial realignment. They were convolved with a canonical hemodynamic response function in SPM12. A normalized individual grey matter mask that was smoothed at 8mm and thresholded at 0.2 was applied on the first level and *t*-contrasts were calculated between relevant conditions (see below). On the second level, one-sample *t*-tests were used across participants. An FDR-correction of $p < .005$ was used. All relevant regressors were modelled as epochs comprising the whole target.

We ran two models, which were different in that only the first model always accounted for dialogue history, so the coherence at first exposure, even when modelling responses to induced PEs at the third exposure. In addition to button presses in response to the cover task, we modelled (1) low-precision within-cause PE (i.e., first exposure to coherent dialogues), (2) low-precision across-cause PE (i.e., first exposure to incoherent dialogues), (3) second exposure to coherent dialogues, (4) second exposure to incoherent dialogues, (5) high-precision within-cause PE (i.e., third exposure to a dialogue with a novel coherent target) after originals had been coherent, (6) high-precision within-cause PE (i.e., third exposure to a dialogue with a novel coherent target) after originals had been incoherent, (7) high-precision across-cause PE (i.e., third exposure to a dialogue with a novel incoherent target) after originals had been coherent, (8) high-precision across-cause PE (i.e., third exposure to a dialogue with a novel incoherent target) after originals had been incoherent, (9) no PE (i.e., third exposure to unchanging dialogue) with coherent originals, (10) no PE (i.e., third exposure to unchanging dialogue) with incoherent originals.

Contrasts between incoherent and coherent dialogues at each time point and within each PE showed that those evoked significantly different brain regions at first and second exposure, but not anymore at third exposure (FDR-corrected $ps > .005$). This was true both for the unchanged dialogues, and for the dialogues that were changed to evoke high-precision PEs.

Due to this lack of differences, we therefore ran a simplified model which is reported in the manuscript. In this model, the dialogue history is not distinguished at third exposure. In addition to the button press during the cover task, we therefore modelled (1) low-precision within-cause PE (i.e., first exposure to coherent dialogues), (2) low-precision across-cause PE (i.e., first exposure to incoherent dialogues), (3) second exposure to coherent dialogues, (4) second exposure to incoherent dialogues, all of which were also part of the first model. Additionally, we modelled (5) high-precision within-cause PEs (i.e., third exposure to a dialogue with a novel coherent target), (6) high-precision across-cause PEs (i.e., third exposure to a dialogue with a novel incoherent target), and (7) no PE (i.e., third exposure to unchanging dialogue). These three latter regressors were collapsed across originally coherent and incoherent dialogues.

We report results from four contrasts in the manuscript. First, the contrast low-precision across-cause PE > low-precision within-cause PE compared coherent and incoherent dialogues at first exposure. Second, high-precision within-cause PE > no PE compared dialogues modified with a new and coherent target at third exposure with those dialogues that were played in the original form for a third time. Third, high-precision across-cause PE > no PE compared dialogues with a new but incoherent target with the unchanged dialogues already used in the previous contrast. Fourth and last, the contrast high-precision within-cause PE > high-precision across-cause PE was tested as a direct comparison between the two high-precision PEs induced at third exposure. Additionally, in the Supplementary Materials S3, we report activity during the second exposure to the original dialogues, contrasting coherent and incoherent dialogues.

2.8. Representational similarity analysis

In addition to the univariate analysis, we conducted Representational Similarity Analysis (RSA). To that end, we specified a GLM in which each cue and each target from each dialogue exposition constituted its own regressor, modelled as an epoch. This was then used to conduct RSA using the CosmoMVPA toolbox (Oosterhof et al., 2016), in which each beta image was correlated with all others, while applying demeaning, producing a 360 × 360 Pearson correlation matrix. This was done separately for the whole brain, but also for the regions of interest hippocampus and inferior frontal gyrus (IFG) comprising the pars opercularis, pars orbitalis, and pars triangularis. These ROIs were obtained by applying the Freesurfer (Fischl, 2012) parcellation to the MNI standard brain.

In the resulting similarity matrices, we extracted the pre-specified correlations of interest (see below). We then corrected the correlations by subtracting the baseline similarity from them. For instance, for the measure of pattern separation, where the target from the first exposure was correlated with the cue from the second exposure within a given dialogue, we subtracted from that value the mean of correlations of the same target with all other cues at second exposure, as well as the same cue with all the other targets at first exposure (Boeltzig, Liedtke, Siestrup, et al., 2025).

We thus extracted three representational similarities of interest. First, we measured how similarly the cues and targets of each dialogue were represented the first time that participants listened to them. We used this as a measure of coherence, assuming that coherent dialogues should have higher neural similarity between cue and target than incoherent dialogues. This measure was therefore used as a manipulation check.

Second, we investigated whether new latent cause inference would induce pattern separation between cues and targets, which was tested for the low-precision PEs induced at first exposure. Specifically, we tested whether after across-cause PEs, inducing new latent cause inference, the targets would be less used for prediction when the same stimulus was played next (at second exposure). Less use for prediction would indicate a less integrated and more separated way of representing cue and target, which would be deemed less relevant for each other. This use for prediction was tapped by correlating the targets from the first exposure with the cue of the second exposure. Due to the implication of the inferior frontal gyrus (IFG) in holding a predictive model (Fujitani et al., 2024), this analysis was carried out for this ROI in addition to the whole brain.

Third, while the previous measure reflects the predictive reinstatement of a dialogue target, we also analysed original reactivation at mismatch, reflecting a reactive reinstatement in response to the detection of a PE. This reactivation was operationalized as the similarity between the original target from the second exposure and the mismatch target from the third exposure. While this similarity reflects basic perceptual similarity, it also captures additional retrieval of the original version while a new and mismatching version is processed (Boeltzig, Liedtke, Siestrup, et al., 2025). We tested whether this reactivation would be stronger when new mismatching information caused an across-cause PE, using the hippocampus as the ROI.

Linear mixed models were used to test these hypotheses. For initial coherence and pattern separation induced by low-precision PEs, only coherence of the original dialogue entered the model. For original reactivation, the coherences of the original and new mismatching dialogues were modelled. In all cases, participants and dialogues were added as random intercepts.

3. Results

3.1. *Manipulation check and overall performance*

The RSA confirmed that the dialogues that we had constructed as incoherent were less similarly represented than coherent dialogues. Specifically, upon first exposure, cue and target were represented less similarly for incoherent than coherent dialogues in the whole brain, hippocampus, and inferior frontal gyrus (βs $>$ .33, *ps* $<$ .001). We therefore assumed that material construction was successful.

Performance on the cover task was good, indicating high levels of attention (*M* = .85*, SD* = .08, range: .65 – 1). Recognition performance generally was robust, and all participants performed significantly above chance (*M* = .71, *SD* = .06, range: .60 - .84). Source recognition, where guessing probability was at 33%, was expectedly lower and more mixed (*M* = .48, *SD* = .11, range: .29 - .77).

3.2. *Processing of PEs*

To compare neural processes at new latent cause inference with model updating, we ran univariate brain analyses, using the four PEs created by the experimental design.

3.2.1. *PEs after low-precision predictions*

Firstly, we contrasted the two low-precision PEs with each other by comparing activity at the first exposure to incoherent dialogues (low-precision across-cause PEs) with activity at the first exposure to coherent dialogues (low-precision within-cause PEs). Given that the dialogues were novel at this time point, predictions were low-precision. After across-cause PEs (incoherent dialogues), new latent cause inference should take place, while within-cause PEs should support the vague prediction and were therefore used as the baseline here.

This contrast (Table 1 and Fig. 3A) revealed activation in the bilateral inferior frontal gyrus (IFG), and in the left anterior dorsal insula. Additionally, there were significant clusters in the bilaterial pre-supplementary motor area, left dorsal premotor cortex, and the right posterior middle frontal gyrus. In the temporal lobe, we found significant clusters in the bilateral superior temporal sulcus and gyrus (STS/STG). Additionally, there was activity in the right intraparietal sulcus, the right tempo-parietal junction, the bilateral posterior dorsal precuneus, and the left cerebellum (Crus 2).

A subset of these regions was still activated when the incoherent dialogues were played the second time. These results can be found in the Supplementary Materials S3.

3.2.2. *High-precision across-cause PEs*

Next, high-precision across-cause PEs were compared to a lack of PEs. Both were modelled at third exposure, where a new target was introduced and incoherent with the cue; or, in the no PE condition, the original dialogue was played for a third time in its unaltered original. In both cases, precise predictions could be made based on the previous exposure to the dialogue. Unexpected and incoherent input, however, would cause new latent cause inference.

Interestingly, these PEs only activated a subset of regions that responded to the low-precision across-cause PE. There was bilateral activation in the IFG, and in the bilateral STS/STG, as well as in the left cerebellar Crus 2 (Table 2 and Fig. 3C).

Table 1
Low-precision across-cause PE > High-precision across-cause PE.

Localization	Cluster	H	Cluster Extent	MNI Coordinates x	y	z	*t*
Inferior Frontal Gyrus (BA 45)	1	R	lm	45	35	-4	7.04
Inferior Frontal Gyrus (BA 44)	1	R	lm	51	23	26	6.34
Inferior Frontal Gyrus (BA 44) / IFS	10	L	31	-45	17	23	4.61
Anterior Dorsal Insula	3	L	119	-30	23	-1	6.65
Pre-Supplementary Motor Area	6	R	145	6	20	53	5.72
Pre-Supplementary Motor Area	6	L	lm	-9	11	50	4.25
Posterior Middle Frontal Gyrus	1	R	998	36	8	35	9.36
Dorsal Premotor Cortex	11	L	25	-42	2	59	4.58
Anterior Superior Temporal Sulcus	9	R	69	51	-4	-19	4.97
Posterior Middle Temporal Gyrus	2	R	lm	66	-43	5	6.66
Temporoparietal Junction	2	R	459	48	-46	26	6.84
Posterior Superior Temporal Sulcus / Gyrus	7	L	77	-63	-52	17	5.43
Intraparietal Sulcus	8	R	106	36	-55	44	5.06
Posterior Dorsal Precuneus	5	R	208	9	-61	41	5.83
Posterior Dorsal Precuneus	5	L	lm	-3	-61	41	5.12
Cerebellum (Crus 2)	4	L	143	-12	-79	-37	6.50

3.2.3. *High-precision within-cause PEs*

Lastly, high-precision within-cause PEs were compared to the no PE condition. Both were situated at the third exposure, where a new target was introduced and coherent with the cue; or, in the no PE condition, the original dialogue was played for a third time in its unaltered original. Again, precise predictions were possible here based on previous familiarity to the material. The new and unpredicted input, still being coherent with the prediction, should cause model updating instead of new latent cause inference.

This analysis revealed large-scale activity, which was partly overlapping with the low-precision across-cause PEs (Table 3 and Fig. 3B). Specifically, there was activation in the right IFG, the left posterior medial frontal gyrus, widespread activity in the bilateral STG, and in the bilateral Crus 2 in the cerebellum. There was also activity in the bilateral precuneus, but in comparison to the low-precision across-cause PEs, the cluster was located more ventrally. Additionally, unique clusters not seen in other PEs were observed in the bilateral vmPFC, the left medial and superior frontal gyrus (corresponding to BA 9 and 8), the left superior colliculus, and the left caudate.

3.2.4. *High-precision within-cause PEs vs High-precision across-cause PEs*

We also compared the two PEs at third exposure after high-precision predictions to directly observe differences between new latent cause inference and model updating. Interestingly, the high-precision across-cause PEs, where we assumed the inference of a new underlying latent cause, did not reveal any significant clusters. High-precision within-cause PEs, on the other hand, evoked higher activity in the bilateral A1 (Heschl's gyrus), the bilateral superior frontal gyrus (BA9), the left middle frontal gyrus, as well as the left angular gyrus and left middle temporal gyrus (Table 3, Fig. 3D).

3.3. *Consequences of PEs*

After establishing the neural signatures of the PEs of interest, we specifically focused on their representational and mnemonic

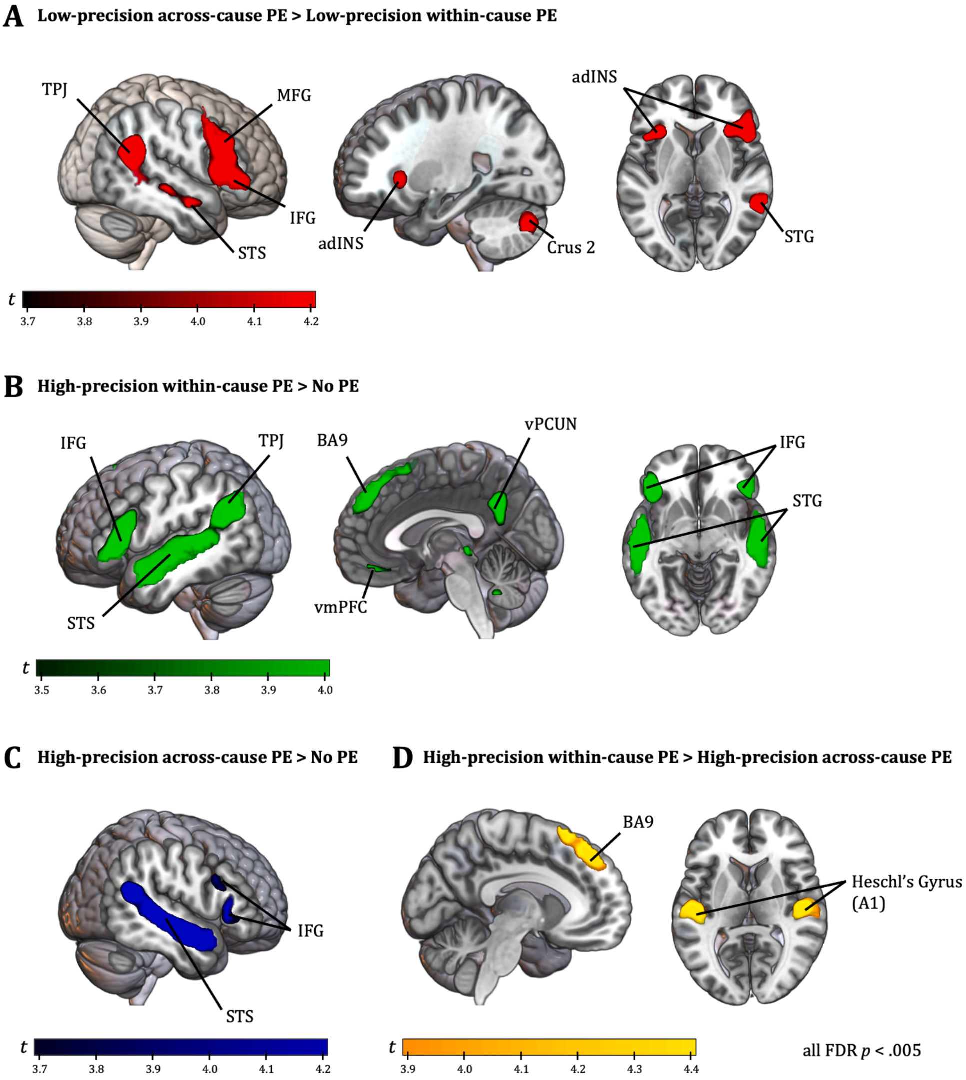

Figure 3. Whole-brain activation for direct contrasts of all PEs.
Note. **A.** Low-precision across-cause and low-precision within-cause PEs were compared by contrasting incoherent and coherent dialogues at first exposure. **B.** To test high-precision within-cause PEs, dialogues that were modified at third exposure with a coherent target were compared to unchanging dialogues at third exposure. **C.** Dialogues that were modified at third exposure using an incoherent target were compared to unchanging dialogues at the third exposure to assess high-precision across-cause PEs. **D.** High-precision within-cause PEs and high-precision across-cause PEs were directly compared as well. The opposite contrast yielded no significant clusters. TPJ = temporoparietal junction, MFG = middle frontal gyrus, IFG = inferior frontal gyrus, STS = superior temporal sulcus, adINS = anterior dorsal insula, STG = superior temporal gyrus, vPCUN = ventral precuneus, vmPFC = ventromedial prefrontal cortex.

Table 2
High-precision across-cause PE > No PE.

Localization	Cluster	H	Cluster Extent	MNI Coordinates			t
				x	y	z	
Inferior Frontal Gyrus (BA 44)	4	L	49	-57	17	20	4.94
Inferior Frontal Gyrus (BA 45)	5	R	158	54	29	8	4.88
Anterior Superior Temporal Gyrus	1	L	869	-54	-10	-10	9.06
Anterior Superior Temporal Gyrus	2	R	808	51	-1	-19	8.79
Superior Temporal Sulcus	1	L	lm	-51	-43	5	6.53
Superior Temporal Sulcus / Gyrus	2	R	lm	51	-25	-4	6.66
Cerebellum (Crus 2)	3	L	39	-18	-79	-37	5.77

consequences.

3.3.1. *The effect of low-precision PEs on integrated representations*

We hypothesized that the inference of a new underlying latent cause within a dialogue, as evident in low-precision across-cause PEs, but not low-precision within-cause PEs, would leave the cue and target of these dialogues representationally more separated instead of integrated. Consequently, and reflecting the lower relevance to the cue, the targets should be predicted less strongly when these dialogues are played again. To test this, we extracted the representational similarity between the target from the first exposure, where either an across-cause (incoherent dialogues) or within-cause (coherent dialogues) PE was induced, and the cue on the subsequent second exposure. This analysis therefore tapped the immediate effect of these PEs on representational structure.

Compared to low-precision across-cause PEs, low-precision within-cause PEs yielded higher similarity between the target at first exposure and cue at second exposure in the whole-brain ROI ($\beta = .10$, $SE = .01$, $p = .013$). In the IFG, this was short of significance ($\beta = .07$, $SE = .01$, $p = .079$). As higher pattern similarity corresponded to higher prediction

Table 3
High-precision within-cause PEs.

Localization	Cluster	H	Cluster Extent	MNI Coordinates			*t*
				x	y	z	
High-precision within-cause PE > No PE							
Medial Frontal Gyrus (BA 9)	3	L	754	-9	47	47	7.97
Superior Frontal Gyrus (BA8)	3	L	lm	-9	38	50	7.14
Inferior Frontal Gyrus (BA 45)	6	R	232	54	32	-4	6.07
Posterior Medial Frontal Gyrus (dorsal premotor cortex)	9	L	35	-39	2	53	4.89
Ventromedial prefrontal cortex	10	L	32	-3	47	-16	4.26
Ventromedial prefrontal cortex	10	R	lm	6	47	-16	4.21
Anterior Superior Temporal Gyrus	1	L	1852	-57	-10	-10	10.78
Posterior Superior Temporal Gyrus	1	L	lm	-54	-40	2	9.47
Temporal Pole	2	R	1082	48	14	-25	9.05
Superior Temporal Gyrus	2	R	lm	57	-25	2	7.88
Anterior Superior Temporal Gyrus	2	R	lm	51	-4	-19	7.12
Temporoparietal Junction	1	L	lm	-57	-64	26	8.50
Temporoparietal Junction	2	R	lm	54	-55	26	5.23
Ventral Precuneus	8	LR	143	0	-55	35	5.19
Cerebellum (Crus 2)	4	L	172	-24	-82	-37	7.36
Cerebellum (Crus 2)	5	R	240	27	-79	-34	7.09
Superior Colliculus	7	L	30	-3	-31	-4	5.31
Caudate	9	L	29	-12	8	14	5.05
High-precision within-cause PE > High-precision across-cause PE							
Heschl's Gyrus (A1)	1	R	141	48	-16	8	6.56
Heschl's Gyrus (A1)	3	L	195	-48	-19	5	5.94
Angular Gyrus	2	L	303	-45	-61	41	6.44
Superior Frontal Gyrus (BA 9)	4	L	388	-12	23	62	5.72
Superior Frontal Gyrus (BA 9)	4	L	lm	-15	44	41	5.66
Superior Frontal Gyrus (BA9)	6	R	22	9	35	59	5.11
Posterior Middle Frontal Gyrus	4	L	lm	-42	17	44	5.42
Middle Temporal Gyrus	5	L	46	-63	-37	-4	5.31

strength, the hypothesis that new latent cause separation evoked by low-precision across-cause PEs was supported on the whole-brain level.

3.3.2. *Original Reactivation during high-precision PEs*

We also predicted that during high-precision across-cause PEs, compared to high-precision within-cause PEs, the originally encoded version would be reactivated more strongly. This would reflect a search for a more appropriate (i.e., belong to the same latent cause) continuation of the stimulus. We therefore extracted the representational similarity between the target from the original dialogues (second exposure) and the target from the mismatch dialogues (third exposure) in the hippocampus.

In line with this prediction, the original targets were reactivated more strongly during high-precision across-cause PEs compared to high-precision within-cause PEs – so when incoming new input was incoherent ($\beta = .14$, $SE = .01$, $p = .035$). This was regardless of whether the original version had induced a within- or across-cause PE ($\beta = .12$, $SE = .01$, $p = .075$; interaction: $\beta = .11$, $SE = .02$, $p = .249$).

3.3.3. *Effects of PEs on memory*

To assess the effects of the PEs on memory, we turned to the behavioural data and analysed recognition memory, as well as confidence and source memory for correct responses, depending on coherence of both the original and the mismatch version.

To first establish whether low-precision PEs at first exposure generally influenced memory, we focused on those dialogues that were never changed. Interestingly, there was no evidence that dialogues with and without a low-precision PE at first exposure were remembered differently well ($\beta = .31$, $SE = .16$, $p = .062$). However, both confidence ($\beta = .37$, $SE = .05$, $p < .001$) and source memory ($\beta = .73$, $SE = .17$, $p < .001$) was better for coherent dialogues, pointing to an advantage for dialogues with a low-precision within-cause PE, compared to a low-precision across-cause PE.

Next, we separately tested memory for the original and the mismatch version. In both cases, we tested whether recognition memory was influenced by coherence of both the original (low-precision PEs) and the mismatch version (high-precision PEs), as well as their interaction.

Memory for the original version (Fig. 4A), so the version that participants heard during the first two exposures to each dialogue, was not affected by coherence of the original or the mismatch, nor by their interaction (*ps* > .173). There was therefore no effect of high-precision or low-precision PEs. However, confidence for original recognition was higher when this original version was coherent, so after low-precision within-cause, compared to low-precision across-cause PEs ($\beta = .38$, $SE = .05$, $p < .001$), while coherence of the mismatch version did not gain significance (*ps* > .380). Interestingly, source memory for the original memory was higher when the mismatch version was coherent and therefore enhanced after high-precision within-cause, compared to high-precision across-cause PEs ($\beta = .41$, $SE = .17$, $p = .017$) – regardless of whether the original itself was coherent or not and with no interaction (*ps* > .451).

For the mismatch version (Fig. 4B), which was introduced upon the third exposure to induce a high-precision PE, recognition memory was once again not affected by any high-precision or low-precision PEs (*ps* > .253). However, confidence and source memory for correct responses showed parallel effects to original memory. Specifically, mismatch targets were recognised more confidently when those mismatch targets were coherent, so after high-precision within-cause PEs ($\beta = .53$, $SE = .06$, $p < .001$).

Additionally, source memory for the mismatch version was higher when the original version had been coherent, so after low-precision within-cause PEs compared to low-precision across-cause PEs ($\beta = .37$, $SE = .18$, $p = .035$), regardless of whether the mismatch version itself was coherent or not and with no interaction (*ps* > .156).

4. Discussion

According to Latent Cause Theory (Gershman et al., 2017), prediction errors (PEs) can trigger either the updating of an existing model with new, unpredicted information if the PE is moderate or, if the PE is large, the inference of a new latent cause, prompting a switch to a new model. The current study, using functional univariate analyses as well as representational similarity analysis, found that these two outcomes differ systematically in terms of neural signatures and representational and mnemonic consequences, which also depend on the precision of the prediction available when the PE occurs. Across all PE conditions, a common network involving inferior frontal, temporal, and cerebellar regions responded robustly to mismatching input. If the mismatching information could be accommodated within the currently active model, the system additionally engaged frontal midline regions to support model updating. If the mismatch instead rendered the current model untenable, a new latent cause was inferred without additional prefrontal control. This inference, however, unfolded differently depending on the precision of the underlying predictions: breaches of low-precision predictions led to a broad bottom-up processing of new information, which

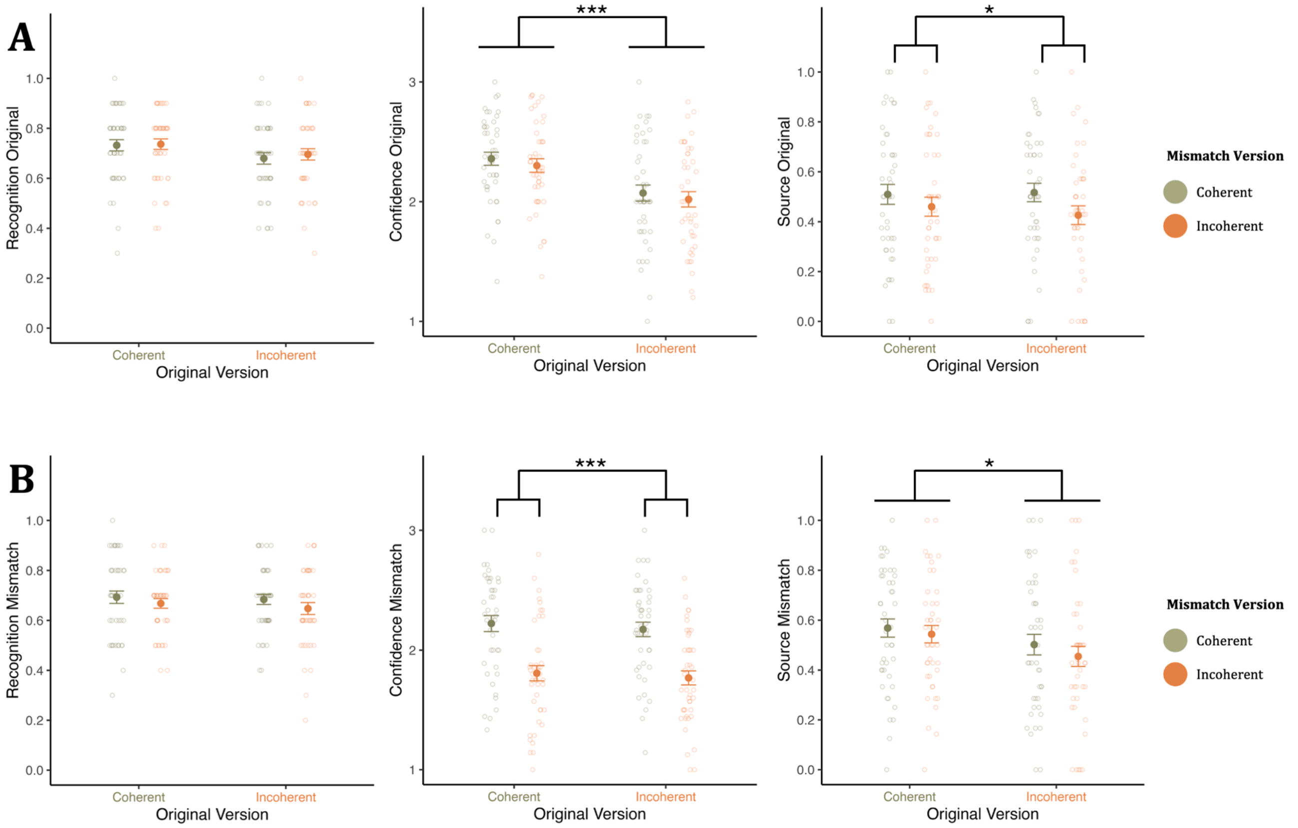

Figure 4. Memory outcomes.

Note. **A.** Memory performance for original versions (played at exposures one and two), depending on whether the original (x-axis) and the mismatch version (coloured) are coherent or incoherent. For confidence for the original, there was an advantage for coherent over incoherent originals, while for source memory for the original, there was an advantage of a coherent mismatching version. **B.** Memory performance for mismatch versions (played at exposure three). Confidence was higher when the mismatch version was coherent, and source memory was better when the original version was coherent.

was encoded separately from its context, reflecting the inference of a new latent cause. In contrast, breaches of specific predictions based on previously encoded episodic memories led to concurrent reinstatement of the prediction instead of bottom-up information gathering.

4.1. General PE response

Before discussing differences between the PEs evoked in the current study, we focus on their commonalities as three regions robustly responded to all PEs, namely IFG, STG/STS, and Crus 2. First, the IFG is involved in PE processing across tasks and contexts (El-Sourani et al., 2020; Li et al., 2019; Schliephake et al., 2021; Wurm & Schubotz, 2012), including episodic PEs (Jainta et al., 2024; Liedtke et al., 2025; Siestrup et al., 2023). Specifically, the IFG may hold the predictive model that will then be compared to reality (Boeltzig, Liedtke, Siestrup, et al., 2025; Fujitani et al., 2024; Liedtke et al., 2025; Sherman et al., 2016). Second, the STG and STS are highly relevant for language processing (Bhaya-Grossman & Chang, 2021) and auditory PEs (Schlossmacher et al., 2022) and may therefore especially have been recruited in the current design that used conversations as stimuli. However, the STS and STG also process socially relevant information (Masson & Isik, 2021), including speaker familiarity (Cordero et al., 2025), goals of actors (Shultz et al., 2011), and theory of mind (Carrington & Bailey, 2009). This makes this region a key hub in the updating of person impressions (Mende-Siedlecki et al., 2013). Finally, the left Crus 2 in the cerebellum has been shown to be involved in mentalizing (Overwalle et al., 2020), social preference predictions (Haihambo et al., 2023), and responds to PEs in sequences performed by others (Li et al., 2024).

Both STS/STG and Crus 2 were presumably triggered by the socially relevant dialogues used in the current study. Participants could form impressions about recurring speakers, which were continuously challenged and updated. Importantly, while the role of STS/STG and Crus 2 in socially relevant PEs and the IFG in domain-general prediction is in line with previous studies, the current results offer a crucial extension: All three regions responded to PEs that updated existing knowledge, and to PEs that led to the inference of a new underlying latent cause – a differentiation that previous studies have not made. The current study therefore places these areas at the centre of all PE-triggered operations within the context of narrative or social material.

4.2. New latent cause inference after low-precision predictions

At the first exposure to new material, participants could form only low-precision predictions about incoming input, based on related episodic memories (Liedtke et al., 2026) or schemas and semantic knowledge (Frank et al., 2020; Zöllner et al., 2022). If the dialogue involved an unpredicted change of topic (incoherent dialogues), necessitating the inference of a new latent cause, these low-precision across-cause PEs activated a large array of brain regions in addition to the PE-general ones discussed above (compared to low-precision within-cause PEs which were used as a baseline here). First, the posterior dorsal precuneus responded in these cases, an area previously associated with PEs evoked by structural changes in narratives (Siestrup et al., 2022; Siestrup & Schubotz, 2023), but also with unexpected or novel social information relevant for knowledge about a person (Dungan et al., 2016; Stanley, 2016). Similarly, the temporoparietal junction (TPJ), a key hub for theory of mind and updating of person knowledge (Mende-Siedlecki, 2018; Olsson et al., 2020) was activated.

This suggests that when low-precision predictions in novel and unfamiliar situations are violated, across-cause PEs are highly meaningful, informing socially relevant person models. For instance, such across-cause PEs may signal that the second speaker has misunderstood the first, prompting an expectation of clarification, or that the exchange involves an unobserved third party, leading the listener to infer a different social configuration and informing social impressions.

The representational consequences of low-precision across-cause PEs are in line with this interpretation. Using representational similarity analysis (RSA) to compare these PEs with low-precision within-cause PEs, the inference of a new latent cause did separate the parts of the dialogue from each other representationally: When the beginning of the dialogue was presented again during a later exposure, its continuation was less strongly reinstated on the whole-brain level, and thereby less strongly predicted (Kim et al., 2014), compared to low-precision within-cause PEs. Increased representational separation between original and new, PE-evoking information after larger PEs has been observed before (Bein et al., 2023; Boeltzig, Liedtke, Siestrup, et al., 2025), and can enhance memory (Kuhl et al., 2010). The current study newly ties pattern separation, and therefore a process akin to event segmentation (Nolden et al., 2024; Zacks & Swallow, 2007), explicitly to latent cause inference. As the two statements are deemed unrelated, they appear to be represented almost as two separate events, as a consequence of new latent cause inference.

This is also in line with the observed effects on recognition memory, as after low-precision across-cause PEs, original memory traces were remembered as well as after low-precision within-cause PEs, despite the latter having a general coherence advantage. Even though this indicates increased relevance of even incoherent social information in a novel situation, there was a meta-mnemonic advantage of the coherent dialogues, represented by elevated confidence.

Additionally, low-precision across-cause PEs led to activation in the anterior insula, likely reflecting the high uncertainty (Sarinopoulos et al., 2010), possibly even aversiveness (Hoskin & Talmi, 2023; Kim et al., 2025) of a situation in which new material is not in line with expectations. Interestingly, the anterior insula and TPJ display increased connectivity in social fear learning (Lindström et al., 2018), implying that participants may be learning by observation from an apparently failed communicative situation.

4.3. New latent cause inference after high-precision predictions

While in the absence of specific predictions, new latent cause inference involved an extended network of brain regions implicated in social processing, fewer regions responded to new latent causes when specific predictions were possible. This was the case when dialogues had already been encoded in their original form, could be predicted upon another exposure, but were modified with a new incoherent dialogue version which was to evoke a high-precision across-cause PE. Only the PE-general areas IFG, STS/STG, and cerebellar Crus 2 responded to these PEs.

There was an immediate representational consequence of high-precision across-cause PEs. Compared to high-precision within-cause PEs, RSA revealed that the originally encoded version was more strongly reactivated in the hippocampus while the mismatch was being processed. As this original version was used for prediction, but challenged by the PE, this increased reactive reinstatement can reflect a continued top-down adherence to the prediction, as opposed to bottom-up information gathering. As the hippocampus is thought to face a trade-off between predicting and encoding, the continued reinstatement can therefore impede the encoding of the new stimulus (Long & Kuhl, 2019; Sherman & Turk-Browne, 2020).

This can then explain that besides the core PE network, no additional regions significantly responded to these PEs. Additional support for this interpretation comes from the univariate contrast between high-precision within-cause and across-cause PEs, where high-precision within-cause PEs elicited reduced A1 activity, signalling less bottom-up processing (Liedtke et al., 2026). Even though this did not lead to a memory disadvantage of these dialogues, confidence was indeed decreased, compared to dialogues with a high-precision within-cause PE. Additionally, reactivation of previously encoded relevant events is often observed at event boundaries (Cohn-Sheehy et al., 2021; Griffiths & Fuentemilla, 2020; Hahamy et al., 2023), drawing attention to the similarities between latent cause inference and event segmentation, also discussed by Gershman et al. (2017).

4.4. Model updating after high-precision predictions

Lastly, high-precision within-cause PEs were evoked when new input was mismatching with the original dialogue version, but not incoherent. These PEs offered an appropriate alternative to the previously encoded stimulus. In this case, Latent Cause Theory (Gershman et al., 2017) would predict no new latent cause inference, but instead the updating of the existing model with new information. Note that as the new targets still featured the same speaker, mismatches were not generated based on an unfamiliar voice, but on unpredicted high-level information.

Brain activity was strongly in line with this prediction. Again, PE-general areas including the IFG, left Crus 2, and both anterior and posterior STG, here extending to the temporal pole, were activated. In addition to left Crus 2, there was a significant cluster also in the right Crus 2, which exhibits high connectivity to mentalizing hubs (Overwalle et al., 2019). These were also activated, specifically the bilateral ventral precuneus and the bilateral TPJ, implicated in the updating of person impressions and theory of mind (Dungan et al., 2016; Mende-Siedlecki, 2018; Stanley, 2016).

Crucially, two higher-order frontal activation clusters were found that were unique to high-precision within-cause PEs, namely a cluster centred on BA9, extending into BA8 and the pre-supplementary motor area, as well as a cluster in the ventro-medial prefrontal cortex (vmPFC). BA9 was hypothesized to be active due to its contribution to integrating new information into a situation model (Ferstl & Cramon, 2001). Situation models are built up during comprehension of texts, but also other media and real-world situations, representing the evolving of meaning (Thye et al., 2024; Zwaan, 2025). BA9 has thus been shown to be active when information was unpredicted but could be reconciled with the overall narrative by re-interpretation, but not if such integration was impossible (Ferstl et al., 2005) or when stories lacked structural coherence (Yarkoni et al., 2008). This role of BA9 in psycholinguistic studies is highly overlapping with its function found in the current study, as the updating of knowledge within a single latent cause as described by Gershman et al. (2017) is akin to the continuous updating of a situation model. Additionally, BA9 has been implicated in maintaining and adjusting an active predictive model, as shown in sequential prediction tasks (Kühn & Schubotz, 2012), where BA9 activity reflected whether an ongoing prediction needed to be stabilized or revised after rare expectancy breaches.

Taken together, these previous findings converge on a common role of BA9 in keeping an interpretive model online and updating it when unexpected information can be integrated. This is precisely the process required for high-precision within-cause PEs in the present study. In addition to BA9, the cluster in the vmPFC points to the resolution of socially ambiguous settings (Lyu et al., 2024), and the need to integrate new information into existing memories (Bowman & Zeithamova, 2018; Spalding et al., 2018; Zeithamova et al., 2012). While the vmPFC has been shown to be crucial in the processing of reward PEs (Rehbein et al., 2023; Wang et al., 2017), the current result suggests that it may be relevant also when updating episodic memories.

These results therefore implicate BA9 and vmPFC in the processing of PEs – but especially when these PEs are of moderate size, where existing models are updated with new information (Gershman et al., 2017). Because the cluster also extended into the pre-supplementary motor area, model updating likely engaged additional processes that manage conflicts between local predictions and the broader hierarchical structure of a sequence (Hertrich et al., 2016; Li et al., 2021).

The integration of new information into existing models was also reflected in increased A1 activation, and as an immediate representational consequence simultaneously less reinstatement of the originally encoded version, both compared to high-precision across-cause PEs. While latent cause inference seemed to trigger a switch to top-down prediction with increased reinstatement of a previously encoded alternative and less A1 activation, the opposite was the case when new relevant information was gathered in a bottom-up fashion and integrated into an existing model. Consequently, even though recognition memory was not increased, confidence and source memory were.

4.5. *Implications for latent cause theory in episodic predictions*

Latent Cause Theory (Gershman et al., 2017) has only recently been applied to PEs in a context of episodic predictions. From the current study, four theoretically relevant conclusions about the nature of new latent cause inference and model updating emerge. First, no brain region was directly implicated in new latent cause inference in the current design. While Latent Cause Theory is per se not strictly tied to a specific biological implementation, the initial finding that in the current design, no clusters reflecting the signalling of a new latent cause crossed the threshold of significance is consistent with previous computational modelling work by Pronoza et al. (2025). In that study, a model was constructed which was capable of replicating previous behavioural results, where the decision about updating or new encoding was based merely on the similarity of new material with an already encoded memory. In that model, no explicit higher-order "decision maker" was necessary, and outcomes could be replicated based solely on internal network dynamics.

Second, neural processes at new latent cause inference differ depending on whether predictions are low-precision or high-precision, with fewer brain regions responding when high-precision predictions can be made. At first glance, this might be inconsistent with the idea that high-precision predictions generally trigger stronger PEs (Greve et al., 2017; Henson & Gagnepain, 2010). However, a more distributed response to PEs is not synonymous with a stronger PE but may rather indicate broad information gathering when little is known about the novel situation. Additionally, the originally encoded memories were more strongly reactively reinstated when an across-cause PE compared to a within-cause PE was induced. This raises the intriguing possibility that these across-cause PEs may have biased the system from bottom-up processing of the unexpected input towards a top-down reinstatement of a previously encoded version of the same episode. This very reinstatement may then have constrained the PE response.

Third, we have previously argued for new latent cause inference after large PEs caused by the modification of a previously encoded episode, resulting in representational separation between the original and new mismatching versions (Boeltzig, Liedtke, Siestrup, et al., 2025). These results are significantly extended in the current study, where we find evidence for new latent cause inference even within the same stimulus, so within a short period of time. This event segmentation is not necessarily predicted by Latent Cause Theory, which argues that temporal proximity is a cue for the maintenance of the same latent cause (Gershman et al., 2017), at least in conditioning designs which the theory is based on. Beyond conditioning settings however, even episodic memories formed closer in time are more likely to be integrated instead of separated (Zeithamova & Preston, 2017) and a recent study showed that a single PE is not enough for event segmentation, but persistence of the new situation causing the PE is also necessary (Güler et al., 2025). However, in these studies, events were either stimulus-response pairings or random associations. It is conceivable that narrative stimuli, with their extension in time, carry persistence of a new latent cause inherently, thus making new latent cause inference more likely (Ben-Yakov et al., 2022). In addition to the general advantages of using naturalistic material in memory research (Maguire, 2022), research on PEs especially needs to approximate real-world situations to account for factors such as temporal structure and persistence of unpredicted events.

Fourth, the results underscore that predictive models can have several sources in any given situation. For instance, when participants first listen to a novel dialogue, it is unclear what they predict and where these predictions stem from. Later, familiar dialogues could be predicted based on a specific episodic memory, as the RSA suggests, and prior research corroborates (Boeltzig, Liedtke, Siestrup, et al., 2025; Haque et al., 2020; Kim et al., 2014; Stawarczyk et al., 2020). However, after low-precision across-cause PEs, the previously encoded dialogue continuations were less predicted upon future exposures – leaving it open what was predicted instead. Previous studies have usually either induced episodic PEs by modifying previously encoded episodes (Kim et al., 2014; Nolden et al., 2025; Stawarczyk et al., 2020), or semantic PEs by violating abstract rules or schemas (Frank et al., 2020; Kalbe & Schwabe, 2020; Pupillo et al., 2023). In real-life prediction, it is conceivable that both sources are tapped in any given context (Cheng et al., 2016), and that one or more relevant episodic memories are leveraged for prediction in novel situations (Liedtke et al., 2026). As PEs based on previous episodic experiences differ in neural correlates from PEs based on semantic knowledge (Varga et al., 2025), it is an interesting avenue for future research to focus on identifying the sources for predictions in naturalistic viewing and investigate whether they modulate mnemonic and representational outcomes.

4.6. *Limitations*

In order to create circumstances in which, grounded in Latent Cause Theory (Gershman et al., 2017), we expected new latent cause inference and updating to be the most likely consequences, we made use of coherent and incoherent dialogues. While the manipulation was successful, as representational similarity at first exposure indicated, it cannot be ruled out that participants made sense of the incoherent dialogues in some way or another. Conversely, in the coherent episodes, it is possible that participants sometimes judged responses to be inappropriate.

As another product of the categorical design, narrative coherence can generally lead to a memory advantage (Arslan & Kominsky, 2026; Cohn-Sheehy et al., 2022), possibly neutralizing the previously shown beneficial effect of new latent cause inference or otherwise large PEs (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025). While there was no effect of coherence on recognition performance (see Haeuser & Kray, 2022 for a similar result), it did affect confidence (as a more meta-cognitive process; Nielsen et al., 2023) and source memory. Due to these two opposing effects, the current set of behavioural memory results are more challenging to interpret.

It is a well-known challenge in the field that comparing findings between studies is inherently difficult. Context can moderate the effects of novelty and surprise (Barto et al., 2013; Kafkas & Montaldi, 2018; Quent et al., 2021; Reichardt et al., 2020), and the stability of the environment can influence how strongly the system is tuned to PEs (Rouhani et al., 2018) or whether predictions are made on a category- or item-level (Kim et al., 2019). In the current study, uncertainty was probably generally higher than in previous studies of PE size, given that new material was continuously introduced and modified in a single session, with more obvious PEs evoked by the incoherent versions. This uncertainty may influence the strength of predictions and render PEs generally smaller (Henson & Gagnepain, 2010). To account for this and other factors which make an exact PE size determination difficult, future studies should focus on modelling the full space of possible PE sizes and consider utilising physiological measures of PE (Brod et al., 2022; Loock et al., 2025).

More broadly, experiments creating episodic PEs rely on participants encoding episodes, and then re-playing these episodes in an identical manner before an unexpected mismatch is induced (Greve et al., 2017;

Kim et al., 2014; Varga et al., 2025; Wahlheim & Zacks, 2019). This naturally simplifies naturalistic predicting. Even though we are often exposed to retellings of the same stories and highly repetitive event, and watch movies, read books, and listen to audiobooks that we already know (O'Brien, 2019), the situation including its context will not directly repeat itself. While it can still be highly effective to leverage episodic memories in these situations, future research should also create conditions where the new situation is similar, but not completely identical, and test whether this modulates signalling and consequences of episodic PEs.

5. Conclusions

The current study elucidates differences between model updating and latent cause inference. Integrating new information into an existing model engaged frontal midline regions (BA9, vmPFC) and enhanced bottom-up processing, whereas inferring a new latent cause relied on a core PE network (IFG, STS/STG, Crus 2). Prediction precision determined whether additional regions were recruited: low-precision predictions elicited broader processing, while reinstated high-precision predictions constrained processing to a narrower, more top-down pattern. Together, these findings contribute to the application of the Latent Cause Theory by Gershman et al. (2017) to naturalistic prediction in humans, where episodic information is tapped in addition to abstract schemas and rules. It also extends prior work showing that mnemonic and representational consequences differ depending on PE size (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025; Liedtke et al., 2025; Pronoza et al., 2025), highlighting more broadly how prediction errors can contribute to the encoding and modification of episodic memories.

These results are also relevant in the broader memory literature. For instance, they can explain why implausible information is less accessible than plausible information (Ford & Nadarevic, 2025) – it might be encoded, but less accessible due to increased pattern separation with relevant cues. Additionally, fake news that is congruent with attitudes may trigger within-cause PEs that lead to updating of previous knowledge. By becoming integrated and by being used for prediction in the future, they exert continued influence on cognition. On the other hand, even truthful news that are incongruent with attitudes may be separated by new latent cause inference from relevant cues and thus be less accessible. Prediction errors therefore are a fruitful avenue to understanding, and ultimately countering, partisan cognition and fake news susceptibility (Kemp et al., 2024; Sinclair et al., 2020). Considering PE size, and thereby the navigation between new latent cause inference and model updating is a crucial avenue for future research in that field.

Funding

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.

Data Availability

The data needed to recreate behavioural and representational similarity analysis, the *t*-maps for univariate contrasts, and the stimulus material are publicly available on OSF: https://osf.io/mgu49/overview?view_only=1af278283f49433d8de09f566a2dff97.

CRediT authorship contribution statement

Marius Boeltzig: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sophie Siestrup:** Writing – review & editing, Software, Formal analysis, Conceptualization. **Inês Bramão:** Writing – review & editing, Methodology, Conceptualization. **Ricarda I. Schubotz:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors have no conflicts of interest to declare.

Acknowledgements

We thank Nina Liedtke for her work in stimulus production and discussions on the methods. We also thank Aleyna Temizel, Anna Sophia Asdonk, Julian von Seydewitz, Juliane Diefenthal, Lena Charlotte Innig, Leon Exeler, Maike Elisabeth Denguth, Pia Niedrée, Thea Maria Bröring, and Yannah Loïs Schröder for their contributions to data collection and stimulus preparation.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2026.121898.

References

Addis, D.R., Wong, A.T., Schacter, D.L., 2007. Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45 (7), 1363–1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016.

Arslan, A., Kominsky, J.F., 2026. Causal coherence improves episodic memory of dynamic events. Cognition 266, 106317. https://doi.org/10.1016/j.cognition.2025.106317.

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

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

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., 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.

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.

Ben-Yakov, A., Smith, V., Henson, R., 2022. The limited reach of surprise: Evidence against effects of surprise on memory for preceding elements of an event. Psychonomic Bulletin & Review 29 (3), 1053–1064. https://doi.org/10.3758/s13423-021-01954-5.

Bhaya-Grossman, I., Chang, E.F., 2021. Speech computations of the human superior temporal gyrus. Annual Review of Psychology 73 (1), 1–24. 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.

Bowman, C.R., Zeithamova, D., 2018. Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. The Journal of Neuroscience 38 (10), 2605–2614. https://doi.org/10.1523/jneurosci.2811-17.2018.

Brod, G., Greve, A., Jolles, D., Theobald, M., Galeano-Keiner, E.M., 2022. Explicitly predicting outcomes enhances learning of expectancy-violating information. Psychonomic Bulletin & Review 29 (6), 2192–2201. https://doi.org/10.3758/s13423-022-02124-x.

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., von Cramon, D.Y., Schubotz, R.I., 2010. Prediction, cognition and the brain. Frontiers in Human Neuroscience 4, 25. https://doi.org/10.3389/fnhum.2010.00025.

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

Chen, J., Cook, P.A., Wagner, A.D., 2015. Prediction strength modulates responses in human area CA1 to sequence violations. Journal of Neurophysiology 114 (2), 1227–1238. https://doi.org/10.1152/jn.00149.2015.

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

Cohn-Sheehy, B.I., Delarazan, A.I., Crivelli-Decker, J.E., Reagh, Z.M., Mundada, N.S., Yonelinas, A.P., Zacks, J.M., Ranganath, C., 2022. Narratives bridge the divide between distant events in episodic memory. Memory & Cognition 50 (3), 478–494. https://doi.org/10.3758/s13421-021-01178-x.

Cohn-Sheehy, B.I., Delarazan, A.I., Reagh, Z.M., Crivelli-Decker, J.E., Kim, K., Barnett, A. J., Zacks, J.M., Ranganath, C., 2021. The hippocampus constructs narrative memories across distant events. Current Biology 31 (22), 4935–4945. https://doi.org/10.1016/j.cub.2021.09.013.

Cole, S., Kvavilashvili, L., 2021. Spontaneous and deliberate future thinking: a dual process account. Psychological Research 85 (2), 464–479. https://doi.org/10.1007/s00426-019-01262-7.

Cordero, G., Paredes-Paredes, J.R., von Kriegstein, K., Díaz, B., 2025. Perceiving speech from a familiar speaker engages the person identity network. PLOS One 20 (5), e0322927. https://doi.org/10.1371/journal.pone.0322927.

Dimsdale-Zucker, H.R., Ranganath, C., 2018. Representational similarity analyses: A practical guide for functional MRI applications. In: Manahan-Vaughan, D. (Ed.), Handbook of Behavioral Neuroscience, Handbook of Behavioral Neuroscience, 28. Elsevier, pp. 509–525. https://doi.org/10.1016/b978-0-12-812028-6.00027-6.

Dungan, J.A., Stepanovic, M., Young, L., 2016. Theory of mind for processing unexpected events across contexts. Social Cognitive and Affective Neuroscience 11 (8), 1183–1192. https://doi.org/10.1093/scan/nsw032.

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.

Ferstl, E.C., von Cramon, D.Y., 2001. The role of coherence and cohesion in text comprehension: an event-related fMRI study. Cognitive Brain Research 11 (3), 325–340. https://doi.org/10.1016/s0926-6410(01)00007-6.

Ferstl, E.C., Rinck, M., von Cramon, D.Y., 2005. Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. Journal of Cognitive Neuroscience 17 (5), 724–739. https://doi.org/10.1162/0898929053747658.

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

Ford, D., Nadarevic, L., 2025. Revisiting the plausibility effect in remembering truth and falsity: An analysis of underlying memory and guessing processes. Journal of Cognition 8 (1), 46. https://doi.org/10.5334/joc.459.

Frank, D., Montaldi, D., Wittmann, B., Talmi, D., 2018. Beneficial and detrimental effects of schema incongruence on memory for contextual events. Learning & Memory 25 (8), 352–360. https://doi.org/10.1101/lm.047738.118.

Frank, D., Montemurro, M.A., Montaldi, D., 2020. Pattern separation underpins expectation-modulated memory. The Journal of Neuroscience 40 (17), 3455–3464. https://doi.org/10.1523/jneurosci.2047-19.2020.

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.

Fujitani, S., Kunii, N., Nagata, K., Takasago, M., Shimada, S., Tada, M., Kirihara, K., Komatsu, M., Uka, T., Kasai, K., Saito, N., 2024. Auditory prediction and prediction error responses evoked through a novel cascade roving paradigm: a human ECoG study. Cerebral Cortex 34 (2), bhad508. https://doi.org/10.1093/cercor/bhad508.

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

Greve, A., Abdulrahman, H., Henson, R.N., 2018. Neural differentiation of incorrectly predicted memories. Frontiers in Human Neuroscience 12, 278. https://doi.org/10.3389/fnhum.2018.00278.

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

Greve, A., Cooper, E., Tibon, R., Henson, R.N., 2019. Knowledge Is Power: Prior Knowledge Aids Memory for Both Congruent and Incongruent Events, but in Different Ways. Journal of Experimental Psychology: General 148 (2), 325–341. https://doi.org/10.1037/xge0000498.

Griffiths, B.J., Fuentemilla, L., 2020. Event conjunction: How the hippocampus integrates episodic memories across event boundaries. Hippocampus 30 (2), 162–171. https://doi.org/10.1002/hipo.23161.

Güler, B., Serin, F., Günseli, E., 2025. Prediction error is out of context: The dominance of contextual stability in structuring episodic memories. Psychonomic Bulletin & Review 1–12. https://doi.org/10.3758/s13423-025-02723-4.

Haeuser, K.I., Kray, J., 2022. How odd: Diverging effects of predictability and plausibility violations on sentence reading and word memory. Applied Psycholinguistics 43 (5), 1193–1220. https://doi.org/10.1017/s0142716422000364.

Hahamy, A., Dubossarsky, H., Behrens, T.E.J., 2023. The human brain reactivates context-specific past information at event boundaries of naturalistic experiences. Nature Neuroscience 26 (6), 1080–1089. https://doi.org/10.1038/s41593-023-01331-6.

Haihambo, N., Ma, Q., Baetens, K., Bylemans, T., Heleven, E., Baeken, C., Deroost, N., Overwalle, F.V., 2023. Two is company: The posterior cerebellum and sequencing for pairs versus individuals during social preference prediction. Cognitive, Affective, & Behavioral Neuroscience 23 (6), 1482–1499. https://doi.org/10.3758/s13415-023-01127-y.

Haque, R.U., Inati, S.K., Levey, A.I., Zaghloul, K.A., 2020. Feedforward prediction error signals during episodic memory retrieval. Nature Communications 11 (1), 6075. https://doi.org/10.1038/s41467-020-19828-0.

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. Neuroscience & Biobehavioral Reviews 68, 602–610. https://doi.org/10.1016/j.neubiorev.2016.06.030.

Hoskin, R., Talmi, D., 2023. Adaptive coding of pain prediction error in the anterior insula. European Journal of Pain 27 (6), 766–778. https://doi.org/10.1002/ejp.2093.

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.

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.

Kafkas, A., Montaldi, D., 2018. How do memory systems detect and respond to novelty? Neuroscience Letters 680, 60–68. https://doi.org/10.1016/j.neulet.2018.01.053.

Kalbe, F., Schwabe, L., 2020. Beyond arousal: Prediction error related to aversive events promotes episodic memory formation. Journal of Experimental Psychology: Learning, Memory, and Cognition 46 (2), 234–246. https://doi.org/10.1037/xlm0000728.

Kemp, P.L., Sinclair, A.H., Adcock, R.A., Wahlheim, C.N., 2024. Memory and belief updating following complete and partial reminders of fake news. Cognitive Research: Principles and Implications 9 (1), 28. https://doi.org/10.1186/s41235-024-00546-w.

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.

Kim, H., Schlichting, M.L., Preston, A.R., Lewis-Peacock, J.A., 2019. Predictability changes what we remember in familiar temporal contexts. Journal of Cognitive Neuroscience 32 (1), 124–140. https://doi.org/10.1162/jocn_a_01473.

Kim, J.-C., Hellrung, L., Nebe, S., Tobler, P.N., 2025. The anterior insula processes a time-resolved subjective risk prediction error. The Journal of Neuroscience 45 (23), e2302242025. https://doi.org/10.1523/jneurosci.2302-24.2025.

Kuhl, B.A., Shah, A.T., DuBrow, S., Wagner, A.D., 2010. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nature Neuroscience 13 (4), 501–506. https://doi.org/10.1038/nn.2498.

Kühn, A.B., Schubotz, R.I., 2012. Temporally remote destabilization of prediction after rare breaches of expectancy. Human Brain Mapping 33 (8), 1812–1820. https://doi.org/10.1002/hbm.21325.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software (13), 82. https://doi.org/10.18637/jss.v082.i13.

Li, C.-W., Guo, F.-Y., Tsai, C.-G., 2021. Predictive processing, cognitive control, and tonality stability of music: An fMRI study of chromatic harmony. Brain and Cognition 151, 105751. https://doi.org/10.1016/j.bandc.2021.105751.

Li, M., Pu, M., Ma, Q., Heleven, E., Baeken, C., Baetens, K., Deroost, N., Overwalle, F.V., 2024. One step too far: social cerebellum in norm-violating navigation. Social Cognitive and Affective Neuroscience 19 (1), nsae027. https://doi.org/10.1093/scan/nsae027. Article.

Li, Q., Liu, G., Yuan, G., Wang, G., Wu, Z., Zhao, X., 2019. Single-trial EEG-fMRI reveals the generation process of the mismatch negativity. Frontiers in Human Neuroscience 13, 168. https://doi.org/10.3389/fnhum.2019.00168.

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. Frontiers in Neuroscience 8, 289. https://doi.org/10.3389/fnins.2014.00289.

Liedtke, N., Boeltzig, Marius, 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.

Liedtke, N., Boeltzig, M., Siestrup, S., Schubotz, R.I., 2026. Shaping Memory from the Start: Initial Prediction Errors during First Encoding. NeuroImage 325, 121660. https://doi.org/10.1016/j.neuroimage.2025.121660.

Lindström, B., Haaker, J., Olsson, A., 2018. A common neural network differentially mediates direct and social fear learning. NeuroImage 167, 121–129. https://doi.org/10.1016/j.neuroimage.2017.11.039.

Long, N.M., Kuhl, B.A., 2019. Decoding the tradeoff between encoding and retrieval to predict memory for overlapping events. NeuroImage 201, 116001. https://doi.org/10.1016/j.neuroimage.2019.07.014.

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.

Loock, K., Kalbe, F., Schwabe, L., 2025. Cognitive mechanisms of aversive prediction error-induced memory enhancements. Journal of Experimental Psychology: General 154 (4), 1102–1121. https://doi.org/10.1037/xge0001712.

Lyu, Y., Su, Z., Neumann, D., Meidenbauer, K.L., Leong, Y.C., 2024. Hostile attribution bias shapes neural synchrony in the left ventromedial prefrontal cortex during ambiguous social narratives. The Journal of Neuroscience 44 (9), e1252232024. https://doi.org/10.1523/jneurosci.1252-23.2024.

Maguire, E.A., 2022. Does memory research have a realistic future? Trends in Cognitive Sciences 26 (12), 1043–1046. https://doi.org/10.1016/j.tics.2022.07.006.

Masson, H.L., Isik, L., 2021. Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. NeuroImage 245, 118741. https://doi.org/10.1016/j.neuroimage.2021.118741.

Mende-Siedlecki, P., 2018. Changing our minds: the neural bases of dynamic impression updating. Current Opinion in Psychology 24, 72–76. https://doi.org/10.1016/j.copsyc.2018.08.007.

Mende-Siedlecki, P., Cai, Y., Todorov, A., 2013. The neural dynamics of updating person impressions. Social Cognitive and Affective Neuroscience 8 (6), 623–631. https://doi.org/10.1093/scan/nss040.

Nielsen, N.P., Gehrt, T.B., Berntsen, D., 2023. Individual differences in autobiographical memory predict memory confidence but not memory accuracy. Journal of Applied Research in Memory and Cognition 12 (4), 542–551. https://doi.org/10.1037/mac0000082.

Nolden, S., Turan, G., Bein, O., Davachi, L., Shing, Y.L., 2025. The impact of mnemonic prediction errors on episodic memory: A lifespan study. Developmental Psychology 61 (10), 1927–1940. https://doi.org/10.1037/dev0001966.

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.

O'Brien, E., 2019. Enjoy It Again: Repeat Experiences Are Less Repetitive Than People Think. Journal of Personality and Social Psychology 116 (4), 519–540. https://doi.org/10.1037/pspa0000147.

Olsson, A., Knapska, E., Lindström, B., 2020. The neural and computational systems of social learning. Nature Reviews Neuroscience 21 (4), 197–212. https://doi.org/10.1038/s41583-020-0276-4.

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, 27. https://doi.org/10.3389/fninf.2016.00027.

Overwalle, F.V., Ma, Q., Heleven, E., 2020. The posterior crus II cerebellum is specialized for social mentalizing and emotional self-experiences: a meta-analysis. Social Cognitive and Affective Neuroscience 15 (9), 905–928. https://doi.org/10.1093/scan/nsaa124.

Overwalle, F.V., Steen, F.V.de, Mariën, P., 2019. Dynamic causal modeling of the effective connectivity between the cerebrum and cerebellum in social mentalizing across five studies. Cognitive, Affective, & Behavioral Neuroscience 19 (1), 211–223. https://doi.org/10.3758/s13415-018-00659-y.

Pronoza, J., Liedtke, N., Boeltzig, M., Schubotz, R.I., Cheng, S., 2025. Competition between memory updating and differentiation emerges from intrinsic network dynamics. bioRxiv. https://doi.org/10.1101/2025.10.31.685486.

Pupillo, F., Ortiz-Tudela, J., Bruckner, R., Shing, Y.L., 2023. The effect of prediction error on episodic memory encoding is modulated by the outcome of the predictions. Npj Science of Learning 8, 18. https://doi.org/10.1038/s41539-023-00166-x.

Quent, J.A., Henson, R.N., Greve, A., 2021. A predictive account of how novelty influences declarative memory. Neurobiology of Learning and Memory 179, 107382. https://doi.org/10.1016/j.nlm.2021.107382.

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

Rehbein, M.A., Kroker, T., Winker, C., Ziehfreund, L., Reschke, A., Bölte, J., Wyczesany, M., Roesmann, K., Wessing, I., Junghöfer, M., 2023. Non-invasive stimulation reveals ventromedial prefrontal cortex function in reward prediction and reward processing. Frontiers in Neuroscience 17, 1219029. https://doi.org/10.3389/fnins.2023.1219029.

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

Reimers, N., Gurevych, I., 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-Networks. arXiv. https://doi.org/10.48550/arxiv.1908.10084.

Rouhani, N., Norman, K.A., Niv, Y., 2018. Dissociable effects of surprising rewards on learning and memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 44 (9), 1430–1443. https://doi.org/10.1037/xlm0000518.

Rowe, E.G., Tsuchiya, N., Garrido, M.I., 2020. Detecting (Un)seen Change: The Neural Underpinnings of (Un)conscious Prediction Errors. Frontiers in Systems Neuroscience 14, 541670. https://doi.org/10.3389/fnsys.2020.541670.

Sarinopoulos, I., Grupe, D.W., Mackiewicz, K.L., Herrington, J.D., Lor, M., Steege, E.E., Nitschke, J.B., 2010. Uncertainty during anticipation modulates neural responses to aversion in human insula and amygdala. Cerebral Cortex 20 (4), 929–940. https://doi.org/10.1093/cercor/bhp155.

Schliephake, L.M., Trempler, I., Roehe, M.A., Heins, N., Schubotz, R.I., 2021. Positive and negative prediction error signals to violated expectations of face and place stimuli distinctively activate FFA and PPA. NeuroImage 236, 118028. https://doi.org/10.1016/j.neuroimage.2021.118028.

Schlossmacher, I., Dilly, J., Protmann, I., Hofmann, D., Dellert, T., Roth-Paysen, M.-L., Moeck, R., Bruchmann, M., Straube, T., 2022. Differential effects of prediction error and adaptation along the auditory cortical hierarchy during deviance processing. NeuroImage 259, 119445. https://doi.org/10.1016/j.neuroimage.2022.119445.

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

Sherman, B.E., Turk-Browne, N.B., 2020. Statistical prediction of the future impairs episodic encoding of the present. Proceedings of the National Academy of Sciences 117 (37), 22760–22770. https://doi.org/10.1073/pnas.2013291117.

Sherman, M.T., Seth, A.K., Kanai, R., 2016. Predictions shape confidence in right inferior frontal gyrus. The Journal of Neuroscience 36 (40), 10323–10336. https://doi.org/10.1523/jneurosci.1092-16.2016.

Shultz, S., Lee, S.M., Pelphrey, K., McCarthy, G., 2011. The posterior superior temporal sulcus is sensitive to the outcome of human and non-human goal-directed actions. Social Cognitive and Affective Neuroscience 6 (5), 602–611. https://doi.org/10.1093/scan/nsq087.

Siestrup, S., Jainta, B., Cheng, S., Schubotz, R.I., 2023. Solidity meets surprise: Cerebral and behavioral effects of learning from episodic prediction errors. Journal of Cognitive Neuroscience 35 (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. 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., 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.

Sinclair, A.H., Stanley, M.L., Seli, P., 2020. Closed-minded cognition: Right-wing authoritarianism is negatively related to belief updating following prediction error. Psychonomic Bulletin & Review 27 (6), 1348–1361. https://doi.org/10.3758/s13423-020-01767-y.

Spalding, K.N., Schlichting, M.L., Zeithamova, D., Preston, A.R., Tranel, D., Duff, M.C., Warren, D.E., 2018. Ventromedial prefrontal cortex Is necessary for normal associative inference and memory integration. The Journal of Neuroscience 38 (15), 3767–3775. https://doi.org/10.1523/jneurosci.2501-17.2018.

Stanley, D.A., 2016. Getting to know you: general and specific neural computations for learning about people. Social Cognitive and Affective Neuroscience 11 (4), 525–536. https://doi.org/10.1093/scan/nsv145.

Stawarczyk, D., Wahlheim, C.N., Etzel, J.A., Snyder, A.Z., Zacks, J.M., 2020. Aging and the encoding of changes in events: The role of neural activity pattern reinstatement. Proceedings of the National Academy of Sciences 117 (47), 29346–29353. https://doi.org/10.1073/pnas.1918063117.

Thye, M., Hoffman, P., Mirman, D., 2024. All the stars will be wells with a rusty pulley": Neural processing of the social and pragmatic content in a narrative. Journal of Cognitive Neuroscience 36 (11), 2495–2517. https://doi.org/10.1162/jocn_a_02228.

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

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.

Wahlheim, C.N., Eisenberg, M.L., Stawarczyk, D., Zacks, J.M., 2021. Understanding everyday events: Predictive-looking errors drive memory updating. Psychological Science 33 (5), 765–781. https://doi.org/10.1177/09567976211053596.

Wahlheim, C.N., Zacks, J.M., 2019. Memory guides the processing of event changes for older and younger adults. Journal of Experimental Psychology: General 148 (1), 30–50. https://doi.org/10.1037/xge0000458.

Wahlheim, C.N., Zacks, J.M., 2025. Memory updating and the structure of event representations. Trends in Cognitive Sciences 29 (4), 380–392. https://doi.org/10.1016/j.tics.2024.11.008.

Wang, Y., Ma, N., He, X., Li, N., Wei, Z., Yang, L., Zha, R., Han, L., Li, X., Zhang, D., Liu, Y., Zhang, X., 2017. Neural substrates of updating the prediction through prediction error during decision making. NeuroImage 157, 1–12. https://doi.org/10.1016/j.neuroimage.2017.05.041.

Weaverdyck, M.E., Lieberman, M.D., Parkinson, C., 2020. Multivoxel pattern analysis in fMRI: A practical introduction for social and affective neuroscientists. Social Cognitive and Affective Neuroscience 15 (4), 487–509. https://doi.org/10.1093/scan/nsaa057.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Yutani, H., 2019. Welcome to the Tidyverse. Journal of Open Source Software 4 (43), 1686. https://doi.org/10.21105/joss.01686.

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.

Zacks, J.M., Swallow, K.M., 2007. Event Segmentation. Current Directions in Psychological Science 16 (2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x.

Zeithamova, D., Dominick, A.L., Preston, A.R., 2012. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron 75 (1), 168–179. https://doi.org/10.1016/j.neuron.2012.05.010.

Zeithamova, D., Preston, A.R., 2017. Temporal proximity promotes integration of overlapping events. Journal of Cognitive Neuroscience 29 (8), 1311–1323. https://doi.org/10.1162/jocn_a_01116.

Zöllner, C., Klein, N., Cheng, S., Schubotz, R.I., Axmacher, N., Wolf, O.T., 2022. Where was the toaster? A systematic investigation of semantic construction in a new virtual episodic memory paradigm. Quarterly Journal of Experimental Psychology 76 (7), 1497–1514. https://doi.org/10.1177/17470218221116610.

Zwaan, R.A., 2025. From words to worlds: Twenty-five years of advances in situation model research. Current Directions in Psychological Science 34 (5), 287–292. https://doi.org/10.1177/09637214251326812.