Contents lists available at ScienceDirect

NeuroImage

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

The benefit of being wrong: How prediction error size guides the reshaping of episodic memories

Marius Boeltzig [a,b,*], Nina Liedtke [a,b], Sophie Siestrup [a,b], Falko Mecklenbrauck [a,b], Moritz F. Wurm [c], Inês Bramão [d], 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] CIMeC – Center for Mind/Brain Sciences, University of Trento, Corso Bettini 31, 38068 Rovereto, Italy
[d] Department of Psychology, Lund University, Allhelgona Kyrkogata 16a, 223 62 Lund, Sweden

ARTICLE INFO

Keywords:
Prediction error
Episodic memory
Memory modification
Memory integration
Pattern separation
Representational similarity analysis

ABSTRACT

Episodic memories are not static – they shift and reshape as our surroundings evolve. One powerful mechanism for change are prediction errors, which arise when predictions about what is going to happen next do not match the actual input. This study investigated how the size of prediction errors – arising from predictions based on episodic memories – affects recognition memory and neural memory representations. In an fMRI experiment, participants listened to a series of naturalistic dialogues. In a later session, a critical segment of the dialogue was altered to introduce a mismatch and thus evoke a prediction error. Importantly, larger prediction errors were linked to increased recognition memory for the original and mismatching targets, and better source memory for the mismatching targets. Representational similarity analysis revealed that larger prediction errors were also associated with stronger reinstatement of the original version during mismatching (unpredicted) input, which promoted memory for both the old and the new version. Additionally, larger prediction errors enhanced the long-term representational stability of the original memory. We argue that these results support the idea that stronger episodic prediction errors lead to a more distinct encoding of new information, which benefits the recognition of both old and new information. This could be achieved by a pattern completion mechanism in which old information is reinstated during mismatching new input.

1. Introduction

To adequately adapt to an ever-changing environment, our brains continuously need to dynamically update knowledge and experiences as new events render old ones outdated. Accordingly, episodic memories are not fixed and static after having been encoded but can undergo fundamental change through a wealth of different processes (Anderson and Hulbert, 2020; Loftus, 2005; Nadel et al., 2012). One powerful trigger for change in episodic memories occurs when they are used for predicting future events but fail to do so accurately. This study examines how the magnitude or size of prediction errors shapes memory outcomes and neural representations of memory traces. Using conversations as naturalistic and temporally structured stimuli, we assess both what is remembered and how this is implemented neurally, as a function of prediction error size.

To operate efficiently even under complex conditions, the brain is thought to constantly make predictions about incoming sensory input based on prior knowledge and experiences (Brown and Brüne, 2012; Bubic et al., 2010; Friston and Kiebel, 2009; Huang and Rao, 2011). If predictions are correct, fewer resources need to be dedicated to the processing of the stimulus. However, if mismatching input is detected, the prediction is wrong, the pre-existing models are therefore insufficient for explaining the current state of the environment, and a prediction error (PE) arises. This PE, which does not necessarily need to reach consciousness, can signal the need to explore the unpredicted stimulus and to update the existing models, so that they better reflect the current state of the environment (Fernández et al., 2016).

Episodic memories are a key source from which such internal models and the predictions they generate can be derived (Forcato et al., 2007; Quent et al., 2021; Sinclair and Barense, 2018). Episodic PEs can have

* Corresponding author at: Department of Psychology, Fliednerstraße 21, 48149 Münster, Germany.
E-mail addresses: marius.boeltzig@uni-muenster.de (M. Boeltzig), nina.liedtke@uni-muenster.de (N. Liedtke), s.siestrup@uni-muenster.de (S. Siestrup), f_meck@uni-muenster.de (F. Mecklenbrauck), moritz.wurm@unitn.it (M.F. Wurm), ines.bramao@psy.lu.se (I. Bramão), rschubotz@uni-muenster.de (R.I. Schubotz).

https://doi.org/10.1016/j.neuroimage.2025.121375
Received 11 April 2025; Received in revised form 20 June 2025; Accepted 15 July 2025
Available online 16 July 2025

several consequences. On the one hand, the original memory that formed the basis of the prediction and now has been identified as outdated by the PE can be weakened (Forcato et al., 2007; Kim et al., 2014; Kim et al., 2019; Vlasceanu et al., 2018). On the other hand, the unpredicted mismatching input may offer new information about the environment and can therefore be encoded (Bein et al., 2021; Brod et al., 2018; Greve et al., 2017; Stawarczyk et al., 2020). How these two outcomes interact is currently not well understood as previous studies have typically examined memory for either original or mismatching new information in isolation or used testing strategies not aimed specifically at one or the other process (Jainta et al., 2022,2024; Siestrup et al., 2022, 2023; Siestrup and Schubotz, 2023; Sinclair et al., 2021; Sinclair and Barense, 2018). Furthermore, neither original weakening (Sinclair et al., 2021; Sinclair and Barense, 2018), nor new learning (Liedtke et al., 2025; Ortiz-Tudela et al., 2023; Varga et al., 2025) are universal outcomes after all PEs, raising the question of what moderates between different mnemonic consequences.

One factor that may help to explain the previous disparate findings is the size of the PE (Bein et al., 2023; Nolden et al., 2024). Therefore, the current study aimed to investigate the effect of a broad spectrum of episodic PE sizes on memory performance and neural representations. Specifically, we set out to extend previous findings and provide a comprehensive test of the Latent Cause Theory (Gershman et al., 2017). Our study uniquely (a) measured episodic PE size continuously using both prior precision and prior accuracy; (b) tested recognition and source memory for both the original and the mismatching new version of each stimulus; and (c) assessed both original memory reinstatement during processing of the mismatching input and the long-term representational consequences for the original memory trace. We briefly outline these aspects and their relevance to the field of episodic PEs in the following.

There are several components contributing to PE size (Greve et al., 2017; Henson and Gagnepain, 2010), two of which are targeted in the current study. First, prior precision refers to how specific a prediction is. When a broader range of outcomes is expected with some probability, the PE will be smaller compared to when one single and highly probable outcome is predicted – provided that the input does not match the prediction. In other words, when a highly likely and strongly predicted event does not materialize, the PE will be larger than when there was less certainty in the prediction and a range of different outcomes was deemed likely. Second, prior accuracy refers to how much the prediction and the input differ, with larger PEs occurring when the input is less similar to the prediction. Prior accuracy therefore denotes how well a prediction matches the input, and larger PEs are produced when prediction and input are very different from each other. These two factors, prior precision and prior accuracy, jointly contribute to PE size, with the largest PEs occurring when predictions are highly specific and highly inaccurate (Henson and Gagnepain, 2010).

In the Latent Cause Theory, Gershman et al. (2017) argue that the size of PE critically determines representational stability of the original memory trace, and thereby also memory performance. They suggest a U-shaped relationship between PE size and original stability. Specifically, after small PEs, the original memory can account for the new input and the original trace is unchanged or even strengthened. No or little new information is encoded. Medium PEs modify the pre-existing memory by integrating new information into the existing model, because the model is still valid enough to apply to the situation but needs adaptation. The neural representation of the original memory trace would therefore change in the process. In the case of large PEs, the model seems not to apply to the situation, leading the system to infer a different underlying latent cause. This triggers the need of establishing a separate new model, while the original model representation remains unchanged, under the assumption that both are valid under different circumstances. Following this theory, the original representations are therefore more stable after small and large PEs, and more vulnerable after medium PEs, while there would be separate encoding of the new event only after larger PEs.

Consequently, this theory would predict better memory for the original and mismatch event after large PEs compared to medium PEs. After small PEs, memory for the original event would be high, while the new event, without having been encoded, could be recognized based on the highly similar original memory trace. There is first evidence supporting these assumptions. Kim et al. (2014) used a continuous neural measure of prior precision and found behavioural weakening of the original memories after moderate PEs, compared to small PEs. These findings were further supported by a recent study, in which medium PEs, quantified by a continuously measured prior accuracy, were found to evoke a distinct neural response, associated with lower original and mismatch version memory (Liedtke et al., 2025). While these fMRI studies only examined prior precision or prior accuracy, respectively, a behavioural study explored their interaction by manipulating prior accuracy continuously and prior precision in two levels. For the original version, memory was lowest after medium PEs, which were produced by the interplay of prior precision and prior accuracy. Similarly, after subtle modifications and large substantial changes, mismatch recognition memory was high (Boeltzig et al., 2025). However, a study measuring both prior precision and prior accuracy in a continuous manner to model a broad space of different PE sizes, albeit necessary to test the Latent Cause Theory, is currently lacking.

Furthermore, existing studies do not allow conclusions about which representational formats support the high memory levels after large PEs. The model by Gershman et al. (2017) predicts integration of new information into an existing model after medium PEs and distinct encoding of the new event after large PEs. Interestingly, evidence and arguments for both integration (Greve et al., 2018; Wahlheim et al., 2019; Wahlheim and Zacks, 2019) and distinct encoding (Bein et al., 2020; Frank et al., 2020; G. Kim et al., 2017) have been presented, without taking PE size into account. The Latent Cause Model argues that these outcomes are not mutually exclusive, but a direct test of this within the same study is missing.

In previous studies, it has also been difficult to disentangle the process of mismatch perception from long-term representational consequences. Some studies (e.g., Stawarczyk et al., 2020) assessed representations only while a prediction was being established, just before mismatching input was presented. As reinstatement of the original before or during mismatching input can have a broad range of consequences (Brunec et al., 2020; Zeithamova and Bowman, 2020), including the original being protected from interference (Chanales et al., 2019; Kuhl et al., 2010), it is difficult to discern the long-term representational consequences of this reinstatement. In that vein, other studies (G. Kim et al., 2014; H. Kim et al., 2019) have measured representational consequences of PEs for the original memory trace within the same session, while it is possible that these only unfold over a longer delay and after consolidation during sleep (Nadel et al., 2012; Schlichting and Preston, 2016). To assess both the time period in which the PE is generated, as well as long-term representational consequences, this study therefore measured original representations one day after PE induction in addition to the mismatch phase, in which a prediction is compared to new input.

To address these aims experimentally, we designed a paradigm that mirrors real-life situations involving ongoing predictions during continuously unfolding events, using auditorily presented naturalistic and socially charged dialogues. Participants first listened to dialogues in their original form. In a second session, the same dialogues were replayed, but some included an altered target statement in the middle. Thus, the identical beginnings of each dialogue enabled the automatic formation of predictions about the continuation, which were then violated by the mismatching input to elicit an episodic PE. We used representational similarity analysis (RSA; Kriegeskorte et al., 2008) to measure prior precision (G. Kim et al., 2014, 2017; Stawarczyk et al., 2020), but also to test for representational similarity between the original target and the mismatching target. In addition to representational similarity due to perceptual similarity in the stimuli themselves, we expected to capture differences in reinstatement of the original target during mismatching input.

We tested two consequences of PE size and original reinstatement at mismatch, based on the predictions of the Latent Cause Theory (Gershman et al., 2017). First, using RSA, we measured long-term representational stability of the original target. To that end, we played the original targets once again in a third fMRI session, to assess whether PEs induced changes to the memory traces. Second, in the fourth session, we tested recognition memory for both the original and mismatch targets, as well as source memory.

For the RSA measures (i.e., prior precision, original reinstatement at mismatch, original stability), we used three different regions of interest (ROIs). First, the inferior frontal gyrus (IFG) is a key hub for the processing of PEs in the domain of episodic memory and beyond (El-Sourani et al., 2020; Jainta et al., 2024; Liedtke et al., 2025; Schliephake et al., 2021; Wurm and Schubotz, 2012). As the IFG has been implicated in maintaining a predictive model (Fujitani et al., 2024), we expected it to reflect prior precision. Second, the hippocampus as the key region involved in episodic memory dynamically switches between retrieval and encoding (Richter et al., 2016) and can protect old memories from interference by retrieving them during novel input (Chanales et al., 2019; Kuhl et al., 2010). Lastly, given that memory retrieval is accompanied by a reinstatement of the activation pattern at encoding (Bosch et al., 2014; Bramão et al., 2022; Thakral et al., 2015), we also assessed whole-brain representations. Therefore, for measuring original reinstatement at mismatch and original stability, both the hippocampus and whole brain were used as ROIs, and for long-term original stability, we used the whole brain.

In line with the outlined research and theory, we had four predictions. First, higher prior precision and lower prior accuracy should lead to better learning of the mismatching information and retention of the original. Second, larger PEs should be related to stronger reinstatement of the original target during mismatching input. Third, this increased reinstatement of the original target should protect it from interference, resulting in increased memory performance. Fourth, larger PEs should lead to a more stable and unchanged original memory trace in the long term.

2. Methods

2.1. Participants

Based on the sample size of a previous study with a similar paradigm (Boeltzig et al., 2025), we recruited 50 participants. After excluding data from two participants due to high movement in the scanner, one participant due to technical problems, and five participants due to non-completion of the study, the final sample size was $N = 42$ (35 female, 7 male; age: M=21.98, $SD = 3.30$, range: 18–31). All participants were right-handed, had normal or corrected-to-normal vision, had no self-reported psychological or neurological diseases, were native speakers of German, and university students.

All participants were informed about their rights, with extensive information on the MRI scanning procedure, and provided written consent to participate in the study. After the experiment, they were debriefed and compensated with money or partial course credit. The Ethics Committee of the Faculty of Psychology and Sports Science at the University of Münster approved the study procedure.

2.2. Material

The material (Boeltzig et al., 2025; Liedtke et al., 2025) consisted of 36 naturalistic and socially charged dialogues in the German language. They were written by the authors, recorded by 20 professional voice actors and were auditorily presented to participants together with a unique content-matching background sound. The dialogues ranged in duration from 21 to 34 s ($M = 27.31$, $SD = 3.02$). Auditory samples of the stimuli can be requested from the corresponding author, and an example can be found in Supplementary Note 1.

Each dialogue was constructed in five versions. Participants first encoded the original, covertly consisting of a head, target, and end. To evoke PEs of different sizes, the target could then be modified in four different ways, constituting either a mismatch in phrasing or meaning, each of smaller or stronger magnitude. The different targets always matched the heads and ends, which never changed. Each participant encoded only one modified version per dialogue and the choice of modification for each dialogue was counterbalanced across participants.

Participants encoded 30 original dialogues, six of which were unchanged, so that no PE was induced. The other 24 dialogues were later presented in a modification, with six dialogues in each modification condition of low or high surface or gist changes. However, the experimental factors were not pre-assigned. Prior precision was measured via the reinstatement of the original target during the presentation of the head in Session 2 (Kim et al., 2014), so neurally, continuously, and on a trial-to-trial basis. Prior accuracy was operationalized through difference ratings obtained from an independent sample ($N = 120$) with no previous exposure to the material (Supplementary Note 2). These ratings could range from 1 for the smallest changes to 7 for the highest changes and were averaged across participants to be used as a measure of prior accuracy in the data analysis.

2.3. Procedure

The study was conducted in four sessions (Fig. 1A), spread out over nine days. For the three MRI sessions, the experiment was implemented in Presentation (Version 23.0, Neurobehavioural Systems) and for the behavioural session, PsychoPy (Peirce et al., 2019) was used. Participants always listened to the stimuli via headphones and self-adjusted the volume at the beginning. Each session lasted between 45 and 75 min.

In Session 1, participants encoded each of the dialogues twice in their original form. They were asked to listen naturally, as if witnessing the conversation, and to imagine the scene. No learning instruction was given and there was no mention of a later memory test. The first encoding run was carried out in the MRI scanner. After a jittered fixation cross (5.5–7 s), the dialogue was presented while a fixation cross remained on the screen. As a cover task, a word was shown after each dialogue, and participants were asked to indicate with a button press whether it had been said in the dialogue or not (4 s time limit). Over the course of the experiment (first encoding in Session 1, two encodings in Session 2, one encoding in Session 3), each dialogue was probed twice with a word that was actually used (from the head or end) and twice with a word that was not used, but was plausible, with the order pseudorandomized. The second encoding run of Session 1 was carried out at a laptop. Instead of indicating whether a specific word occurred, participants were asked to respond to five questions capturing impressions of each dialogue, which were obtained for a separate investigation.

Two days later, in Session 2, 24 dialogues were presented in a modified version in order to evoke a PE, and six dialogues were presented in their original version. There were also six additional new dialogues presented to facilitate a univariate analysis of brain responses to PEs (Liedtke et al., 2025), which are not analysed here. Importantly, all dialogues always started as they had in the previous session and the mismatch only pertained to the target statement around the middle of the dialogue. This allowed establishing a prediction of the original target while the dialogue unfolded, thus laying the ground for a PE. Participants were not informed about the dialogues being modified and the cover task was the same as in the scanned part of Session 1. Each dialogue was presented twice, in two blocks which were randomized independently from each other, and the full session took place in the fMRI scanner.

On the next day, in Session 3, the original versions were played once again, with all parameters and instructions being the same as in Session

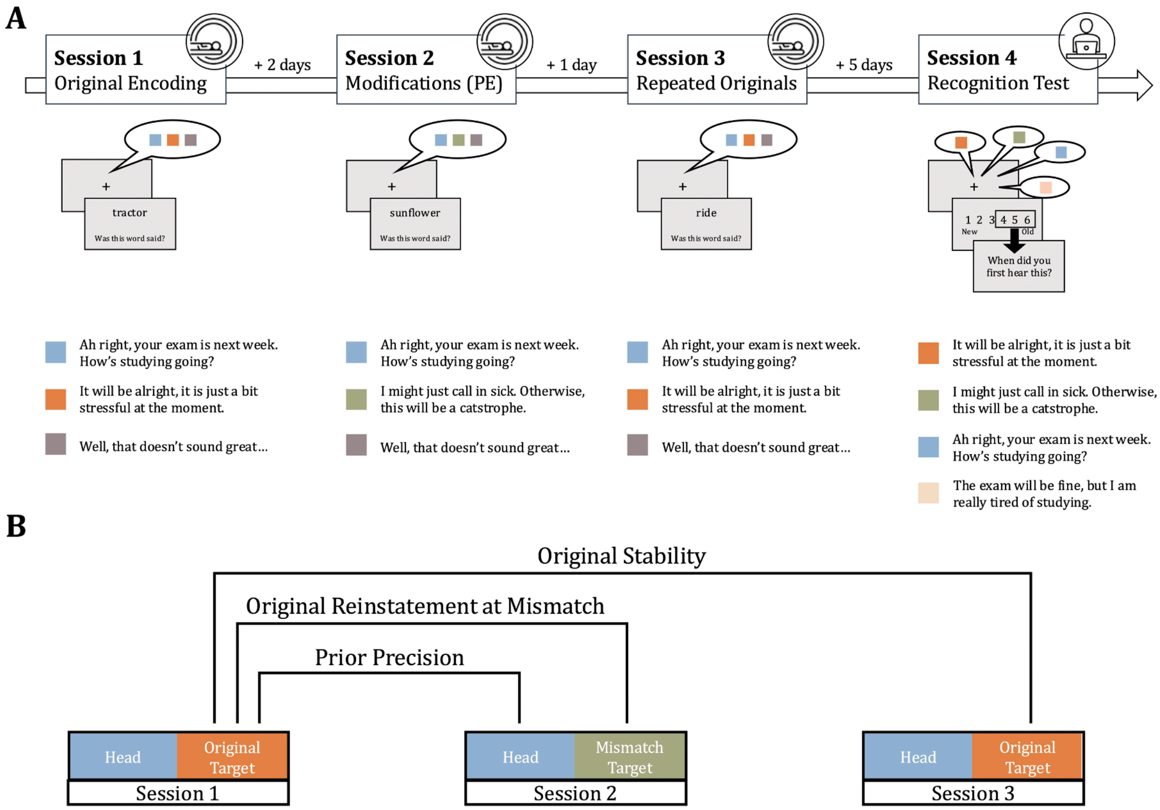
A
Session 1
Original Encoding
+ 2 days
Session 2
Modifications (PE)
+ 1 day
Session 3
Repeated Originals
+ 5 days
Session 4
Recognition Test
tractor
Was this word said?
sunflower
Was this word said?
ride
Was this word said?
1 2 3 4 5 6
New Old
When did you first hear this?
Ah right, your exam is next week. How's studying going?
It will be alright, it is just a bit stressful at the moment.
Well, that doesn't sound great...
Ah right, your exam is next week. How's studying going?
I might just call in sick. Otherwise, this will be a catstrophe.
Well, that doesn't sound great...
Ah right, your exam is next week. How's studying going?
It will be alright, it is just a bit stressful at the moment.
Well, that doesn't sound great...
It will be alright, it is just a bit stressful at the moment.
I might just call in sick. Otherwise, this will be a catstrophe.
Ah right, your exam is next week. How's studying going?
The exam will be fine, but I am really tired of studying.
B
Original Stability
Original Reinstatement at Mismatch
Prior Precision
Head
Original Target
Session 1
Head
Mismatch Target
Session 2
Head
Original Target
Session 3

Fig. 1. Experimental Procedure & Representational Similarities of Interest.
Note. **A.** The experiment took place in four sessions. In Session 1, participants were made familiar with the original versions of the dialogues, played once in the scanner, and another time outside of the scanner. In Session 2, the dialogues began in the same way, but instead of the original target (in orange), the head was followed by a mismatch target (in green). In Session 3, we again played the original versions to compare the original representations before and after the PE. As a cover story to keep attention high, after each dialogue in Sessions 1 – 3, a word was shown on the screen and participants had to decide whether it had been part of the dialogue or not. In Session 4, recognition memory was tested for the original and mismatch targets, for the head, and for similar lures. Participants responded on a scale from 1 = definitely new to 6 = definitely old. If participants indicated that the dialogue was old, they were additionally asked when they first heard this version of the dialogue, as a measure of source memory. Note that the example is a translated excerpt of a dialogue – the head (in blue) and end (in grey) consisted of more than one utterance by more than one speaker. **B.** The three representational similarities of interest. Prior precision was defined by the similarity between the original target from Session 1 and the head in Session 2. Original reinstatement at mismatch was the similarity between the two targets. Original stability was defined as the similarity between the original targets in Sessions 1 and 3. Note that each of these representational similarities will also be influenced by perceptual similarity of the material and other factors such as context similarity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

1. This session therefore allowed for a comparison of original memory representations before and after the PE induction.

In Session 4, which took place five days later in the behavioural lab, a surprise memory test was conducted. Participants listened to 144 short auditory probes taken from the dialogues (2.4–6.9 s, $M = 4.55$ s) and decided whether they had heard this exact utterance during any of the experimental sessions, on a scale from 1 = definitely new to 6 = definitely old. To minimize strategic answering, participants were asked to make an independent decision for each probe and disregard their previous answers on similar probes. Each dialogue was tested with four probes. For the dialogues in which a PE was evoked, these were the original and mismatch target, both of which should be classified as old by participants, and two similar lures from unused dialogue modifications which were employed to introduce a need for calibrated recognition decisions but were not analysed. The dialogues that were played always in the same original version had no mismatch target, and a probe from the head was presented instead. The test was covertly structured in four blocks, each of which contained one of the probes from each dialogue to space out material from the same dialogue. The order of probes was counterbalanced and controlled for in the analyses (Supplementary Note 3).

To obtain a measure of source memory, when participants responded that they had heard a probe before, they also indicated when they first heard this probe. They had two options to choose from, namely Session 1 or Session 2, along with the days of the week on which they happened.

2.4. MRI data acquisition

The three MRI sessions were carried out 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, and the head was cushioned out to avoid head movement. The right index and middle fingers were placed on a button box.

Each scanning session began with the acquisition of a high-resolution T1-weighted anatomical image using a 3-D magnetization rapid gradient echo sequence (192 slices of 1 mm, repetition time = 2140 ms, echo time = 2.28 ms, flip angle = 8°, field of view = 256 × 256 mm^2). The functional images to measure BOLD contrasts were obtained along the AC-PC plane in interleaved order using a gradient-echo planar imaging sequence (33 slices of 3 mm, repetition time = 2000 ms, echo time = 30 ms, flip angle = 90°, field of view = 192 × 192 mm^2). A temporal high-pass filter of 128 s was applied, removing low-frequency noise from the time series.

Participants were excluded from analysis if in one of the three sessions, continuous movement on any axis exceeded 4 mm (zero participants) or if spikes (i.e., movement between two consecutive acquisitions) exceeded 2 mm (two participants). Especially within the context of three separate fMRI sessions with each participant, we followed the recommendation to exclude participants only in extreme movement cases (Poldrack et al., 2024).

2.5. MRI data preprocessing for representational similarity analysis

To increase sensitivity of the representational similarity analysis (RSA) measures, we refrained from normalization and instead used every participant's anatomy to extract individually tailored regions of interest (ROIs). To that end, the T1-weighted anatomical images from the first session were used for a freesurfer (Fischl, 2012) parcellation before extracting the bilateral inferior frontal gyrus (IFG; defined by the pars opercularis, pars orbitalis and pars triangularis; voxel count in 3 × 3 × 3 resolution: $M = 667.76$, $SD = 59.51$) and the bilateral hippocampus (voxel count: $M = 225.48$, $SD = 22.08$) using the Desikan-Killiany Atlas. Additionally, a whole brain analysis, including all voxels (voxel count: $M = 131101.81$, $SD = 1370.35$) was performed.

SPM 12 (Friston et al., 2007) in MATLAB (Version R2024a; MathWorks Inc.) was used to prepare the functional data for the RSA. Only slice time correction to the temporal middle slice and movement correction were applied to the functional images from all three sessions. We slice-time corrected to the middle slice to minimize both temporal and spatial interpolation errors, as the temporal middle slice was also the spatial middle slice. The biggest error would therefore apply to the top and bottom slices with less tissue instead of the more crucial middle slices (Sladky et al., 2011). To facilitate the analysis of the hippocampus as a small region, data was not smoothed (Dimsdale-Zucker and Ranganath, 2018; Weaverdyck et al., 2020). Next, to have all images in the same individual space, the anatomy from Session 1 was co-registered along with the individual ROIs and all functional images to the mean functional image from Session 1. Normalized Mutual Information values indicated good realignment for both sessions for all participants ($M = 1.34$, $SD = 0.03$).

This data was then used to specify a generalized linear model where the head and the target of each dialogue from all sessions formed its own regressor using its full duration. In Session 2, where each dialogue was played twice, we therefore averaged the data across the two presentations. In addition to the 96 heads and 96 targets, there were three additional parameters denoting each session, and six movement parameters for each session as regressors of no interest.

The resulting beta images were then used for RSA using the CosmoMVPA toolbox (Oosterhof et al., 2016). For each of the ROIs and for each participant, a 192 × 192 similarity matrix was calculated by extracting the beta patterns for the 192 conditions (30 dialogues with one head and one target in each session, plus the six new dialogues in Session 2 with a head and a target, which are not used for analysis), demeaning each beta pattern (by subtracting the mean of all voxels from each voxel), and correlating the beta patterns for all pairwise combinations using a Pearson correlation coefficient.

From these matrices for each ROI, the three similarities of interest were extracted (Fig. 1B). First, as a measure of prior precision, we used similarities between the target from Session 1 and the head from Session 2 of each dialogue. Second, we extracted the similarity between the original target from Session 1 and mismatch target from Session 2. With this similarity, we attempted to capture differences in reinstatement of the original target while the mismatch target is being processed. Lastly, as a measure of original stability, the similarity between the original targets of Session 1 and 3 was used. Note that each of these similarities will also depend on other factors, most notably the perceptual similarity between the stimuli and overall context similarity. However, this was constant for stimuli of all PE sizes.

Each of these extracted similarities for each participant and each dialogue was then baseline-corrected to account for coincidental and overall similarity in the stimulus material. Specifically, we subtracted from each similarity of interest the mean of the similarities of the involved cues or targets with all other cues or targets from the relevant session. For instance, prior precision was measured by the similarity between the target of a dialogue in Session 1 with the head of that dialogue in Session 2. That similarity of interest within the same dialogue was contrasted against the mean of 58 other similarities from the matrix, namely 1) the similarity between the target in Session 1 with the 29 other heads of Session 2 (all except the head belonging to the same dialogue), and 2) the similarity between the head of the dialogue in Session 2 with the 29 other targets of Session 1 (all except the target belonging to the same dialogue). The mean of these 58 similarities was subtracted from the similarity of interest. The reported values are therefore not the raw correlation coefficients from the RSA but reflect how much more similar two given events are than what would be expected if they were not related to each other (see Shao et al., 2023 for a similar measure).

2.6. Behavioural data preprocessing

Participants responded to the probes during the recognition test in Session 4 using a scale from 1 = definitely new to 6 = definitely old. For the original and mismatch targets, which had actually been played before, we transformed these answers to a binary variable where 0 = incorrect (when participants chose options 1–3) and 1 = correct (when 4–6 were chosen) to use a dependent variable free from confidence judgements. In addition to recognition of original and mismatch targets separately, we also analysed memory for both versions of the same dialogue. In that variable, trials were only coded as correct when both versions were correctly recognized. Source memory judgements were coded as 0 = incorrect or 1 = correct, using trials in which participants correctly classified the probe as old.

2.7. Data analysis

The current analysis is based on the same data presented in a previous paper (Liedtke et al., 2025). The data used for the current analysis is available on OSF (https://osf.io/wnzs3/?view_only=8a7e9cf626aa42eeb9cb83c5e386ef41).

Data analysis was carried out in six steps. First, we tested the effect of prior accuracy, as measured by the difference ratings, on recognition and source memory for the original and mismatch targets. Second, the effect of prior precision on the same dependent variables was tested. In a third step, the two factors were combined to test whether they interact in predicting memory. Fourth, it was tested whether prior accuracy and prior precision predicted original reinstatement at mismatch, so how similar the original and mismatch target were represented. As a fifth step, the effect of this original reinstatement on memory was tested. Sixth and last, we tested whether prior precision and prior accuracy influence long-term original stability, so the similarity between original targets in Sessions 1 and 3.

Note that the difference rating was used to operationalize prior accuracy. Bigger difference ratings signified lower prior accuracy, and therefore a larger PE. We opted not to inverse the difference ratings to harmonize interpretation and facilitate plotting along with prior precision. The regression coefficients and the figures can therefore be understood in a way that larger values signify larger PEs.

In all steps, when predicting the binary recognition and source memory outcomes, we used multilevel logistic regressions. The results therefore indicate effects on likelihood of correct memory judgements. In addition to the independent variables of interest, we also added random intercepts for each dialogue, to control for differences in memorability of the stimuli, and for each participant, accounting for baseline memory capacity. Additionally, we added covariates that arose from the experimental design, chosen in a data-driven approach (Supplementary Note 3). In steps four and six, where neural similarities were used as dependent variables, standard multilevel regression models were used, also including the random intercepts for participants and dialogues. For each model, *p*-values were FDR-corrected for the number of ROIs it was tested in. As we expected a U-shaped relationship between PE size and memory, polynomial regressions were tested in addition to linear regressions. In those cases where the PE variables were significant in both a U-shaped and linear model, we compared these models in an ANOVA and report conditional R^2 values. All models were calculated using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2025). We report the critical results in the manuscript, while the full models including conditional R^2 values as well as beta coefficients and partial R^2 for each predictor are shown in Supplementary Note 4, including the covariates of no interest.

In terms of representational similarities, for prior precision, we were interested in the IFG, while we limited the analysis of original reinstatement at mismatch to the hippocampus and whole brain ROIs, and the analysis of original stability to the whole brain. However, before conducting analyses in each step, for each representational similarity of interest (Fig. 1B), we tested whether it was reliably reflected within the relevant ROIs. Specifically, we assessed whether the similarity between the relevant parts of the same dialogue (e.g., original target in Session 1 and head in Session 2 as a measure of prior precision) was higher than similarities between these parts and all other dialogues (e.g., Session 1 target with all other Session 2 heads, and Session 2 head with all other Session 1 targets) in each given ROI. That same measure, so the mean of these unrelated similarities, was also used in the baseline correction described above. Using multi-level regression models with participants and dialogues as random intercepts, we thus contrasted the relevant similarities with the baseline similarities. Only if the relevant similarities stemming from the same dialogue were significantly larger than the baseline similarities in a given ROI, it was used for analysis of each effect.

3. Results

Responses to the cover task in the scanner revealed high attention as participants responded correctly to 89.36 % of the questions. On the recognition memory test, participants also performed well on all probe types (original targets: $M = 0.80$, $SD = 0.12$; mismatch targets: $M = 0.70$, $SD = 0.15$; heads: $M = 0.75$, $SD = 0.17$, lures: $M = 0.76$, $SD = 0.09$). A comparison between the changing dialogues (so those with a PE) and the unchanging dialogues revealed that PEs generally led to lower original recognition performance on the targets (unchanging: $M = 0.88$, $SD = 0.16$; changing: $M = 0.83$, $SD = 0.12$; $t(41) = 2.07$, $p = .044$, $d = 0.37$). A descriptive reduction of source memory did not gain significance (unchanging: $M = 0.83$, $SD = 0.23$; changing: $M = 0.78$, $SD = 0.15$; $t(41) = 1.91$, $p = .063$, $d = 0.37$).

3.1. Low and high prior accuracy are associated with increased mismatch target recognition memory

We first tested the effect of PE size on behavioural memory outcomes. PE size was measured via two distinct components, prior precision and prior accuracy, and we first analysed how the latter influenced memory. Prior accuracy was operationalized via difference ratings between the original and modified dialogue from an independent sample (Supplementary Note 2). Differences were measured on a scale from 1 to 7 where higher ratings correspond to lower prior accuracy, and therefore larger PEs.

For original targets, presented first in Session 1 and later used for prediction, leading to a PE, prior accuracy had no significant effect on the likelihood of recognition. This was evident both in a linear ($b = -0.11$, $OR = 0.90$, $p = .093$) and in a quadratic model (base term: $b = -0.65$, $OR = 0.52$, $p = .153$; quadratic term: $b = 0.07$, $OR = 1.08$, $p = .229$; Fig. 2A).

For the likelihood of recognising the mismatch target, presented in Session 2 to induce a PE, prior accuracy was not a significant linear predictor ($b = -0.05$, $OR = 0.95$, $p = .363$). However, a significant U-shaped relationship emerged between prior accuracy and mismatch target recognition (base term: $b = -1.96$, $OR = 0.14$, $p < .001$; quadratic term: $b = 0.26$, $OR = 1.30$, $p < .001$; Fig. 2A). This U-shaped model outperformed the linear model significantly ($X^2(1) = 25.08$, $p < .001$; linear $R^2_{cond} = .18$, U-shaped $R^2_{cond} = .22$). Consequently, mismatch target recognition was highest for both small and large changes, but lower for medium changes. To more directly test whether medium changes led to decreased mismatch target recognition, we ran a follow-up model in which the difference ratings used to measure prior accuracy were split into three categories (small, medium, strong differences) at the 33rd and 66th percentiles. Tukey-corrected post-hoc tests confirmed that medium differences were recognised worse than both small ($p = .001$) and large ($p = .005$) differences, while there was no difference between the two ends of the spectrum ($p = .929$).

When assessing the combined recognition memory for both original and mismatch target in the same dialogue, there was also a significant U-shaped relationship (base term: $b = -1.29$, $OR = 0.28$, $p < .001$; quadratic term: $b = 0.17$, $OR = 1.18$, $p < .001$). After small and large changes, participants were therefore also more likely to recognize both versions of the same dialogue. Again, prior accuracy as a linear regressor did not gain significance ($b = -0.09$, $OR = 0.91$, $p = .085$).

While recognition of the mismatch target was more likely after small and large changes, source memory for the mismatch target benefited only from larger changes. Prior accuracy was significant in both a linear ($b = 0.66$, $OR = 1.93$, $p < .001$) and a quadratic model (base term: $b = 1.71$, $OR = 5.56$, $p < .001$; quadratic term: $b = -0.15$, $OR = 0.86$, $p = .014$). In both models, higher differences (i.e., lower prior accuracy) led to better mismatch source memory. The quadratic model also accounted for a plateau towards stronger changes (Fig. 2B) and performed better than the linear model, even though the explained variance was almost identical ($X^2(1) = 5.96$, $p = .015$; linear: $R^2_{cond} = 0.32$, U-shaped: $R^2_{cond} = 0.32$). In contrast, original source memory was not significantly impacted by prior accuracy, neither in a quadratic (base term: $b = -0.82$, $OR = 0.44$, $p = .085$; quadratic term: $b = 0.11$, $OR = 1.12$, $p = .090$) nor in a linear model ($b = -0.02$, $OR = 0.98$, $p = .753$).

In summary, high and low prior accuracy were associated with better mismatch target recognition. Low prior accuracy (i.e., larger rated differences) also led to better source memory of the mismatch target. Original recognition memory and original source memory were not affected by prior accuracy.

3.2. Higher prior precision leads to better original and mismatch target recognition

After establishing the effect of prior accuracy, we next turned to prior precision as the second PE component and tested its behavioural consequences. Prior precision was operationalized by measuring the neural similarity between the original target presented in Session 1 and the head of the same dialogue presented in Session 2. Higher similarity indicates stronger reinstatement of the target, and therefore a stronger prediction of the original target (Kim et al., 2014). Once the mismatch target was then played, this prediction turned out to be incorrect, leading to a PE that was larger when the prediction was stronger.

We predicted that the IFG would reflect prior precision. The baseline contrast (see 2.7 Data Analysis) confirmed this, as sections from the same dialogues were significantly more similar than sections from unrelated dialogues ($b = 0.01$, $p = .038$). Note that this was not the case for the other ROIs (whole brain: $b = 0.005$, $p = .126$; hippocampus: $b = 0.004$, $p = .304$).

Higher values of prior precision led to better original ($b = 2.13$, $OR = 8.45$, $p = .001$) and mismatch target recognition ($b = 1.30$, $OR = 3.68$, $p = .017$; Fig. 2C). Consequently, recognizing both the original and mismatch targets for the same dialogue was also predicted by prior precision in the IFG ($b = 2.24$, $OR = 9.37$, $p < .001$).

There were no significant effects of prior precision on source memory

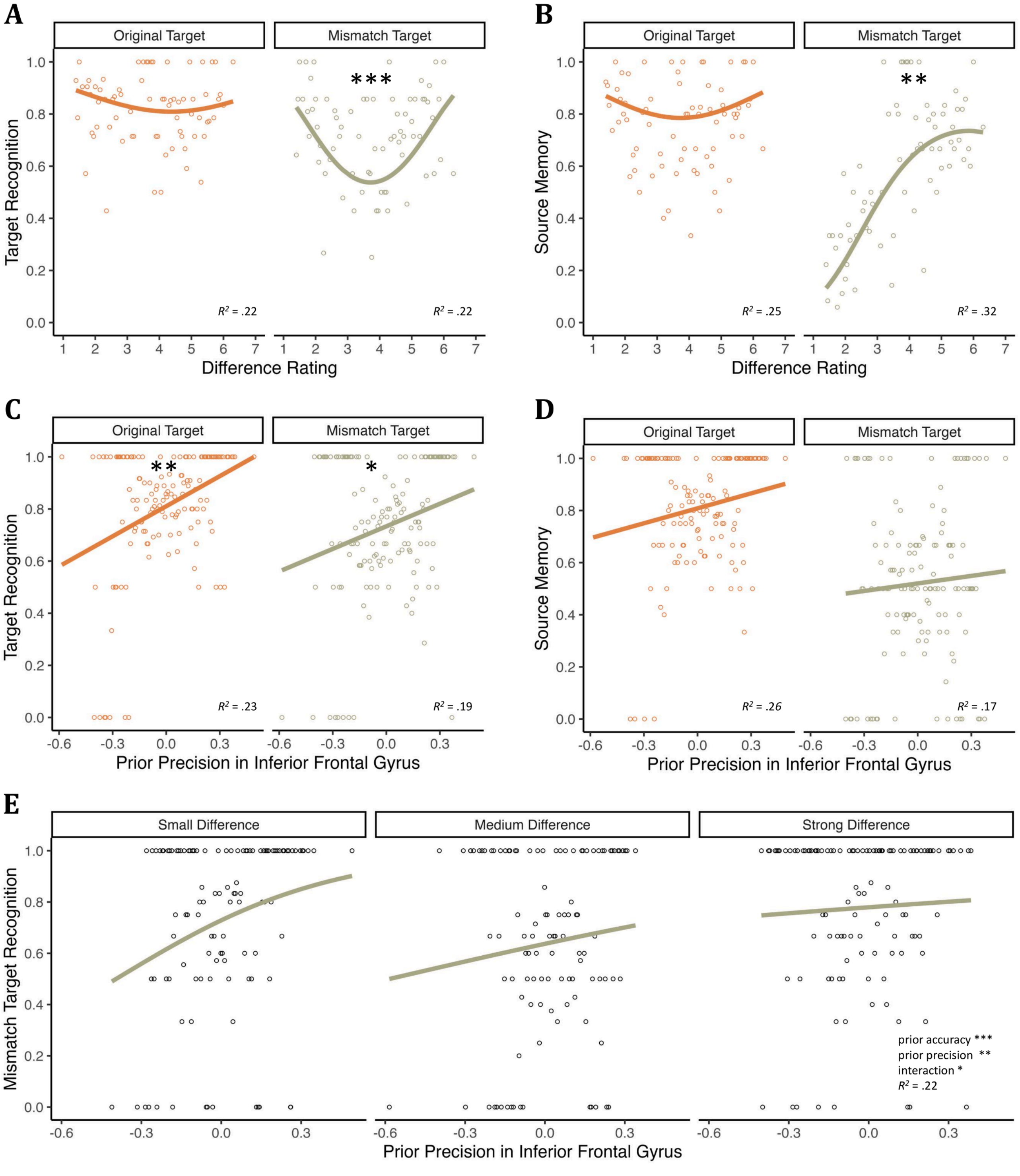

Fig. 2. Influence of Prediction Error Size on Memory Outcomes.

Note. All plots show memory performance aggregated across participants for a given level of the PE size measures. Note that the analysis was not conducted on aggregated data but using each trial within multi-level models. Conditional R^2 is indicated to evaluate the full model including random factors, and significance levels for the plotted variables are indicated with * < 0.05, ** < 0.01, *** < 0.001. **A.** The difference rating was used as a measure of prior accuracy. Higher ratings indicate stronger changes, so lower prior accuracy, and therefore larger prediction errors. The difference rating had an effect on the mismatch target, but not the original. The *p*-values are shown for the base term and the quadratic term in these polynomial regression models. **B.** The difference rating had a significant effect on source memory for the mismatch target, but not on the original target. **C.** Prior precision was measured via the original target reinstatement in the IFG during the head of the episode. Higher prior precision was associated with better memory for both the original and mismatch target. **D.** There was no significant influence of prior precision on source memory for either original or mismatch target. **E.** Joint influence of prior accuracy and prior precision on recognition memory for the mismatch target. The difference rating as a measure of prior accuracy is split into three distinct categories (small, medium, and strong differences). Both prior accuracy and prior precision had a significant effect and interacted with each other.

for the original ($b = 0.76$, $OR = 2.14$, $p = .258$) or the mismatch target ($b = -0.18$, $OR = 0.83$, $p = .758$; Fig. 2D). Note that the alternative polynomial regressions did not gain significance for recognition or source memory, indicating that there was no curvilinear relationship between prior precision and memory outcomes.

In summary, larger PEs, as measured by prior precision in the IFG, enhanced recognition memory for both the original and the mismatch target.

3.3. Both prior accuracy and prior precision predict mismatch target memory

Mismatch target recognition was enhanced both by low and high prior accuracy, and by stronger prior precision. We therefore conducted a follow-up analysis to determine whether these two factors jointly contributed to the likelihood of mismatch target recognition. Because the model using the base and quadratic term of the difference rating with an interaction with prior precision failed to converge, we instead included the difference rating in three discrete categories, based on a split at the 33rd and 66th percentiles (small, medium, strong differences). In this model, prior precision significantly predicted mismatch target memory ($b = 3.35$, $OR = 28.59$, $p = .001$). Regarding the effect of prior accuracy, medium changes led to significantly lower mismatch target recognition than the baseline low changes ($b = -0.70$, $OR = 0.50$, $p < .001$) while low and strong changes did not significantly differ ($b = -0.06$, $OR = 0.94$, $p = .754$). There was also a significant interaction between prior precision and the large changes ($b = -3.12$, $OR = 0.04$, $p = .026$). Descriptively, prior precision was especially beneficial when changes were small, as compared to larger changes (Fig. 2E). However, in a Tukey-adjusted post-hoc comparison of slopes between the levels of the categorical difference rating, this difference was short of significance ($p = .066$).

3.4. Large PEs lead to stronger original reinstatement at mismatch

After establishing the behavioural consequences of PEs of different sizes, we next turned to how they are achieved neurally. First, we assessed whether PE size influences how similarly original and mismatch target are represented (i.e., the similarity between the targets in Session 1 and Session 2). We expected that in addition to representational similarity due to perceptual similarity in the stimuli, this measure would also capture differences in reinstatement of the original

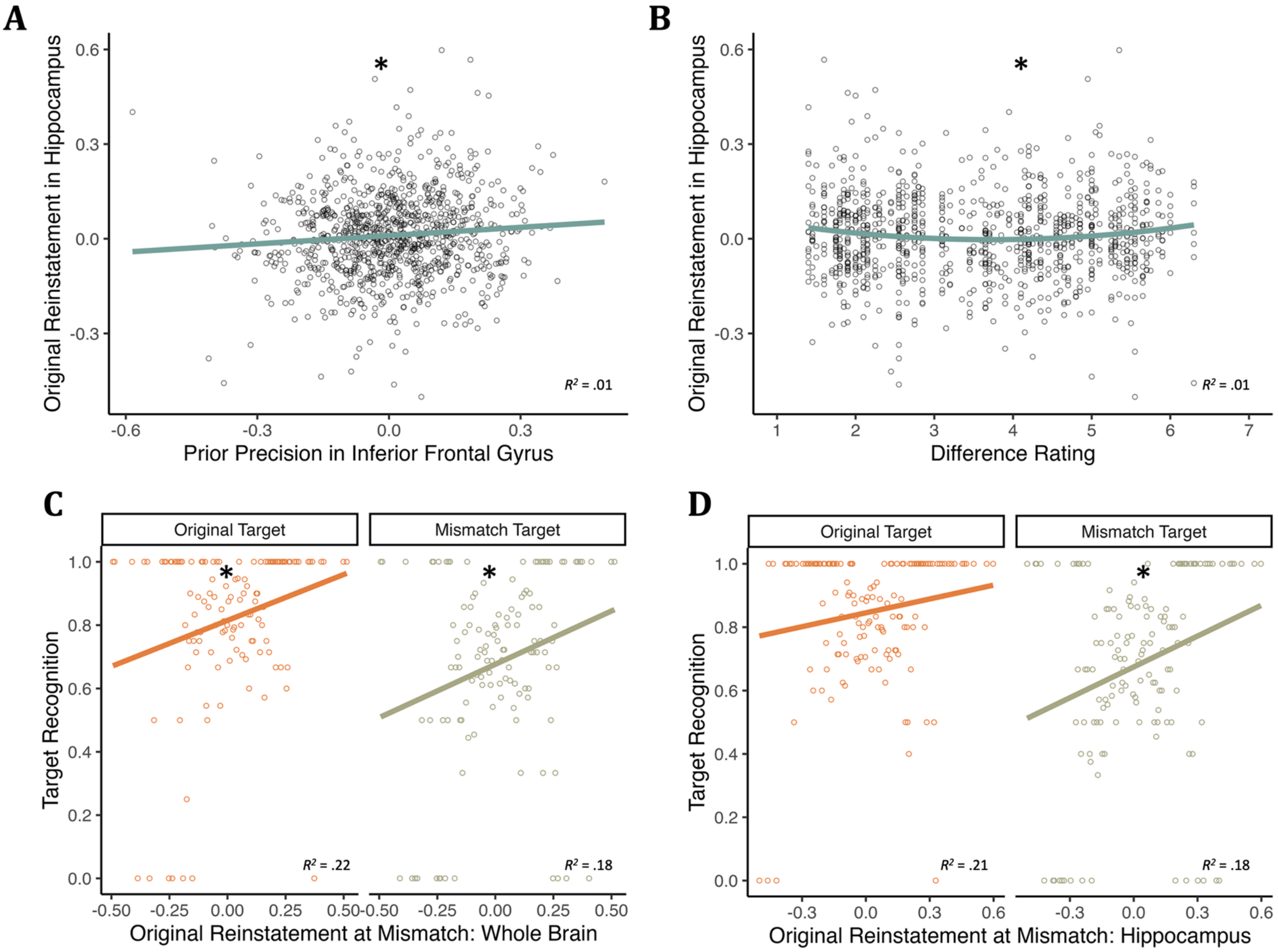

Fig. 3. Predictors and Consequences of Original Reinstatement at Mismatch.

Note. The first two plots show how prediction error size influences how much the original target is reinstated while the mismatch target is played in Session 2. Higher values indicate stronger original reinstatement. Each dot represents a single trial (panels A and B) or an aggregation of all participants per PE size level (panels C and D). Conditional R^2 is indicated to evaluate the full model including random factors, and significance levels for the plotted variables are indicated with * < 0.05, ** < 0.01, *** < 0.001. **A.** Prior precision had a significant effect on original reinstatement. The plot shows this relationship in the hippocampus, but the whole brain also showed this effect. **B.** The difference rating, as a measure of prior accuracy, had a significant curvilinear relationship with reinstatement in the hippocampus. Original and mismatch target were perceptually more similar after smaller changes, but there was more reinstatement also after large changes. **C.** Stronger original reinstatement in the whole brain is associated with higher recognition likelihood of original and mismatch targets. **D.** For the hippocampus, stronger original activation only significantly predicted mismatch recognition memory.

target during the mismatch target.

Testing first the validity of this measure in the two ROIs, we found that this similarity was significantly above the baseline in both regions (hippocampus: $b = 0.01$, $p < .001$; whole brain: $b = 0.01$, $p < .001$).

We then tested whether higher prior precision led to more original reinstatement at mismatch. This was the case in both ROIs (whole brain: $b = 0.18$, $p < .001$; hippocampus: $b = 0.09$, $p = .004$; Fig. 3A). Stronger prior precision was therefore associated with higher similarity between the targets.

In addition to prior precision, we also tested the effect of prior accuracy measured by the difference rating on original reinstatement during the mismatch phase. There were no linear effects of prior accuracy in the two ROIs ($ps > .658$). However, due to the curvilinear effect of prior accuracy on mismatch target memory, we also tested the influence of prior accuracy on this similarity using a quadratic model. In this model, a curvilinear prior accuracy was significant in the hippocampus (base term: $b = -0.05$, $p = .022$; quadratic term: $b = 0.01$, $p = .024$; Fig. 3B). This suggests that the hippocampus exhibited more similar activation between original target and mismatch target not only when the two versions were judged to be similar, but also when they were rated as highly dissimilar, possibly due to stronger reinstatement of the original target during mismatch target input. No such relationship was found for the whole brain (base term: $b = -0.01$, $p = .744$; quadratic term: $b = 0.00$, $p = .690$).

In summary, the original and mismatch targets were represented more similarly if prior precision was high, or when prior accuracy was low or high. Thus, in addition to perceptually similar targets, also highly unsimilar ones were represented similarly, possibly due to original reinstatement.

3.5. *Higher original reinstatement at mismatch leads to better recognition memory*

The previous analyses indicated that after larger PEs (i.e., higher prior precision and smaller prior accuracy), there was more original target reinstatement during the mismatch. We next aimed to test the behavioural implications of this reinstatement. Higher original reinstatement in the whole brain was associated with better original target recognition ($b = 2.11$, $OR = 8.24$, $p = .014$), but no such effect was found for the hippocampus ($b = 0.35$, $OR = 1.42$, $p = .600$). Better mismatch target memory was observed after more original reinstatement during the mismatch phase in the whole brain ($b = 1.38$, $OR = 3.96$, $p = .035$) and the hippocampus ($b = 1.44$, $OR = 4.23$, $p = .022$). Memory for the combined original and mismatch target recognition was also predicted by activity in both ROIs (whole brain: $b = 1.68$, $OR = 5.35$, $p = .008$; hippocampus: $b = 1.42$, $OR = 4.13$, $p = .008$; Fig. 3C and D). There were no effects of this similarity on source memory ($ps > .270$).

In summary, higher original reinstatement at mismatch generally increased recognition memory, but not source memory. As was seen in the analysis step before, higher representational similarity was not only caused by similar targets, but also by highly dissimilar ones, and by higher prior precision.

3.6. *Larger PEs lead to higher long-term original stability*

As a last step, we were interested in long-term stability of the original representations after experiencing a PE. To that end, we computed similarities between the original targets in Session 1 and Session 3, so before and after the PE. For the whole brain, which was used for analysis here, this similarity was significantly above baseline ($b = 0.02$, $p < .001$).

We first tested whether whole-brain original stability was influenced by PE size. First, prior precision in the IFG, which had been predictive of memory for both original and mismatch target, positively predicted original stability in the whole brain ($b = 0.07$, $p = .007$; Fig. 4A). Second, prior accuracy was predictive of whole brain original stability, with larger differences leading to fewer changes to the representation of the original ($b = 0.01$, $p = .0497$; Fig. 4B). A curvilinear model did not exhibit a significant effect of prior accuracy (base term: $b = 0.02$, $p = .380$, quadratic term: $b = -0.001$, $p = .556$).

When adding both factors into one model, it was mainly prior accuracy predicting original stability ($b = 0.01$, $p = .034$) as prior precision was below the threshold of significance ($b = 0.13$, $p = .070$). There was no interaction ($p = .381$).

There was no direct effect of original stability on recognition or source memory ($ps > .226$).

In summary, the original representations were more stable after bigger PEs, either when prior precision was high, or when prior accuracy was low.

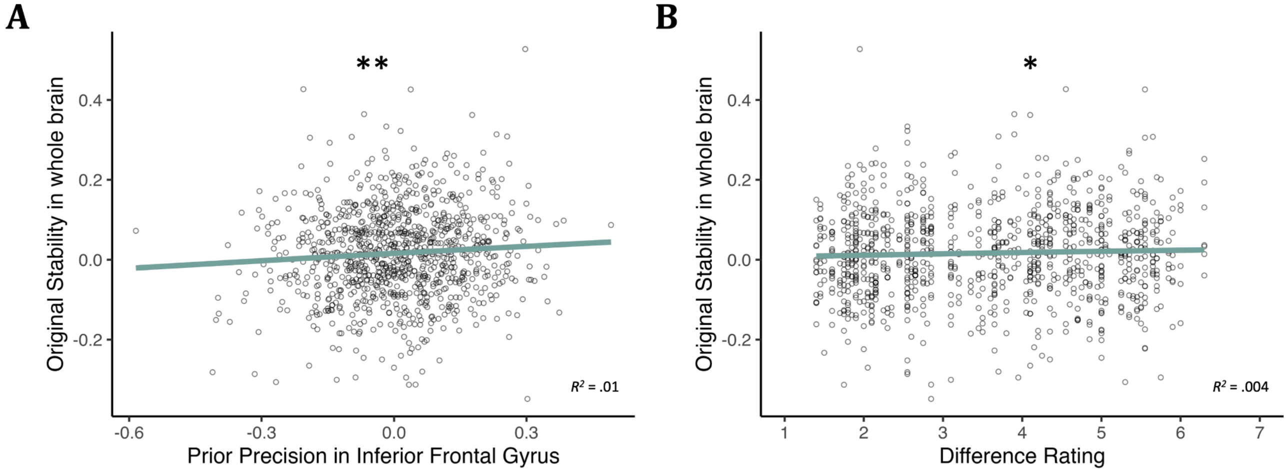

Fig. 4. Influence of Prediction Error Size on the Stability of the Original.
Note. Original stability was measured through representational similarity of the original targets before and after the prediction error. Higher values indicate higher similarity, and therefore less change. For both components of prediction error size, larger prediction errors are significantly associated with stronger stability in the whole brain. Each dot represents a single trial. Conditional R^2 is indicated to evaluate the full model including random factors, and significance levels for the plotted variables are indicated with * < 0.05, ** < 0.01, *** < 0.001. **A.** Higher prior precision in the IFG is associated with higher original stability. **B.** The difference rating as a measure of prior accuracy is associated with higher original stability.

4. Discussion

To better understand how episodic memories can undergo change, this study specifically examined the size of episodic prediction errors (PEs) and its mnemonic and representational consequences. Leveraging naturalistic and socially charged stimuli to continuously manipulate episodic PE size over a broad range, we found evidence for size-dependent effects of prior precision and prior accuracy on original and mismatch target memory (see Fig. 5 for an overview). While higher prior precision generally led to better original and mismatch target recognition, both low and high prior accuracy were associated with better mismatch target memory. Larger PEs also triggered stronger original reinstatement during exposure to the mismatch, which further was beneficial for original and mismatch target memory. Lastly, larger PEs led to higher stability of the original memory trace. As we will argue, and in line with the Latent Cause Theory (Gershman et al., 2017), the findings highlight that the size of PEs decisively shapes memory outcomes and representations, suggesting that larger PEs lead to more distinctive encoding of unpredicted information.

Two factors were measured to capture differences in PE size. First, prior accuracy (Henson and Gagnepain, 2010) was measured using difference ratings from a separate sample. Importantly, previous studies (Boeltzig et al., 2025; Liedtke et al., 2025) used difference ratings cast by the same participants, after they had been exposed to the material over multiple sessions. The independent ratings used here therefore crucially extend previous findings.

First, we tested the direct influence of prior accuracy, as measured by the difference ratings, on recognition and source memory for both original and mismatch targets. The hypothesis that original memory traces, would remain intact after small and large PEs, but weakened after medium PEs (Boeltzig et al., 2025; Gershman et al., 2017; G. Kim et al., 2014) was not directly supported. Prior accuracy had no direct effect on recognition memory. It is possible that encoding in two different spatial contexts (Imundo et al., 2020), consolidation during sleep (Abel et al., 2023; Nadel et al., 2012; Schlichting and Preston, 2016), and replay after the mismatch dialogues could have strengthened the originals, covering any PE-induced weakening. However, as discussed below, there was an effect of prior accuracy on reinstatement of the original during the mismatch, which in turn influenced original recognition memory, constituting an indirect effect of prior accuracy on memory for the original.

Memory for the new and mismatching version on the other hand was directly affected by prior accuracy. Recognition performance was improved for dialogues that underwent small or large changes compared to medium changes. Notably however, source memory increased linearly with the difference rating, with the highest source memory after large PEs. This supports an account of distinct encoding of a new memory trace predicted by the Latent Cause Theory (Gershman et al., 2017), as participants not only recognized the probe, but could also leverage detailed memory traces to indicate where they know it from. After smaller PEs, where the Latent Cause Theory predicts little or no learning, we found high recognition memory, but low source memory, indicating that participants may have failed to encode the mismatch version but correctly responded to it due to the high overlap with the original. This is in congruence with previous studies finding good memory for the mismatch version after subtle PEs that did not affect the overall gist of the episodes (Boeltzig et al., 2025; Liedtke et al., 2025; Siestrup and Schubotz, 2023). Additionally, the current findings align with work showing that both highly expected and highly unexpected events are remembered well (Brod et al., 2022; Kesteren et al., 2012; Quent et al., 2022). Our findings suggest that there are differences between those two ends of the spectrum, with detailed memories after unexpected events, and more generalized memories after expected ones.

Next, we tested how prior precision affected memory. Prior precision

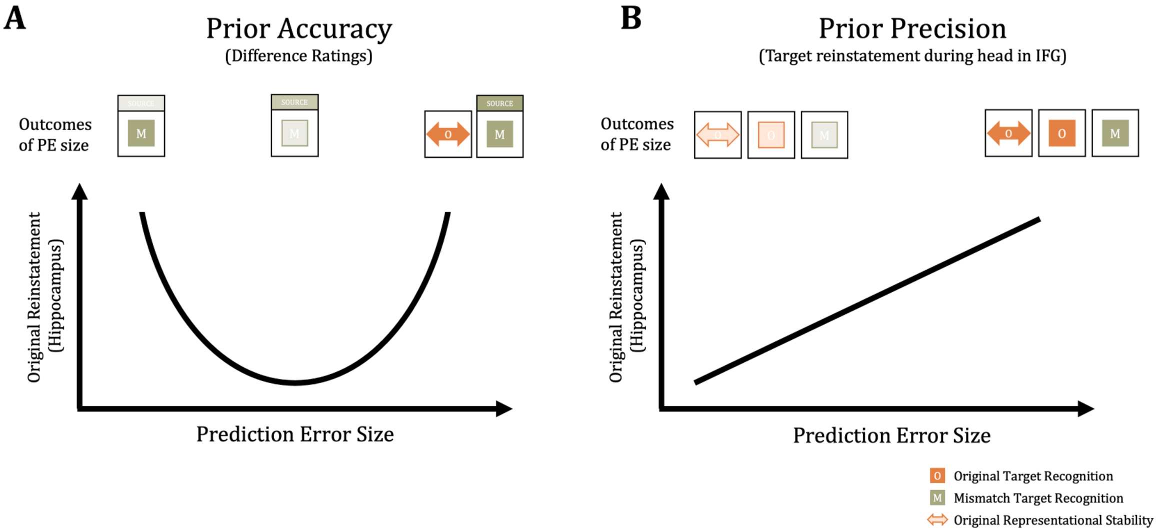
A
Prior Accuracy
(Difference Ratings)
Outcomes of PE size
SOURCE
M
SOURCE
M
O
SOURCE
M
Original Reinstatement (Hippocampus)
Prediction Error Size
B
Prior Precision
(Target reinstatement during head in IFG)
Outcomes of PE size
O
O
M
O
O
M
Original Reinstatement (Hippocampus)
Prediction Error Size
O Original Target Recognition
M Mismatch Target Recognition
⇔ Original Representational Stability

Fig. 5. Overview of Results.

Note. This schematic overview summarises the main findings of the study. It plots reactivation of the original in the hippocampus against prediction error size separately for both prior accuracy (A) and prior precision (B). Additionally, the outcomes of PE sizes are shown at the top of the figure. **A.** Prior accuracy refers to how similar original and mismatch targets are and was measured via difference ratings from a separate sample. Small PEs, evoked by similar mismatch targets, led to high representational similarity to the original. However, also very different targets did so, pointing to a role of pattern completion to reinstate the original target alongside the incoming mismatch target. Highly similar targets (low PEs) led to high mismatch recognition memory, but low source memory. Medium similarity (medium PEs) was associated with low mismatch recognition memory. Large PEs led to high mismatch target recognition memory and high mismatch source memory. Additionally, original representations were more stable in the comparison before vs after the PE. Note that prior accuracy had no effect on original target recognition memory. **B.** Prior precision refers to how strong a prediction was before the mismatching input. It was measured as the reinstatement of the original target while the head was playing before the mismatching target. Higher prior precision, so more reinstatement of the predicted original target in the IFG, was associated with subsequently higher reinstatement in the hippocampus as well. Weaker predictions (i.e. smaller PEs) were associated with lower original and mismatch target recognition, and lower original representational stability. After stronger predictions (i.e., larger PEs) all these measures were higher. Prior precision had no effect on source memory. Together, the findings suggest distinct encoding of mismatching information after large PEs, enabling higher recognition and source memory, as well as original representations undergoing less change.

as the second factor contributing to PE size (Henson and Gagnepain, 2010) was measured continuously and neurally. Supporting previous research consistently implicating the inferior frontal gyrus (IFG) in the processing of PEs based on episodic memories and in other domains (El-Sourani et al., 2020; Jainta et al., 2024; Liedtke et al., 2025; Schliephake et al., 2021; Wurm and Schubotz, 2012), only the IFG robustly reinstated the original targets above baseline, whereas neither the whole brain nor the hippocampus did so. Our findings therefore extend previous work by suggesting that the IFG may contribute to higher-order, context-dependent predictions, particularly under structured input conditions (Fujitani et al., 2024), as a precondition to detecting mismatching input (Sherman et al., 2016), also when PEs are based on episodic memories.

In this study, prior precision measured in the IFG was linearly associated with both original and mismatch target memory. At first glance, this is inconsistent with the Latent Cause Theory (Gershman et al., 2017), which predicts a U-shaped relationship. Interestingly, while the current study, as well as Stawarczyk et al. (2020), found a linear increase in memory with increasing prior precision (i.e., larger PEs), G. Kim et al. (2014), with the same measure, found a decrease. However, these authors argue that their stronger PEs corresponded to moderate reinstatement, which makes the memory prone to weakening. While it is difficult to make statements about the absolute level of reinstatement in either study, the disparate results could be explained with overall differences in prediction strength due to implicit learning of random associations in the study of G. Kim et al. (2014) and socially charged, neatly separated, and twice-encoded dialogues here. Thus, G. Kim et al. (2014) might have captured low to medium levels of prior precision, whereas this experiment and Stawarczyk et al. (2020), also using naturalistic events, might have tapped medium to strong levels, together forming a U-shape of learning from PEs. This possible explanation should be tested in future studies.

In addition to testing how both factors separately contributed to recognition memory, we also tested whether they jointly contribute to mismatch target learning (this was not tested for the original target, as prior accuracy did not gain significance there). In a previous behavioural study, prior precision and prior accuracy interacted to predict original memory, with lower recognition for medium PEs, while for mismatch target learning, only prior accuracy was relevant (Boeltzig et al., 2025). In the current study, there was instead an interaction between prior precision in the IFG and prior accuracy for the new version. Specifically, the data suggested that prior precision was especially influential when prior accuracy was high, suggesting a compensatory relationship between these two determinants of PE size. Crucially, the measure of prior precision in this study was brain-derived and continuous, while it was previously manipulated via encoding frequency. The more graded measure may therefore have been capable of teasing out the interaction for mismatch target. However, differences in encoding frequency and the repetition of the originals after mismatch versions in this study may have prevented the significant interaction for original memory that had been found in the previous study (Boeltzig et al., 2025). Future studies should therefore systematically investigate the role of these factors for the influence of prior precision and prior accuracy on memory for the original and mismatch target.

Moving from behavioural to neural outcomes of PEs, we also tested how original representational stability was affected by PEs of different sizes. This was measured via representational similarity between the original targets in Session 1 and Session 3, so before and after the PEs. This stability linearly increased with larger PEs, pointing to more unchanged original memory traces after larger PEs. Note however that these effects were relatively subtle.

After establishing that large PEs were associated with increased original and mismatch target recognition, mismatch source memory, and higher original representational stability, we tested for mechanistic implementations of these effects. Specifically, we measured the similarity between original and mismatch targets to gauge how strongly the original is being reinstated while the mismatching input is processed.

For prior accuracy, we observed a U-shaped pattern, where the mismatching target was represented more similarly to the original target when the rated differences between them were either small or large. High similarity after small differences can be assumed to derive from the targets being perceptually similar and thus producing similar representations. However, after large differences, there was also higher representational similarity, which points to a role of reinstatement of the original target in the face of mismatching new input.

This higher hippocampal reinstatement of the original target when strongly mismatching input was detected points to a pattern completion process. This refers to the retrieval of other elements of an episode when cued with a single element. It has been shown to take place in the neocortex but also in the hippocampus (Horner et al., 2015; Joensen et al., 2024; Johnson et al., 2009), where it may particularly be driven by subfield CA3 (Grande et al., 2019). Pattern completion is also evident in dynamic naturalistic stimuli (Sun et al., 2025) and could therefore play a role in the processing of PEs. Specifically, large differences (i.e., low prior accuracy) may trigger enhanced pattern reinstatement, while medium differences may fail to do so.

Importantly, this prior accuracy-driven U-shaped reinstatement is independent from the linear effect of prior precision on hippocampal similarity. Prior precision in the IFG reflects a prediction preceding the mismatch with a linear relationship, where stronger predictions (i.e., stronger reinstatement of the original target) led to stronger reinstatement in the hippocampus during the mismatch. Therefore, for both prior accuracy and prior precision, larger PEs led to stronger or more sustained activation of the original memory while the mismatch was unfolding, resulting in their co-activation. Note however that even though both effects were significant, they explained only a small share of variance, highlighting the role of other influences in addition to the subtle effect of prediction errors.

These findings suggest that both prior precision and prior accuracy shape original reinstatement during mismatching input as a two-stage process: Prior precision precedes the mismatch and therefore is blind to any perceptual similarity between prediction and input. It therefore reinstates the original in anticipation of a continuation. Prior accuracy may then act as an additional retrieval cue – especially when differences are either small or large - possibly via an automatic pattern completion process that facilitates direct comparison between the prediction and the new input. This proposed mechanistic account of episodic PEs via temporally distinct triggers of pattern completion warrants further investigation. Especially the use of EEG could clarify the temporal dynamics of original target reinstatement and test whether prior precision and prior accuracy make distinct and independent contributions to this process.

The increased reinstatement of the original during mismatching input was positively related to original and mismatch target recognition. This is consistent with previous findings of original reactivation during mismatching but related new input reducing interference between similar events and thus strengthening original and mismatch target without creating a trade-off (Chanales et al., 2019; Kuhl et al., 2010). Strong original predictions and a clear mismatch with new input seem to signal that the original should be preserved without interfering with the encoding of the new episode.

The result that there is no trade-off, and therefore little or no interference between the two versions of the same event raises the question of how this is implemented representationally. The Latent Cause Theory (Gershman et al., 2017) predicts that after large PEs, the new event will be encoded separately, while after medium PEs, new information will be integrated into the old memory trace. Both integration (Greve et al., 2018; Stawarczyk et al., 2020; Wahlheim and Zacks, 2019, 2025) and distinct encoding or pattern separation (Bein et al., 2021; Frank et al., 2020; G. Kim et al., 2017) have been suggested to take place after PEs.

Importantly, the size of the PE has rarely been explicitly considered as a determining factor in these accounts. The data presented here supports the Latent Cause Theory (Gershman et al., 2017) and suggests a differentiated and continuous view of consequences of episodic PE size on representational outcomes. Instead of integration or distinct encoding as a universal outcome after all PEs, both processes can play out, depending on the size of the PE (Bein et al., 2023). After large PEs, we found high recognition memory for original and mismatch targets, high source memory for the mismatch target, and high original representational stability, indicating less change in the memory representations compared to smaller PEs.

These results are interesting against the backdrop of how the mismatching events are stored in memory. The Latent Cause Theory (Gershman et al., 2017) predicts that new information is integrated into pre-existing models after moderate PEs but stored separately after large PEs. In contrast, the model by Wahlheim and Zacks (2025) posits integration to be taking place after PEs generally. In the wider literature, memory integration is regarded as a process where similar events are stored in overlapping memory traces (Chanales et al., 2019). This allows the formation of new connections between events (Schlichting and Preston, 2015; Zeithamova and Preston, 2010) and generalisation over individual episodes (Bowman and Zeithamova, 2018; Mack et al., 2018). This may be to the detriment of unique features of each episode, such as source or detail memory (Carpenter and Schacter, 2017; but see Boeltzig et al., 2023 and de Araujo Sanchez and Zeithamova, 2023). However, similarity between events can also be resolved by pattern separation, where episodes are pushed apart representationally, preserving their unique features (Brunec et al., 2020; Kumaran et al., 2016; Zeithamova and Bowman, 2020).

After large PEs, we observed more stable (i.e., unchanged) original memory representations and high source memory, both of which are incompatible with the standard view of integration laid out above (Carpenter and Schacter, 2017; Gershman et al., 2017). Instead, it is more plausible that original and mismatch targets were distinctly encoded (Bein et al., 2021; Frank et al., 2020). As discussed above, we found that larger PEs lead to stronger original target reinstatement, which can support distinct encoding (Kuhl et al., 2010), providing a possible mechanism for the observed pattern of results. As we did not measure the mismatch target representation before PE induction, we cannot draw conclusions about a reduction of similarity, which would be the consequence of pattern separation. We therefore use the term of distinct encoding coined in previous literature (Kuhl et al., 2010).

After medium PEs, in contrast, recognition memory and source memory were lower, and original stability was reduced. While the latter two factors are often used as markers of integration, a direct behavioural or neural measure for integration is lacking in this study. However, integration is compatible with the results, as new evidence is integrated into the original trace (Gershman et al., 2017).

The results therefore support the view that seemingly disparate accounts of integration and distinct encoding can be reconciled by considering PE size as a critical factor in both theoretical and empirical approaches. As laid out, strong PEs were associated with increased source memory and more stable original representations. It is therefore unlikely that large PEs promoted integration in the standard view, involving neural changes to the original memory trace and a loss of episodic detail that facilitates generalization (Brunec et al., 2020; Carpenter and Schacter, 2017, 2018a, 2018b; Mack et al., 2016; Varga et al., 2019; Zeithamova and Bowman, 2020). However, in the framework of Wahlheim & Zacks (2025), integration refers to the ability to remember both versions of an episode, along with the PE itself and the temporal sequence of versions. This definition does not hinge on specific assumptions about neural representations, nor does it make predictions regarding the role of PE size. Our data therefore is not inconsistent with this account in the context of large PEs. Like many of the studies informing the Wahlheim & Zacks framework (Hermann et al., 2021; Kemp et al., 2024; Stawarczyk et al., 2020; Wahlheim et al., 2021; Wahlheim and Zacks, 2019), the current study used highly naturalistic materials, making it plausible to assume that the PEs involved were relatively large. It is therefore an interesting question whether different notions of what integration entails can explain the disparity between the frameworks of Wahlheim & Zacks (2025) and Gershman et al. (2017).

In summary, the current set of results provides support for the model by Gershman et al. (2017) in the realm of episodic predictions (Boeltzig et al., 2025; Liedtke et al., 2025), extending the previous findings to neural representations and using continuous measures of both prior precision and prior accuracy. The reinstatement of the original target seems to play a crucial role in this process, protecting the original from being modified with new information. Resonating with this, Bein et al. (2023) included reinstatement of the original memory as one of the moderators between integration and separation, which this research strongly supports and extends by observing effects of original reinstatement due to both prior precision and prior accuracy.

These findings have implications for applied memory research. For eye witnesses or people exposed to fake news, it is highly relevant to preserve their original memory and prevent the integration of potentially false details (Granhag et al., 2012; Hope and Gabbert, 2019). The higher propensity of memory updating after medium PEs may also be leveraged for fake news correction. Further work with relevant material (Kemp et al., 2022; 2024) could further explore whether it is possible to reliably create situations favouring medium PEs to correct and update previously encoded fake news.

4.1. *Limitations & future research*

The naturalistic dialogues were used to create a situation of ongoing prediction during dynamically unfolding events. As a byproduct, predictions are likely transcending the boundaries of the epochs that we chose as, for instance, prediction of the ends may have mixed with processing of the targets. However, this predictive activity is relevant, as it is likely dialogue-specific, and therefore meaningful. Furthermore, this raises the interesting question of where the predictive horizon ends and how far in advance predictions are made under which circumstances (Brunec and Momennejad, 2022), which is an important task for further research, especially within dynamic and continuous stimuli.

Studies considering multiple determinants of the size of episodic PEs are rare, and a proposed third factor, namely precision of the new input (Greve et al., 2017; Henson and Gagnepain, 2010), has not been addressed here. Future research should therefore manipulate or measure all of these, for the purpose of creating a broad range of PE sizes and identifying potential unique contributions to original and mismatch target memory. Other factors orthogonal to PE size such as level (item- vs category-level; H. Kim et al., 2019), qualitative type (Liedtke et al., 2025; D. Varga et al., 2025), or outcome of the prediction (Pupillo et al., 2023) have further been shown to impact PE effects. Future work should combine these factors to promote a more nuanced understanding of outcomes following PEs, recognizing that not all PEs have the same effects on memory.

To increase empirical certainty concerning memory integration and pattern separation after different PE sizes, alternative behavioural tests should be considered. The focus in this study was on recognition memory, but future studies could also prioritise source memory, which was tied here to recognition decisions, as well as detail memory, both of which have been used as indicators of integration (Carpenter and Schacter, 2017; but see Boeltzig et al., 2023 and de Araujo Sanchez and Zeithamova, 2023). Integration can also be assessed by testing for indirect links across episodes (Preston et al., 2004). Additionally, rejection of similar lures has been used as an indicator of pattern separation (Bein et al., 2020; Frank et al., 2020) but is challenging with verbal material where the exact phrasing is often not encoded (Poppenk et al., 2008). The lures used in the current study were furthermore unplayed modifications of the original targets, thus being more similar to the original than the mismatch target. An ideal test would use similar lures to both versions.

5. Conclusion

The presented study offers three main insights. First, the size of the episodic PE, quantified by continuous measures of prior precision and prior accuracy, is crucial for memory outcomes: larger PEs are associated with more detailed memory. Second, increased original reinstatement via pattern completion during the mismatching input can be caused by higher prior precision, but also by lower prior accuracy when the new input is very different from the original. This increased original reinstatement is beneficial for memory and leads to stronger representational stability of the original. Third, the evidence is consistent with a distinct encoding of episodes after stronger PEs, and less distinct (and possible integrative) encoding after moderate PEs. While more work is needed to assess representational consequences of episodic PEs, this study with its unique continuous measures of prior precision and prior accuracy and measurement of both original and mismatch target recognition, underlines the importance of considering a continuous measure of PE size in future empirical and theoretical works concerning the effect of PEs on episodic memory.

More broadly, these findings shed light on how memories can change after encoding. Such modification can occur automatically, as episodic memories help predict and make sense of unfolding events. The current study suggests that memories remain stable when they explain the environment very badly or very well – and that change is most likely when the quality of prediction falls in between these poles. Prediction errors, in this sense, are a powerful mechanism for memory modification, helping us to keep knowledge and experiences updated as the environment around us changes.

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.

CRediT authorship contribution statement

Marius Boeltzig: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Nina Liedtke:** Writing – review & editing, Software, Resources, Project administration, Methodology, Data curation, Conceptualization. **Sophie Siestrup:** Writing – review & editing, Software, Formal analysis, Conceptualization. **Falko Mecklenbrauck:** Writing – review & editing, Software, Formal analysis. **Moritz F. Wurm:** Writing – review & editing, Software, Formal analysis. **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.

Acknowledgments

We thank Monika Mertens, Britt Hasslöver, Lena Puder, Leon Exeler, Kinan Elachkar, Michelle Bellstedt Juliane Diefenthal, Tabea Krause, Lana Steuernagel, and Leon Exeler for their work in study organization, testing, and stimulus construction. We also want to thank Michael Borgard for leading the stimulus production, and all speakers appearing in them. We thank the Biological Psychology Group of the University of Münster, the Lund Memory Lab, and Linn Petersdotter for their valuable feedback.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2025.121375.

Data availability

The data is publicly available, and the link is shared in the manuscript and data availability statement

References

Abel, M., Nickl, A.T., Reßle, A., Unger, C., Bäuml, K.-H.T., 2023. The role of sleep for memory consolidation: does sleep protect memories from retroactive interference? Psychon. Bull. Rev. 30 (6), 2296–2304. https://doi.org/10.3758/s13423-023-02264-8.

Anderson, M.C., Hulbert, J.C., 2020. Active forgetting: adaptation of memory by prefrontal control. Annu. Rev. Psychol. 72 (1), 1–36. https://doi.org/10.1146/annurev-psych-072720-094140.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 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. Neurosci. Biobehav. Rev. 153. https://doi.org/10.1016/j.neubiorev.2023.105368. Article 105368.

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

Bein, O., Reggev, N., Maril, A., 2020. Prior knowledge promotes hippocampal separation but cortical assimilation in the left inferior frontal gyrus. Nat. Commun. 11 (1), 4590. https://doi.org/10.1038/s41467-020-18364-1.

Boeltzig, M., Johansson, M., Bramão, I., 2023. Ingroup sources enhance associative inference. Commun. Psychol. 1. https://doi.org/10.1038/s44271-023-00043-8. Article 40.

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.

Bosch, S.E., Jehee, J.F.M., Fernández, G., Doeller, C.F., 2014. Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. J. Neurosci. 34 (22), 7493–7500. https://doi.org/10.1523/jneurosci.0805-14.2014.

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

Bramão, I., Jiang, J., Wagner, A.D., Johansson, M., 2022. Encoding contexts are incidentally reinstated during competitive retrieval and track the temporal dynamics of memory interference. Cereb. Cortex 32 (22), 5020–5035. https://doi.org/10.1093/cercor/bhab529.

Brod, G., Greve, A., Jolles, D., Theobald, M., Galeano-Keiner, E.M., 2022. Explicitly predicting outcomes enhances learning of expectancy-violating information. Psychon. Bull. Rev. 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. Learn Instr. 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. Front. Hum. Neurosci. 6. https://doi.org/10.3389/fnhum.2012.00147. Article 147.

Brunec, I.K., Momennejad, I., 2022. Predictive representations in hippocampal and prefrontal hierarchies. J. Neurosci. 42 (2), 299–312. https://doi.org/10.1523/jneurosci.1327-21.2021.

Brunec, I.K., Robin, J., Olsen, R.K., Moscovitch, M., Barense, M.D., 2020. Integration and differentiation of hippocampal memory traces. Neurosci. Biobehav. Rev. 118, 196–208. https://doi.org/10.1016/j.neubiorev.2020.07.024.

Bubic, A., von Cramon, D.Y., Schubotz, R.I., 2010. Prediction, cognition and the brain. Front. Hum. Neurosci. 4. https://doi.org/10.3389/fnhum.2010.00025. Article 25.

Carpenter, A.C., Schacter, D.L., 2017. Flexible retrieval: when true inferences produce false memories. J. Exp. Psychol.: Learn. Mem. Cogn. 43 (3), 335–349. https://doi.org/10.1037/xlm0000340.

Carpenter, A.C., Schacter, D.L., 2018a. False memories, false preferences: flexible retrieval mechanisms supporting successful inference bias novel decisions. J. Exp. Psychol.: Gen. 147 (7), 988–1004. https://doi.org/10.1037/xge0000391.

Carpenter, A.C., Schacter, D.L., 2018b. Flexible retrieval mechanisms supporting successful inference produce false memories in younger but not older adults. Psychol. Aging 33 (1), 134–143. https://doi.org/10.1037/pag0000210.

Chanales, A.J.H., Dudukovic, N.M., Richter, F.R., Kuhl, B.A., 2019. Interference between overlapping memories is predicted by neural states during learning. Nat. Commun. 10 (1). https://doi.org/10.1038/s41467-019-13377-x. Article 5363.

de Araujo Sanchez, M.A., Zeithamova, D., 2023. Generalization and false memory in acquired equivalence. Cognition 234. https://doi.org/10.1016/j.cognition.2023.105385. Article 105385.

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. Elsevier, pp. 509–525. https://doi.org/10.1016/b978-0-12-812028-6.00027-6. Vol. 28.

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

Fernández, R.S., Boccia, M.M., Pedreira, M.E., 2016. The fate of memory: reconsolidation and the case of prediction error. Neurosci. Biobehav. Rev. 68, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004.

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

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

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

Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D., 2007. Statistical Parametric Mapping: the Analysis of Functional Brain Images. Academic Press.

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

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. Cereb. Cortex 34 (2). https://doi.org/10.1093/cercor/bhad508. Article bhad508.

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

Grande, X., Berron, D., Horner, A.J., Bisby, J.A., Düzel, E., Burgess, N., 2019. Holistic recollection via pattern completion involves hippocampal subfield CA3. J. Neurosci. 39 (41), 8100–8111. https://doi.org/10.1523/jneurosci.0722-19.2019.

Granhag, P.A., Ask, K., Rebelius, A., Öhman, L., Giolla, E.M., 2012. 'I saw the man who killed Anna Lindh!' An archival study of witnesses' offender descriptions. Psychol. Crime Law 19 (10), 1–11. https://doi.org/10.1080/1068316x.2012.719620.

Greve, A., Abdulrahman, H., Henson, R.N., 2018. Neural differentiation of incorrectly predicted memories. Front. Hum. Neurosci. 12. https://doi.org/10.3389/fnhum.2018.00278. Article 278.

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

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

Hermann, M.M., Wahlheim, C.N., Alexander, T.R., Zacks, J.M., 2021. The role of prior-event retrieval in encoding changed event features. Mem. Cogn. 49 (7), 1387–1404. https://doi.org/10.3758/s13421-021-01173-2.

Hope, L., Gabbert, F., 2019. Memory at the sharp end: the costs of remembering with others in forensic contexts. Top. Cogn. Sci. 11 (4), 609–626. https://doi.org/10.1111/tops.12357.

Horner, A.J., Bisby, J.A., Bush, D., Lin, W.-J., Burgess, N., 2015. Evidence for holistic episodic recollection via hippocampal pattern completion. Nat. Commun. 6 (1). https://doi.org/10.1038/ncomms8462. Article 7462.

Huang, Y., Rao, R.P.N, 2011. Predictive coding. Wiley Interdiscip. Rev.: Cogn. Sci. 2 (5), 580–593. https://doi.org/10.1002/wcs.142.

Imundo, M.N., Pan, S.C., Bjork, E.L., Bjork, R.A., 2020. Where and how to learn: the interactive benefits of contextual variation, restudying, and retrieval practice for learning. Q. J. Exp. Psychol. 74 (3), 413–424. https://doi.org/10.1177/1747021820968483.

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

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

Joensen, B.H., Ashton, J.E., Berens, S.C., Gaskell, M.G., Horner, A.J., 2024. An enduring role for hippocampal pattern completion in addition to an emergent nonhippocampal contribution to holistic episodic retrieval after a 24 h delay. J. Neurosci. 44 (18), e1740232024. https://doi.org/10.1523/jneurosci.1740-23.2024.

Johnson, J.D., McDuff, S.G.R., Rugg, M.D., Norman, K.A., 2009. Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. Neuron 63 (5), 697–708. https://doi.org/10.1016/j.neuron.2009.08.011.

Kemp, P.L., Alexander, T.R., Wahlheim, C.N., 2022. Recalling fake news during real news corrections can impair or enhance memory updating: the role of recollection-based retrieval. Cogn. Res.: Princ. Implic. 7 (1), 85. https://doi.org/10.1186/s41235-022-00434-1.

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. Cogn. Res.: Princ. Implic. 9 (1). https://doi.org/10.1186/s41235-024-00546-w. Article 28.

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

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

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

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

Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2. https://doi.org/10.3389/neuro.06.004.2008. Article 4.

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

Kumaran, D., Hassabis, D., McClelland, J.L., 2016. What learning systems do intelligent agents need? Complementary learning systems theory updated. Trends Cogn. Sci. (Regul. Ed.) 20 (7), 512–534. https://doi.org/10.1016/j.tics.2016.05.004.

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

Loftus, E.F., 2005. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. Learn. Mem. 12 (4), 361–366. https://doi.org/10.1101/lm.94705.

Mack, M.L., Love, B.C., Preston, A.R., 2016. Dynamic updating of hippocampal object representations reflects new conceptual knowledge. Proc. Natl. Acad. Sci. 113 (46), 13203–13208. https://doi.org/10.1073/pnas.1614048113.

Mack, M.L., Love, B.C., Preston, A.R., 2018. Building concepts one episode at a time: the hippocampus and concept formation. Neurosci. Lett. 680, 31–38. https://doi.org/10.1016/j.neulet.2017.07.061.

Nadel, L., Hupbach, A., Gomez, R., Newman-Smith, K., 2012. Memory formation, consolidation and transformation. Neurosci. Biobehav. Rev. 36 (7), 1640–1645. https://doi.org/10.1016/j.neubiorev.2012.03.001.

Nolden, S., Turan, G., Güler, B., Günseli, E., 2024. Prediction error and event segmentation in episodic memory. Neurosci. Biobehav. Rev. 157. https://doi.org/10.1016/j.neubiorev.2024.105533. Article 105533.

Oosterhof, N.N., Connolly, A.C., Haxby, J.V., 2016. CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Front. Neuroinform. 10. https://doi.org/10.3389/fninf.2016.00027. Article 27.

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

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

Poldrack, R.A., Mumford, J.A., Nichols, T.E., 2024. Handbook of Functional MRI Data Analysis. Cambridge University Press.

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

Preston, A.R., Shrager, Y., Dudukovic, N.M., Gabrieli, J.D.E., 2004. Hippocampal contribution to the novel use of relational information in declarative memory. Hippocampus 14 (2), 148–152. https://doi.org/10.1002/hipo.20009.

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 Sci. Learn. 8. https://doi.org/10.1038/s41539-023-00166-x. Article 18.

Quent, J.A., Greve, A., Henson, R.N., 2022. Shape of U: the nonmonotonic relationship between object–location memory and expectedness. Psychol. Sci. 33 (12), 2084–2097. https://doi.org/10.1177/09567976221109134.

Quent, J.A., Henson, R.N., Greve, A., 2021. A predictive account of how novelty influences declarative memory. Neurobiol. Learn. Mem. 179. https://doi.org/10.1016/j.nlm.2021.107382. Article 107382.

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

Richter, F.R., Chanales, A.J.H., Kuhl, B.A., 2016. Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. NeuroImage 124, 323–335. https://doi.org/10.1016/j.neuroimage.2015.08.051.

Schlichting, M.L., Preston, A.R., 2015. Memory integration: neural mechanisms and implications for behavior. Curr. Opin. Behav. Sci. 1, 1–8. https://doi.org/10.1016/j.cobeha.2014.07.005.

Schlichting, M.L., Preston, A.R., 2016. Hippocampal–medial prefrontal circuit supports memory updating during learning and post-encoding rest. Neurobiol. Learn. Mem. 134, 91–106. https://doi.org/10.1016/j.nlm.2015.11.005.

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. https://doi.org/10.1016/j.neuroimage.2021.118028. Article 118028.

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

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

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

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

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

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

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

Sladky, R., Friston, K.J., Tröstl, J., Cunnington, R., Moser, E., Windischberger, C., 2011. Slice-timing effects and their correction in functional MRI. NeuroImage 58 (2), 588–594. https://doi.org/10.1016/j.neuroimage.2011.06.078.

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. Proc. Natl. Acad. Sci. 117 (47), 29346–29353. https://doi.org/10.1073/pnas.1918063117.

Sun, L., Li, S., Ren, P., Liu, Q., Li, Z., Liang, X., 2025. Pattern separation and pattern completion within the hippocampal circuit during naturalistic stimuli. Hum. Brain Mapp. 46 (2). https://doi.org/10.1002/hbm.70150. Article e70150.

Thakral, P.P., Wang, T.H., Rugg, M.D., 2015. Cortical reinstatement and the confidence and accuracy of source memory. NeuroImage 109, 118–129. https://doi.org/10.1016/j.neuroimage.2015.01.003.

Varga, D., Raykov, P., Jefferies, E., Ben-Yakov, A., Bird, C., 2025. Hippocampus responds to mismatches with predictions based on episodic memories but not generalised knowledge. BioRxiv. https://doi.org/10.1101/2025.02.04.636427.

Varga, N.L., Gaugler, T., Talarico, J., 2019. Are mnemonic failures and benefits two sides of the same coin?: investigating the real-world consequences of individual differences in memory integration. Mem. Cogn. 47 (3), 496–510. https://doi.org/10.3758/s13421-018-0887-4.

Vlasceanu, M., Drach, R., Coman, A., 2018. Suppressing my memories by listening to yours: the effect of socially triggered context-based prediction error on memory. Psychon. Bull. Rev. 25 (6), 2373–2379. https://doi.org/10.3758/s13423-018-1481-2.

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

Wahlheim, C.N., Smith, W.G., Delaney, P.F., 2019. Reminders can enhance or impair episodic memory updating: a memory-for-change perspective. Memory 27 (6), 849–867. https://doi.org/10.1080/09658211.2019.1582677.

Wahlheim, C.N., Zacks, J.M., 2019. Memory guides the processing of event changes for older and younger adults. J. Exp. Psychol.: Gen. 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 Cogn. Sci. (Regul. Ed.) 29 (4), 380–392. https://doi.org/10.1016/j.tics.2024.11.008.

Weaverdyck, M.E., Lieberman, M.D., Parkinson, C., 2020. Multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. Soc. Cogn. Affect. Neurosci. 15 (4), 487–509. https://doi.org/10.1093/scan/nsaa057.

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.

Zeithamova, D., Bowman, C.R., 2020. Generalization and the hippocampus: more than one story? Neurobiol. Learn. Mem. 175. https://doi.org/10.1016/j.nlm.2020.107317. Article 107317.

Zeithamova, D., Preston, A.R., 2010. Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. J. Neurosci. 30 (44), 14676–14684. https://doi.org/10.1523/jneurosci.3250-10.2010.