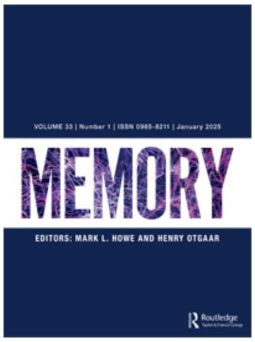

Memory

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pmem20

Prediction errors lead to updating of memories for conversations

Marius Boeltzig, Nina Liedtke & Ricarda I. Schubotz

To cite this article: Marius Boeltzig, Nina Liedtke & Ricarda I. Schubotz (2025) Prediction errors lead to updating of memories for conversations, Memory, 33:1, 73-83, DOI: 10.1080/09658211.2024.2404498

To link to this article: https://doi.org/10.1080/09658211.2024.2404498

MEMORY
2025, VOL. 33, NO. 1, 73–83
https://doi.org/10.1080/09658211.2024.2404498

Prediction errors lead to updating of memories for conversations

Marius Boeltzig, Nina Liedtke and Ricarda I. Schubotz

Department of Psychology, University of Münster, Münster, Germany; Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany

ABSTRACT
Previous research has established that the brain uses episodic memories to make continuous predictions about the world and that prediction errors, so the mismatch between generated predictions and reality, can lead to memory updating. However, it remains unclear whether prediction errors can stimulate updating in memories for naturalistic conversations. Participants encoded naturalistic dialogues, which were later presented in a modified form. We found that larger modifications were associated with increased learning of the modified statement. Moreover, memory for the original version of the statement was weakened after medium-strong prediction errors, which resulted from the interplay of modification extent and strength of previous memory. After strong prediction errors, both original and modification were well-remembered. Prediction errors thus play a role in keeping representations of statements and therefore socially relevant knowledge about others up to date.

ARTICLE HISTORY
Received 26 February 2024
Accepted 9 September 2024

KEYWORDS
Episodic memory; prediction error; memory updating; memory for conversations

Our daily lives are marked by social interactions and representing social information is a crucial cognitive skill (Oliva, 2022; Son et al., 2021). We gather information about others from several sources (Anzellotti & Young, 2020) and this knowledge can guide the decision of whether to trust them (Kroneisen & Bell, 2022). However, little is known about how information about others that is deduced from socially relevant episodic memories is kept up to date. Generally, updating of memory and knowledge is thought to be stimulated by so-called prediction errors (PEs). According to the predictive coding framework, PEs occur when the expectations, continuously and automatically generated by the brain on the basis of memory and knowledge, do not match the actual experience. As a result, an update is made to optimise future predictions (Friston & Kiebel, 2009; Quent et al., 2021).

The current study set out to test whether and how PEs can alter memory for statements in naturalistic and socially relevant conversations. PEs contribute to event segmentation of verbally presented stories (Kumar et al., 2023), but their role in updating memories of conversations remains unaddressed. Strikingly, previous studies found memory for statements to be relatively poor (Brown-Schmidt & Benjamin, 2018), especially memory for the surface form, so the exact phrasing of statements (Pezdek & Prull, 1993; Poppenk et al., 2008). Nevertheless, conversations constitute a major source for social learning (Atir et al., 2022) and the content and style of statements can contain relevant information about others (e.g., Rosenblum et al., 2020). Conversations and the PEs they contain may therefore be a driver of updating what we know and remember about others.

To test whether PEs play a role in the updating of memories for conversations, we investigated three central aspects against the backdrop of previous research.

First, the updating of memories by PE can consist of the old memory fading or a new memory being encoded. On the one hand, PEs were found to lead to a weakening (Forcato et al., 2007) or pruning (Kim et al., 2014, 2017; Vlasceanu et al., 2018) of the original memory that was used for the incorrect prediction. On the other hand, PEs induce learning of the information that was not predicted (e.g., Bein et al., 2021; Greve et al., 2017; Wahlheim & Zacks, 2019), which can also lead to the whole episode or details from it later being mistaken for the original memory (Jainta et al., 2022; Shao et al., 2023; Siestrup et al., 2022; Siestrup & Schubotz, 2023; Sinclair & Barense, 2018). Interestingly, previous studies usually examined these two effects separately, making claims about the relationship between the two potential memory traces difficult. We therefore investigated both memory for the original and the modification in the same study for the first time.

CONTACT Marius Boeltzig marius.boeltzig@uni-muenster.de Department of Psychology, Fliednerstraße 21, 48149 Münster, Germany
Supplemental data for this article can be accessed online at https://doi.org/10.1080/09658211.2024.2404498.

Second, there is uncertainty about the relationship between PE strength and the likelihood of learning from PE. According to the Predictive, Interactive Multiple Memory Systems framework (PIMMS; Henson & Gagnepain, 2010), PE size is affected by prior accuracy, referring to the difference between prediction and actual input (i.e., how much the stimulus changed), and prior precision, denoting the familiarity with the situation and specificity of predictions (i.e., how exact predictions are). Consequently, the biggest PEs will be achieved when strong expectations (high prior precision) turn out to be very wrong (low prior accuracy).

Bigger PEs, evoked by higher prior precision and lower prior accuracy, are associated with higher levels of new learning of the unexpected information (Greve et al., 2017). When it comes to the impact on the original memory used for the prediction, however, the relationship between PE strength and memory accuracy has been suggested to be U-shaped. According to the Latent Cause Theory (Gershman et al., 2017), small PEs carry no substantially new information and therefore have little impact on the original. Medium PEs are big enough to be considered important in the future and therefore lead to an updating of the original episode, which is changed and adapted so that it now reflects the changed state of the environment – to the detriment of details from the old and now outdated episode. Big PEs, finally, may be produced by a situation where the environment has changed so much that a new underlying latent cause is inferred. Consequently, a new memory trace is needed to reflect the new latent cause, leaving the original unaltered (Gershman et al., 2017). Similarly to this account, the Nonmonotonic Plasticity Hypothesis states that memories are rendered vulnerable to distortion specifically when moderately reactivated (Kim et al., 2014). As prediction is implemented through reinstatement, this account would also predict original weakening after medium PEs.

Previous studies investigating original weakening have not specifically employed the PIMMS framework, leaving it unclear whether the interplay of prior accuracy and prior precision can yield these moderate PEs where the original is adapted and therefore weakened in its fidelity. Furthermore, prior accuracy has usually been manipulated in a categorical manner (see Rouhani et al., 2018 for a quantitative manipulation using reward learning), making claims about the full relation between PE size and memory outcomes difficult. We therefore aimed to observe the continuous effect of PE strength, achieved by combining prior accuracy and prior precision, on original weakening and new learning.

Third, in addition to prior accuracy and precision, the type of modification may be crucial (Siestrup & Schubotz, 2023). Changes to the surface form (i.e., the phrasing of the statement) may not have mnemonic consequences, as they introduce no crucial new information like gist changes altering the meaning of the statement do. Additionally, if the phrasing of statements is poorly remembered (Pezdek & Prull, 1993; Poppenk et al., 2008), there would be less precise predictions and smaller PEs (Reichardt et al., 2020).

To examine these three aspects, participants were presented with naturalistic dialogues, which were later played with a modification inducing a PE. PE size was manipulated by prior accuracy, assessed on a continuous scale of the amount of change between the original and modified statement, and prior precision, induced by different encoding frequencies of the original. PE quality was manipulated by using two different modification types, namely surface and gist. A recognition test was used to assess memory both for the original information that was used for the prediction and for the new information brought forward in the modification.

We expected that lower prior accuracy and higher prior precision as proxies for PE size would lead to new learning. Original weakening was expected after mid-sized PEs. Furthermore, we expected gist modifications to have a stronger influence on memories than surface changes.

Methods

Participants

Based on sample sizes of previous studies investigating PEs with naturalistic stimuli (Jainta et al., 2022; Siestrup et al., 2022, 2023; Siestrup & Schubotz, 2023; Sinclair & Barense, 2018), 44 participants were recruited. A power analysis confirmed that a sample size of 34 was sufficient to detect a medium effect ($\beta = .40$) in a linear regression with 80% power. After drop-out of four participants who did not attend all five experimental sessions, the final sample of $N = 40$ consisted of 28 women, 11 men and one non-binary person with a mean age of 23.40 years (SD = 2.96, range: 18-31). Participants, most of whom were students, received course credit or financial compensation. All participants provided informed consent after reading information on the procedure of the study, their right to leave the experiment at any point, and the possibility to have their data deleted after participation. The procedure was approved by the ethics committee of the Faculty of Psychology at the University of Münster. The study was not preregistered.

Material

The authors wrote 80 dialogues in the German language, 36 of which were used for this experiment. Each dialogue included two speakers discussing topics centred around the daily life of young adults and allowed forming impressions of the characters by judging their interests and communication with and about others. Anecdotal observations confirmed that participants had favourite and disliked characters and formed a wealth of opinions on their behaviour, confirming the social nature of the dialogues. The dialogues were recorded by 20 professional voice actors and mixed together with a unique background sound matching the setting of the dialogue. The resulting dialogues were between 21 and 34 seconds long (M = 27.31, *SD* = 3.02).

Five versions were constructed for each dialogue, one original and four modifications (see Figure 1a). The versions shared a common head and end, and only differed in one statement in the middle of the dialogue, the target. In comparison to the original version, two versions of the target deviated on a surface level either to a low degree (e.g., use of a synonym) or a high degree (e.g., a larger-scale paraphrase) while preserving the content. The other two versions were low or high gist changes, where the meaning of the utterance was changed slightly or significantly, respectively (see Table 1 for an example). The ends were constructed so that they integrated well with the original and all modifications.

Design

Three experimental factors were manipulated and hypothesised to influence the size and quality of the prediction error, and therefore memory outcomes. First, prior accuracy was operationalised as the amount of difference between original and modified version of the dialogue and was measured through individual participant ratings at the end of the experiment. Second, prior precision was manipulated through encoding frequency of the original version with the dialogues being presented three or five times. Third, qualitatively different PE types were created by modifications containing either a change in language that preserved the meaning (surface change), or a change in content (gist change).

To establish episodic memories, participants first encoded original versions of the dialogues. After that, without warning, two thirds of the dialogues were played in a version containing a modified target statement. Specifically, the dialogue began just as it had before, but would eventually continue in a changed and unpredicted way, therefore eliciting a PE. Participants were later tested for their memory of both original and modified version.

To assess the overall effect of PEs, not all dialogues were changed (see Supplementary Figure 1.1). Only the 24 changing dialogues were first encoded in the original

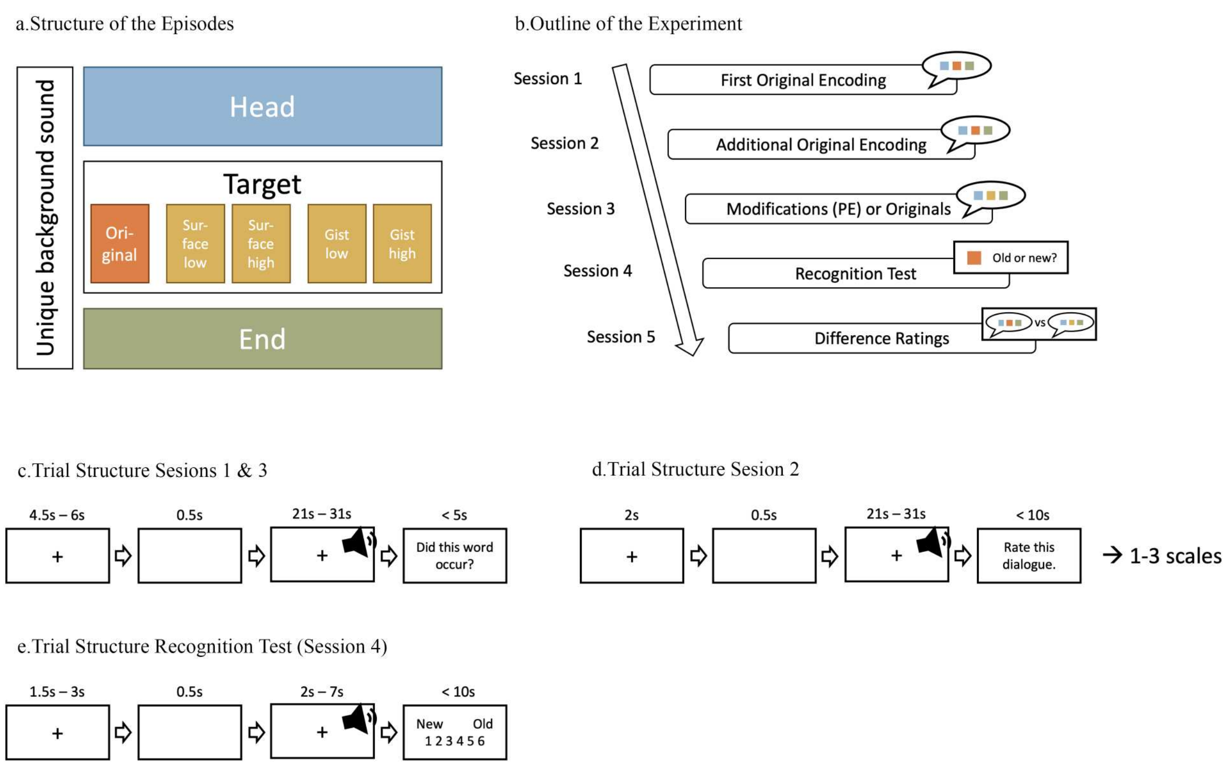

Figure 1. Method of the experiment. (a) The stimuli were dialogues covertly consisting of heads, targets, and ends together with unique background sounds. All dialogues were first presented with the original target and some dialogues were later modified by using one of the four alternative targets, which were smaller or bigger changes on a surface or gist level. (b) The experiment was conducted on five consecutive days. In Sessions 1 & 2, the dialogues were encoded and re-encoded. In Session 3, some of the dialogues were modified in one of the described forms, while the others were repeated in the original. In Session 4, there was a recognition test where, among others, both the original and the modified target were tested. In Session 5, participants rated the difference between the original and modified dialogues that they had heard to obtain an individual measure of prior accuracy. (c) In Sessions 1 and 3, there was a jittered fixation cross (4.5-6s) and a blank screen (0.5s) before the dialogue was played. Participants were then asked to determine whether a given word occurred in the dialogue or not. (d) In Session 2, after a fixation cross (2s) and blank screen, participants listened to the dialogue and then rated it on 1–3 of five scales that assessed individual perceptions of the dialogues. (e) In Session 4, a recognition test was conducted. After a jittered fixation cross (1.5-3s) and a blank screen, participants listened to an excerpt of the dialogue (the targets or head) and then decided whether they had heard exactly that statement before. They used a scale from 1 = definitely new to 6 = definitely old.

Table 1. Translated example dialogue.

	The dialogue plays at a pub, there is a talking crowd and music in the background.
Head	A: Cheers! Here's to winning! B: Cheers! What a game. It hasn't been this much fun for a while. A: Yes, and the atmosphere was amazing. Great support from everyone for 90 minutes. B: Do you think we're safe from relegation now?
Target	**O:** A: After this performance, I have no doubts that we are. **SL:** A: After this match, I have no doubts that we are. **SH:** A: They played so well, so we are safe for sure. **GL:** A: Well, even though the performance was good, I still have some doubts. **GH:** A: No, definitely not. It will still be tough all the way till the last game.
End	B: Well, today, we will celebrate. Great start to the week.

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

version and later presented in one of the four modifications (with a quarter of the dialogues in each modification category). In addition, there were six repeated unchanging dialogues which were never altered and always played in the original version. Furthermore, in Session 3, six new and previously unheard unchanging dialogues were presented once in their original versions.

Procedure

The experiment was conducted on five consecutive days (see Figure 1b). It was implemented in PsychoPy (Peirce et al., 2019) and run online on Pavlovia (pavlovia.org). As the experiment was conducted during a term break, participants could choose to do the experiment via Zoom ($n = 35$) or in the lab ($n = 5$). In both cases, each session was conducted by an experimenter who ensured a quiet environment, fixed technical issues, and provided clarification on the instructions. Participants listened to the stimuli through headphones and self-regulated the volume at the beginning of each session.

Participants were told that we were interested in perceptions of everyday conversations and that they should listen to the dialogues as if witnessing them directly. They were also instructed to imagine the setting visually. The assignment of dialogues to modification and encoding frequency conditions was counterbalanced across participants by randomly assigning dialogues to lists and rotating these lists through the experimental conditions. The stimulus orders were randomised in each session. Each session had one to three self-paced breaks, according to session length. Presentation times and trial structures are displayed in Figure 1c–e.

In Sessions 1 and 2, the original versions of the six repeated unchanging and 24 changing dialogues were established. In Session 1, each dialogue was presented only once. After each dialogue presentation (see Figure 1c), there was an attention check in which participants were cued with a word and were asked to decide whether it was used in the dialogue. For half the dialogues, the word actually did occur (pseudo-randomly chosen from the head, target, or end), while non-occurring but plausible words were presented for the other half.

In Session 2, while still establishing the original versions of the memories (see Figure 1d), prior precision was manipulated through encoding frequency as half the dialogues were played two or four times, respectively. To space out repetitions of the same dialogue, the twice-encoded dialogues were either played in block one and three or two and four, and the other dialogues were played once in each block. Instead of attention checks, each dialogue was followed by rating scales (concerning autobiographical associations, valence, arousal, typicality, social norms) to capture individual impressions of the dialogue. Depending on how often an individual dialogue was presented, one to three scales were administered after each presentation. Responses on these scales were not analysed for this paper.

Crucially, in Session 3, we aimed to elicit PEs by playing the 24 changing dialogues in a modified version (one quarter in each of the versions shown in Figure 1a). The now-familiar dialogues would start in the same way as before, given that the head was identical in all versions, but would then unexpectedly continue in a different way, as the target was modified. Participants were not warned about these changes and were told that this session would be just like Session 1, with the only difference that they would hear some completely new dialogues (i.e., the six singular unchanging dialogues). Each dialogue was again followed by an attention check (see Figure 1c), for which a non-occurring word was used when an occurring one was tested in Session 1, and vice-versa.

In Session 4, participants were tested on their memory in a surprise recognition test (see Figure 1e). Participants were instructed that they would hear short excerpts (2.4–6.9s, $M = 4.55$) from the dialogues. These excerpts usually constituted the full target statement, but to reduce variability in duration between the excerpts, longer targets were shortened without cutting off the speakers abruptly.

Participants were asked to indicate whether they had heard that exact statement before, responding on a scale from 1 = definitely new to 6 = definitely old (Brady et al., 2023). They were instructed that it did not matter when they heard that statement, so that modifications from Session 3 would also be accepted as old. Furthermore, they were advised that some of these probes would be similar to ones they had heard before and were asked to assess whether the statements were phrased exactly the same way, not taking into account previous decisions in regard to similar statements. To space out similar probes, the task was covertly divided into five blocks where no more than one probe referring to the same dialogue would appear in each block. The order of probe types for each dialogue (e.g., whether a dialogue was tested on original or modification memory first) was counter-balanced across dialogues and participants and controlled for in the analyses if necessary.

A total of 192 probes was presented. Each changing dialogue was tested using five different probes, including excerpts from the original target, which was presented in the first two sessions, the modified target which was heard in Session 3, and the head, which was always unchanged. There were also two previously unheard targets from other modifications, which acted as similar lures. If the Surface Low modification was presented in Session 3, these lures were for instance Surface High and Gist high. The assignment of the two lures to each dialogue was counterbalanced. Performance on lures was not further analysed, instead, their function was to introduce the need for careful memory calibration. The unchanging dialogues that were played in the same form in Sessions 1–3 had the same probe types, except for the modification target, as those dialogues were never modified. Unchanging dialogues that were singularly played in Session 3 were tested only with a probe from the head and from the original target.

As perfect responding to this collection of probes would produce more old than new responses, we also added 36 novels, which were statements from dialogues that were not used in this experiment. These novels were distributed across the blocks so that each block would feature the same number of old and new probes (see Supplementary Figure 1.1).

Finally, in Session 5, to get a measure of prior accuracy, participants were asked to rate the difference between the two dialogue versions that they had heard before. Each of the 24 changing dialogues was first presented in its modified form (as heard in Session 3) and then in the original (from Sessions 1 and 2). Participants rated the magnitude of difference on a scale from 1 = very small difference to 7 = large difference. After this session, they were debriefed on the purpose of the experiment, thanked, and compensated.

Data analysis

The data was analysed using R (R Core Team, 2023) and is available on OSF (https://osf.io/8dnsg/?view_only=3fc9fd3393474fd8b82969623c6ce6f9). Our main interest was the pattern of original and modification memory depending on prior accuracy (i.e., difference ratings), prior precision (i.e., encoding frequency), and change type (i.e., surface or gist). Due to the inclusion of this latter factor as well as potential qualitative differences between changing and unchanging dialogues, the key analysis was limited to changing dialogues. However, after first establishing overall memory levels for unchanging dialogues, we tested whether hearing a modification would impair memory for the original episode by comparing unchanging with changing dialogues on original target memory and subsequently on memory for the unchanged head.

For this and the following analyses, we chose to use weighted accuracy as a dependent variable. Participants cast their responses on a scale from 1 = definitely new to 6 = definitely old. However, as it was doubtful whether confidence for wrong responses would be meaningful, we instead set wrong answers to 0 and coded correct answers as 1–3 according to confidence judgements.

After thus assessing the general effect of modifications on original memory, the key analyses focused on the changing dialogues that were modified over the course of the experiment. Two separate models tested weighted accuracy for the original and the modified targets, representing memory for the original version from Sessions 1 and 2 and for the modification from Session 3. The use of weighted accuracy allowed us to fit linear mixed models in order to include individual participants as well as individual dialogues as random intercepts and to include potentially confounding nuisance variables in the models.

Even though we employed careful counterbalancing and randomisation, we opted to control for influences on memory that arose from the complex multi-session design. These nuisance variables were selected in a data-driven approach where we ran a separate regression with the experimental factors and each of the possible nuisance variables. To make sure we would only use the most important variables and avoid over-fitting, those variables that gained significance were additionally subjected to a stepwise backward regression. Only the surviving variables then entered the analyses. This procedure was done separately for original and modification memory. A full list of the tested nuisance variables can be found in Supplementary Note 2.

Participants varied considerably in their use of the difference rating scale. The maximum value of each participant on the seven-point scale ranged from four to seven and medians ranged from one to five. To address these differences in conceptions of smaller and bigger changes, we z-standardized the difference ratings for each participant. Bigger values on this scale therefore reflect bigger relative changes in the subjective perception of each participant.

The models were fitted using the *lme4* package (D. Bates et al., 2015) and the significance of the individual predictors was tested using the *car* package (Fox & Weisberg, 2019), following the approach of previous studies in the field (e.g., Calderon et al., 2021; Wahlheim et al., 2019). The *emmeans* package (Lenth, 2023) was used to carry out follow-up tests, specifically comparing slopes across different factor levels in interaction effects, using the *emtrends* function.

In the interest of brevity and readability, only significant results are fully qualified with test statistics and effect sizes. However, the full models can be consulted in Supplementary Note 3.

Results

Performance in attention checks in Sessions 1 and 3 was high (Session 1: M = .89, SD = .07, Session 3: M = .92, SD = .05) with even the lowest-performing participant responding above chance level. Participants can therefore be assumed to have paid attention to the dialogues. Furthermore, across dialogue conditions, old probes in Session 4 were recognised with above-chance performance (original targets: M = .83, SD = .08; modified targets: M = .64, SD = .16; heads: M = .82, SD = .10), *ps* < .001, *ds* > 0.84, and new probes were rejected with above-chance accuracy (lures: M = .78, SD = .07; novels: M = .99, SD = .01), *ps* < .001, *ds* > 3.91. Overall memory can therefore be assumed to be robust.

Memory for unchanging dialogues

To assess baseline memory for verbal statements, we turned to the unchanging dialogues, which were always presented in the same version. We collapsed across head and target probes, as they were not qualitatively different in these dialogues. Correct acceptance of those probes was significantly above chance (M = .76, SD = .14), $t(39) = 11.81$, $p < .001$, $d = 1.87$, 95% CI [1.10, 2.63]. Furthermore, unheard lures were rejected with above-chance performance (M = .80, SD = .13), $t(39) = 14.94$, $p < .001$, $d = 2.36$, 95% CI [1.53, 3.20]. Thus, there was robust above-chance memory for unchanging dialogues.

The repeated unchanging dialogues were encoded three or five times in Sessions 1 and 2 and then repeated in Session 3, while the singular unchanging dialogues were played only once during Session 3. Encoding frequency had a significant impact on heads and original targets, $F(1.92, 75.05) = 40.76$, $p < .001$, $\eta^2 = .32$. There were significant differences between one encoding (M = .64, SD = .21) and dialogues encoded more often (3 encodings: M = .87, SD = .15; 5 encodings: M = .90, SD = .13), *ps* < .001, *ds* > 1.11. There was no significant difference between three or five encoding opportunities, $p = .523$. Thus, memory was enhanced when dialogues were encoded multiple times.

Effects of modifications on original targets

To establish whether PEs induced an overall weakening of original targets, unchanging dialogues were compared to changing dialogues with surface or gist modifications. A 3 × 2 within-subject ANOVA was carried out, with the factors of modification type (unchanging, surface changing, gist changing) and encoding frequency (three times, five times) predicting weighted accuracy.

The ANOVA showed a significant effect of encoding frequency, $F(1, 39) = 5.87$, $p = .020$, $\eta^2 = .01$, with better memory for dialogues played five times (M = 2.28, SD = 0.52) compared to dialogues played three times (M = 2.16, SD = 0.55). There was also a significant effect of modification type, $F(1.84, 71.84) = 3.86$, $p = .029$, $\eta^2 = .03$, suggesting that modifying the dialogue had an impact on original memory. However, in the Tukey-corrected post-hoc tests, neither the comparison between unchanging (M = 2.35, SD = 0.67) and gist-changed dialogues (M = 2.13, SD = 0.46), $p = .056$, nor that between unchanging and surface-changed ones (M = 2.18, SD = 0.44), $p = .121$, gained significance. There was no significant interaction, $p = .603$ (see Supplementary Table 3.1 for full model).

Even though the individual comparisons did not gain significance, the main effect of modification type suggests that there is an overall weakening of the original after PEs, while gist and surface modifications seem to have a similar effect.

Effects of modifications on heads

The destabilisation of original targets through PEs could potentially have the side effect of deterioration of the heads, which were connected to the targets but not modified themselves. However, this was not the case. While encoding frequency was a significant predictor of head recognition, $F(1, 39) = 13.82$, $p < .001$, $\eta^2 = .04$ (three encodings: M = 2.06, SD = 0.53; five encodings: M = 2.26, SD = 0.55), the modification type, $p = .956$, and the interaction, $p = .857$, were not significant (see Supplementary Table 3.2). The potential weakening effect of PEs was therefore limited to the targets and did not spill over to the heads.

Effects of PE size and type on original targets

After establishing that modification of statements had an overall weakening effect on original targets, we tested whether the three experimental factors would influence this deterioration. Using only the changing dialogues, it was tested whether prior accuracy (i.e., different degrees of modifications as rated by participants), prior precision (i.e., three or five encodings), and modification type (i.e., surface, gist) would have an influence on original target recognition. This analysis therefore tested weighted accuracy for the original version of the episode depending on the kind of modification that was administered in Session 3.

In addition to nuisance variables (see Supplementary Table 3.3), only the interaction between the difference rating and encoding frequency reached significance, $\beta = .31$, $F(1, 913.11) = 4.50$, $p = .034$. As can be seen in Figure 2a, the slope is negative for three encodings, with higher difference ratings associated with lower memory, while it is positive for five encodings, with higher difference ratings leading to better memory. The fact that the slopes were significantly different was confirmed in a direct comparison ($p = .039$), while in separate models for three and five encodings, the difference rating did not gain significance (three encodings: $p = .259$; five encodings: $p = .182$).

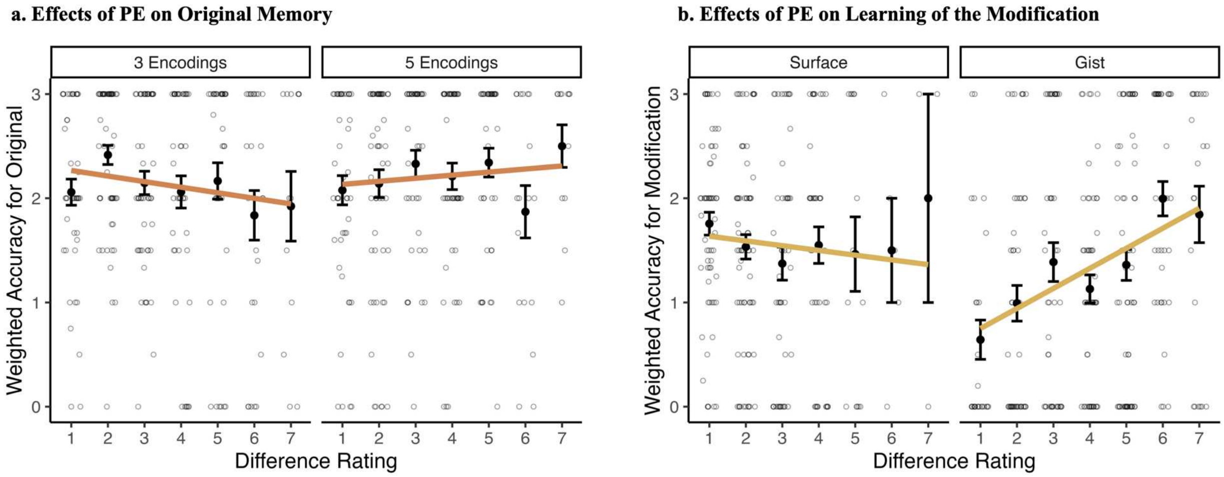

Figure 2. Memory outcomes for original and modification. (a) Weighted accuracy ratings for the original version of the statement (as encoded in Sessions 1 and 2) are plotted against difference rating (i.e., prior accuracy) and encoding frequency (i.e., prior precision). Bold dots represent means for each combination of factors and error bars are standard errors. The grey dots are individual participants' means for this combination of factors. (b) Weighted accuracy ratings for the modified version of the statement (as encoded in Session 3) are plotted against difference rating (i.e., prior accuracy) and modification type. There was a significant interaction between difference rating and modification type. Note that few surface modifications got high difference ratings, thus the large error bars.

In summary, the effect of the difference rating flipped depending on how well the original was encoded. While this interaction was significant, the effect of difference rating within each encoding frequency level was not.

Effects of PE size and type on modification recognition

Complementing the results on original target recognition, which reflects memory for the original version, we also assessed how new learning of the modification is influenced by the three experimental factors. In addition to nuisance variables (see Supplementary Table 3.4), the effect of modification type gained significance, $\beta = .33$, $F(1, 876.45) = 11.01$, $p < .001$, with overall higher memory after surface ($M = 1.57$, $SD = 1.23$) than gist changes ($M = 1.35$, $SD = 1.30$). Furthermore, the interaction between the difference rating and modification type gained significance, $\beta = .18$, $F(1, 917.55) = 9.69$, $p = .002$, while the overall effect of the difference rating was short of significance, $\beta = .10$, $p = .070$. A comparison of the slopes confirmed a significant difference between the modification types ($p = .001$), which can be seen in Figure 2b, where the slope is markedly flatter for surface modifications than gist modifications. In separate exploratory models for gist and surface, it was confirmed that the effect of difference ratings was only significant with the gist modifications, $\beta = .11$, $F(1, 384.54) = 16.13$, $p < .001$, but not the surface modifications, $p = .225$.

In summary, higher difference ratings led to better learning of gist modifications, while surface modifications did not benefit from larger differences. Furthermore, surface modifications were recognised better than gist modifications in general. Encoding frequency of the original (i.e., prior precision) was not of significance for new learning.

Discussion

Conversations are at the heart of our daily lives. Who said what to whom and how contributes substantially to our social knowledge and memory (Kumar et al., 2023). Here, we investigated whether memories of naturalistic and socially relevant conversations are subject to change in response to prediction errors (PEs) as has been shown for other memories. Our findings confirm this, indicating learning of the new information brought forward in the modified statement as well as some degree of weakening of the old version of the statement.

Our general main finding is that prediction errors contribute to learning when listening to verbal statements within inherently social material. This embodies a crucial extension of previous research that has used complex events, which were however neither verbal nor social (e.g., Siestrup et al., 2022; Sinclair & Barense, 2018). Information about others plays an important role in our lives and therefore needs to be represented in memory (Kroneisen & Bell, 2022; Oliva, 2022) and flexibly kept up to date. In the social domain, PEs have been shown to modulate the updating of overall person impressions (Mende-Siedlecki et al., 2013) and the perception of attractiveness (Mihara et al., 2023). Our findings connect this line of research with work on the social influences on flexible use of memories (Boeltzig et al., 2023) by showing that PEs are involved in flexibly pruning and extending our stock of socially relevant episodic memories. We also extend findings from a previous study on socially shared information as a trigger for memory pruning after PEs (Vlasceanu et al., 2018) by using a design where not only the

PE-inducing modifications were delivered socially, but also the initial encoding situation, and by adding new learning as a consequence of PEs in the social domain. Future work may therefore also establish whether PEs can drive social contagion, a process where false memories are introduced through conversations (Andrews & Rapp, 2014; Harris et al., 2017).

Independent of the social and verbal domain, our results extend insights into new learning after PEs. We found more new learning, pertaining to the new and unexpected parts of the episodes that evoked a PE, after lower prior accuracy (i.e., bigger modifications), which is in line with the PIMMS framework and previous research (e.g., Bein et al., 2021; Greve et al., 2017). However, by using the continuous measure of participant-generated difference ratings, we showed a linear relationship between prior accuracy and modification learning for the first time outside reward learning (Rouhani et al., 2018).

The effect of prior accuracy on new learning was specific to gist changes and not found for surface modifications. Interestingly, previous studies using visual narratives sometimes found effects of changes not influencing the overall gist (Siestrup & Schubotz, 2023; Sinclair & Barense, 2018). With regards to poor memory for statement phrasing, so the surface form of the episodes (Pezdek & Prull, 1993), and theories that gist and surface information may be encoded separately (Reyna et al., 2016), it is possible that no predictions about the exact wording were generated and could therefore not have been violated. The overall high performance on surface modifications may be explained by their high similarity to the original, which could have led to confusions between the two versions. We aim to test this possibility in the future by introducing a source memory test.

A final result concerning new learning after PEs worth highlighting is that it was not increased by higher prior precision. It is possible that the chosen encoding frequencies (three or five times) are already at the higher end of prior precision where they were relevant for original weakening, but not new learning. Future work could address this and manipulate prior precision continuously instead of utilising two levels. Furthermore, even though the approach of manipulating prior precision via encoding frequency has also been taken in the related field of fake news correction (Kemp et al., 2024) other manipulations of prior precision are conceivable, such as the inherent logic and predictability of the episode.

Turning from effects of new learning to effects on the original episode, which was used for the incorrect prediction, our study finds an overall detrimental effect of PEs on original memories. This extends results from previous studies (Kim et al., 2014, 2017; Vlasceanu et al., 2018) and reflects the need to weaken seemingly outdated information. However, pruning of particularly social information is an interesting finding as it may be relevant to represent outdated information about others to understand their pasts. For instance, forgetting the former spouse of your friend after their divorce would hardly be beneficial. More research is therefore needed to investigate the role of anticipated future usefulness of old information in memory pruning.

The overall detrimental effect of PEs on original memories was modulated by the interplay of prior accuracy and prior precision. This pattern of results is fully consistent with previous frameworks (see Figure 3 for a depiction of this). The prior is less precise after three encodings than after five encodings, as people are less familiar with the dialogues. According to the PIMMS framework (Henson & Gagnepain, 2010), when this imprecise prior is combined with high prior accuracy (i.e., a modification that is similar to the original), the PE is small (see left quarter of Figure 3). The Latent Cause Model (Gershman et al., 2017) would therefore predict that there is no substantial new information that needs to be integrated, and congruently, there is no weakening of the original in our data after these small PEs.

When prior accuracy decreases with bigger modifications (i.e., higher difference ratings), PE size is thought to increase (second quarter in Figure 3; Henson & Gagnepain, 2010). Similarly, with a more precise prior after five encodings, even small deviations should lead to a bigger PE (third quarter in Figure 3). These predictions of the PIMMS framework are reflected in our data, given that we observed more pronounced original weakening in both cases. The combinations of low prior precision with low prior accuracy, as well as high prior precision and high prior accuracy therefore seemed to have created medium PEs, causing weakening of the original episode in our data.

This is in congruence with the framework by Gershman et al. (2017), because medium PEs trigger updating of the old memory trace without establishing a new trace. As medium PEs may go along with medium reactivation of the original memory trace, a weakening of the original at this point is also in line with the Nonmonotonic Plasticity Hypothesis (Kim et al., 2014). Our data therefore supports previous frameworks using a more vigorous quantitative manipulation of PE size.

Large PEs were designed to be induced by high prior precision (i.e., five encodings) and low prior accuracy (i.e., high difference ratings; fourth quarter in Figure 3). In our data, at this highest level of PE, original retention increased again in comparison with medium PEs, which importantly went along with high modification learning as well. Following the Nonmonotonic Plasticity Hypothesis, the original should remain stable due to its strong reactivation, which is congruent with our data. The pattern is also in congruence with the framework by Gershman et al. (2017), in which large PEs signal a change of the underlying latent cause, which means that both the old and the new memory are valid and should be stored in separate traces. Instead of this account of pattern separation, it has also been suggested that the two traces may become integrated (Greve et al., 2018; Ritvo et al., 2019). While our

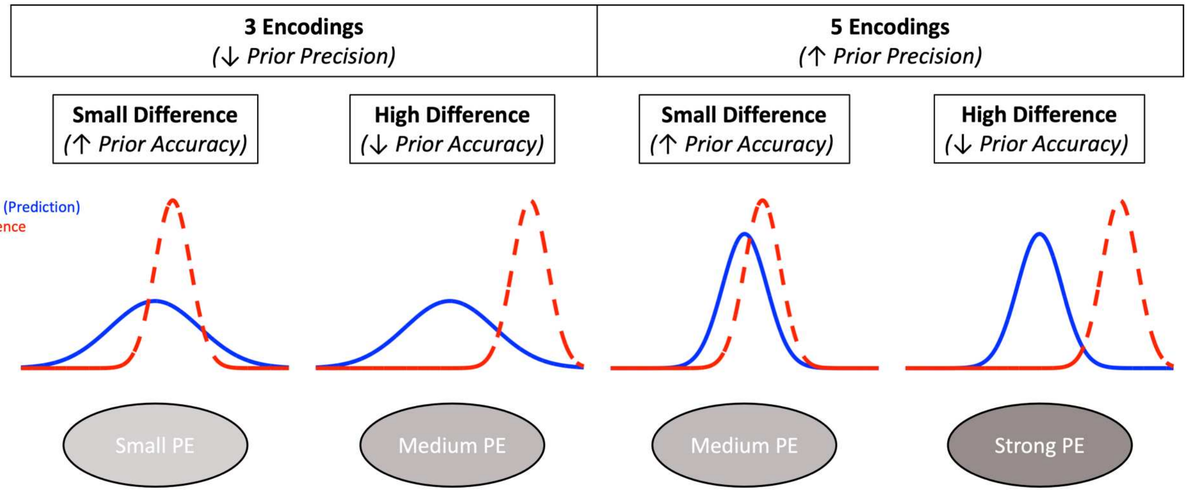

Figure 3. Explanation of the results within the PIMMS framework. Interpretation of our results in light of the PIMMS framework (Henson & Gagnepain, 2010). The prior is generally less precise, meaning that expectations are broader after three than after five encodings (hence the flatter blue line on the left than on the right). When an imprecise prior after three encodings is combined with a small difference between original and modification (the dotted red line's maximum is close to the blue line's on the leftmost quarter), there is a lot of overlap between prediction and outcome, so there is a small PE. If the difference is bigger (as the red curve moves away from the blue one, second quarter), the overlap is reduced and the PE is at a medium level. Similarly, with a precise prior after five encodings, even a small difference will lead to a medium-sized PE (third quarter). Lastly, when the prior is very precise and the modification big, there is a diminishing amount of overlap between prediction and input and the PE will be at its strongest (fourth quarter).

current behavioural data cannot distinguish between integration and separation, future neuroimaging studies may shed light on this issue (Bein et al., 2023).

The finding of original and modification memory being high after big PEs is also consistent with a theoretical account placing conscious awareness of the modification at its centre (Wahlheim & Zacks, 2019). This account states that there should be reduced interference between original and modified events when the change is noticed by a person and can later be recalled. We did not test participants on change perception and recollection, but it is possible to speculate that they were especially likely to notice the change and remember it later when episodes were well-encoded and modifications were large. As these were the conditions in which participants remembered both original and modification well, reduced interference between the episodes could offer an alternative explanation of this finding.

Collectively, the conditions of original weakening found in our data are congruent with previous accounts and support the role of medium-sized PEs in the weakening of original memories. After strong PEs, both original and modification were recalled well.

Interestingly, PEs did not impair recognition of the head of the dialogue that preceded the modification but was not altered itself. A forgetting caused by the weakening of associated elements (Joensen et al., 2020) was therefore not found here. This localisation of the effect tightly connects the observed memory modifications to the evoked PEs and makes overall decay of the episode an unlikely explanation. However, previous studies have found unimpaired (Sinclair & Barense, 2018) or even enhanced (Sinclair et al., 2021) correct original details after PEs. The robust memory for heads could therefore be seen as evidence for this enhancing effect of PEs. Future work should therefore attempt to establish how localised the effects of PEs are, also because this could delineate PEs from surprise, which can also enhance memory retroactively (Congleton & Berntsen, 2020).

An interesting general point that emerges from the data is that there was solid recognition memory for statements, which appears to be in contrast to previous research (Brown-Schmidt & Benjamin, 2018). The social-interactional information (e.g., E. Bates et al., 1978; Keenan et al., 1977), but also the expressively delivered and succinct dialogues may have increased memory. Furthermore, multiple encoding opportunities not being given in previous studies, our findings newly suggest that these may enhance memory for verbal statements. The results therefore suggest a fresh look at memory for conversations, especially as podcasts, streaming platforms, and taped lectures have joined the already popular radio plays and audiobooks, making repeated exposure to the same statements more common.

Conclusion

Using social and verbal material for the first time in the field, we showed that episodic PEs can lead to learning of new and unexpected information, as well as the weakening of the original and apparently outdated details. New learning linearly benefited from larger gist changes. Consistent with prior research and theoretical accounts, we found that medium PEs are especially likely to weaken the original, while strong PEs lead to the maintenance of both original and modified episodes.

While we have identified important influences on memory updating in the form of prior accuracy and prior precision, future work should further promote our understanding of other boundary conditions. Especially in the social domain, it could be highly maladaptive to forget information just because the situation has changed, pointing to a role of future anticipated usefulness, or past importance of the information. More work is therefore needed in the field that makes use of personally relevant, socially coloured, naturalistic, and complex stimulus material.

Acknowledgements

We thank Sophie Siestrup, Falko Mecklenbrauck, Anoushiravan Zahedi, and Inês Bramão for their helpful comments in the conceptualisation process. We also thank Larissa Stähler, Britt Hasslöver, Lana Steuernagel, and Lena Puder for their work during data collection as well as Lana Steuernagel, Tabea Krause, and Michelle Bellstedt for their help during stimulus construction.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

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

Data availability statement

The data, including a codebook, is available on OSF (https://osf.io/8dnsg/?view_only=3fc9fd3393474fd8b82969623c6ce6f9). Samples of the stimuli can be provided upon request to the corresponding author.

ORCID

Marius Boeltzig http://orcid.org/0009-0007-8930-0171
Nina Liedtke http://orcid.org/0009-0003-8537-2777
Ricarda I. Schubotz http://orcid.org/0000-0001-5802-6869

References

Andrews, J. J., & Rapp, D. N. (2014). Partner characteristics and social contagion: Does group composition matter? *Applied Cognitive Psychology*, *28*(4), 505–517. https://doi.org/10.1002/acp.3024

Anzellotti, S., & Young, L. L. (2020). The acquisition of person knowledge. *Annual Review of Psychology*, *71*, 1–22. https://doi.org/10.1146/annurev-psych-010419-050844

Atir, S., Wald, K. A., & Epley, N. (2022). Talking with strangers is surprisingly informative. *Proceedings of the National Academy of Sciences*, *119*(34), Article e2206992119. https://doi.org/10.1073/pnas.2206992119

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, E., Masling, M., & Kintsch, W. (1978). Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(3), 187–197. https://doi.org/10.1037/0278-7393.4.3.187

Bein, O., Gasser, C., Amer, T., Maril, A., & Davachi, L. (2023). Predictions transform memories: How expected versus unexpected events are integrated or separated in memory. *Neuroscience & Biobehavioral Reviews*, *153*, Article 105368. https://doi.org/10.1016/j.neubiorev.2023.105368

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

Boeltzig, M., Johansson, M., & Bramão, I. (2023). Ingroup sources enhance associative inference. *Communications Psychology*, *1*(1), Article 40. https://doi.org/10.1038/s44271-023-00043-8

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

Brown-Schmidt, S., & Benjamin, A. S. (2018). How we remember conversation: Implications in legal settings. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 187–194. https://doi.org/10.1177/2372732218786975

Calderon, C. B., Loof, E. D., Ergo, K., Snoeck, A., Boehler, C. N., & Verguts, T. (2021). Signed reward prediction errors in the ventral striatum drive episodic memory. *The Journal of Neuroscience*, *41*(8), 1716–1726. https://doi.org/10.1523/JNEUROSCI.1785-20.2020

Congleton, A. R., & Berntsen, D. (2020). It took me by surprise: Examining the retroactive enhancement effect for memory of naturally unfolding events. *Journal of Applied Research in Memory and Cognition*, *9*(3), 300–309. https://doi.org/10.1016/j.jarmac.2020.03.003

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

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed). Sage.

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

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

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

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

Harris, C. B., Barnier, A. J., Sutton, J., & Khan, T. (2017). Social contagion of autobiographical memories. *Journal of Applied Research in Memory and Cognition*, *6*(3), 319–327. https://doi.org/10.1016/j.jarmac.2017.07.006

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

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

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of*

Experimental Psychology: General, *149*(2), 230–248. https://doi.org/10.1037/xge0000648

Keenan, J. M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 549–560. https://doi.org/10.1016/S0022-5371(77)80018-2

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

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences*, *111*(24), 8997–9002. https://doi.org/10.1073/pnas.1319438111

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

Kroneisen, M., & Bell, R. (2022). Memory as a cognitive requirement for reciprocal cooperation. *Current Opinion in Psychology*, *43*, 271–277. https://doi.org/10.1016/j.copsyc.2021.08.008

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

Lenth, R. V. (2023). *emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans.

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

Mihara, M., Izumika, R., & Tsukiura, T. (2023). Remembering unexpected beauty: Contributions of the ventral striatum to the processing of reward prediction errors regarding the facial attractiveness in face memory. *NeuroImage*, *282*, Article 120408. https://doi.org/10.1016/j.neuroimage.2023.120408

Oliva, A. (2022). CA2 physiology underlying social memory. *Current Opinion in Neurobiology*, *77*, Article 102642. https://doi.org/10.1016/j.conb.2022.102642

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

Pezdek, K., & Prull, M. (1993). Fallacies in memory for conversations: Reflections on Clarence Thomas, Anita Hill, and the like. *Applied Cognitive Psychology*, *7*(4), 299–310. https://doi.org/10.1002/acp.2350070404

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

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

R Core Team. (2023). *R: A language and environment for statistical computing*. https://www.R-project.org/.

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

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, *5*(1), 1–9. https://doi.org/10.1016/j.jarmac.2015.12.003

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

Rosenblum, M., Schroeder, J., & Gino, F. (2020). Tell it like it is: When politically incorrect language promotes authenticity. *Journal of Personality and Social Psychology*, *119*(1), 75–103. https://doi.org/10.1037/pspi0000206

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

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

Siestrup, S., Jainta, B., Cheng, S., & Schubotz, R. I. (2023). Solidity meets surprise: Cerebral and behavioral effects of learning from episodic prediction errors. *Journal of Cognitive Neuroscience*, *35*(2), 291–313. https://doi.org/10.1162/jocn_a_01948

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

Siestrup, S., & Schubotz, R. I. (2023). Minor changes change memories: Functional magnetic resonance imaging and behavioral reflections of episodic prediction errors. *Journal of Cognitive Neuroscience*, *35*(11), 1823–1845. https://doi.org/10.1162/jocn_a_02047

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

Sinclair, A. H., Manalili, G. M., Brunec, I. K., Adcock, R. A., & Barense, M. D. (2021). Prediction errors disrupt hippocampal representations and update episodic memories. *Proceedings of the National Academy of Sciences*, *118*(51), Article e2117625118. https://doi.org/10.1073/pnas.2117625118

Son, J.-Y., Bhandari, A., & FeldmanHall, O. (2021). Cognitive maps of social features enable flexible inference in social networks. *Proceedings of the National Academy of Sciences*, *118*(39), Article e2021699118. https://doi.org/10.1073/pnas.2021699118

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. *Psychonomic Bulletin & Review*, *25*(6), 2373–2379. https://doi.org/10.3758/s13423-018-1481-2

Wahlheim, C. N., Smith, W. G., & Delaney, P. F. (2019). Reminders can enhance or impair episodic memory updating: A memory-for-change perspective. *Memory (Hove, England)*, *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. *Journal of Experimental Psychology: General*, *148*(1), 30–50. https://doi.org/10.1037/xge0000458