

Marius Boeltzig

Learning From The Unexpected:
How Prediction Error Size Shapes Episodic Memory

2026

Psychologie

Learning From The Unexpected:
How Prediction Error Size Shapes Episodic Memory

Inaugural-Dissertation

zur Erlangung des Doktorgrades

im Fachbereich Psychologie und Sportwissenschaft

der Universität Münster

Vorgelegt von

Marius Boeltzig

Aus Wolfenbüttel

- 2026 -

Dekanin:	Prof. Dr. Ricarda Schubotz
Erste Gutachterin:	Prof. Dr. Ricarda Schubotz
Zweite Gutachterin:	Prof. Dr. Kristina Liefke
Tag der mündlichen Prüfung:	_____
Tag der Promotion:	_____

To my mother.

And to mental time travel.

Acknowledgements

These 41 months of PhD have been marked by an incredible amount of luck. They have been challenging and not exactly devoid of work, but more rewarding, insightful, and fun than I could have imagined. The operatic length of these Acknowledgements may serve as a testimony to how many people have contributed to this overwhelming amount of luck (as well as to my unshakable tendency to use many words where few words would do just as well).

First, as I could not have done this or, in fact, anything else without them being the fantastic people they are, my parents Grit and Birko. I will forever admire them for their courage to leave a dissolving country behind in 1990, and even more for working harder than I will ever have to, to build themselves and their sons a life defined by opportunity instead of constraint. They taught me to take pride in work, and at the same time gave me all the freedom in the world to decide who and what I wanted to be. In fact, they actively enabled me and cheered me along, in whatever I decided to do – and some of the choices I made really must be classed as "whatever". They were there, always, wherever – in churches in tiny villages, in the bone-breaking cold at the shores of the Kama, and in the car or on the plane to wherever I had decided to live that week. This PhD is their achievement as much as it is mine.

While I used to be able to follow this and all other passions thanks to my mother, I now do it for her. And for my family, Birko, Benni, Lisa, Felix, Antje, and everyone else who has the privilege to be able to mentally travel in time and experience, again and again, what a wonderful person my mother was. Without wanting to rub it in, anyone who did not meet her really did miss out.

A quality that my doctoral supervisor, Ricarda Schubotz, shares with my parents is how much she trusted me. She agreed to every project, trip, and idea that I had, and often trusted my intuition more than I did. While many supervisors are reluctant to let their students do anything that is not strictly project-related, she helped me start studies I barely dared to dream about. Whenever I tell people about my confusing split between place of employment and residence, I stop to appreciate my luck: There is no second doctoral supervisor who would agree to such an arrangement, which made nothing easier for her, but everything better for me. Her perspectives challenged my thinking, sharpened my use of language, and her enthusiasm for the brain managed to convince even me. Her patience in explaining, in my fourth year of doctoral studies (in cognitive neuroscience!), how to find Heschl's gyrus is unmatched.

Thanks to Ricarda, I am a better scientist than I was 41 months ago.

Without Inês Bramão, there is every chance I would not be a scientist at all, and I would most certainly not research memory. While substantial credit must be given to Mikael Johansson, who I believe capable of igniting enthusiasm for memory research even in squirrels and household plants, Inês' insistence about the superiority of experimental methods and her simultaneous readiness to engage with my ideas, which must have seemed to her like a particularly obscure branch of sociology, got me to write my Master's thesis with her. While others saw their supervisor once a month, barely a day passed without Inês supporting me one way or another – and I have rarely met a person who is so generous with their help, even when she has no obligation or benefit to do so. I do not believe that Inês has ever read a text or listened to a presentation with less than 100% attention, as her questions and comments are always highly relevant and challenging, without her ever forgetting to be appreciative. With Inês' help, I returned to a place that feels like home, and if things continue to go as marvellously as they have, we will keep working together in that very place.

As science and working out now comfortably occupy around 90% of my inner monologue, the path to Linn Petersdotter, who is deeply involved in all of them, is invariably short. When Linn and I warmed up to each other during the neuroscience course in 2020 (before that, I had found her enthusiasm for physical exercise rather suspicious), a conversation started that has not really ended ever since. Linn is the person who many people miss in their working life: Somebody you can share everything with – any silly detail about your day, any small success and big frustration, someone who knows who and what you are talking about without lengthy explanations. I am lucky to share a work life with Linn, to run projects, write papers, discuss research, and (yet another person who has blessed me with saying yes to every crazy idea) to organise the Kyiv Winter School of Memory together. But I count myself even luckier to be her friend. Without Linn's energy and warmth, my life would be quieter and less exciting. Without her relentless flow of ever-new goals, my VO2 max would be much lower, and the *overall functionality of my body* (which is obviously the only reason why people do CrossFit) would be much worse.

There are many other people that helped me in seeing this thesis come to fruition while keeping my sanity. First and foremost, Nina Liedtke has shared every step of the way with me in the initial stages of the project. Co-writing 80 dialogues is a good way of getting to know each other very well very quickly. I am lucky to have had

someone as reliable, positive, and kind as Nina as a project partner. The same is true for my other two office mates, Sophie Siestrup and Falko Mecklenbrauck. They supported us in fMRI analysis, explained how to get batteries from the university (the former being considerably easier than the latter), and enthusiastically contributed to assigning names to made-up characters. Nina, Sophie, and Falko have been incredibly supportive when I was in Münster, but also generously took care of things for me when I was away, and showed admirable tolerance for my clean Hochdeutsch which they falsely (but benignly) claim to feature mushy vowels. Rosari Selvan has been great moral support at every step of the way, inside and outside the lab. Monika Mertens made experiments much easier to run and kept drop-out and in-scanner movement low, by applying her Monika magic that one can only dream to acquire. The whole Biological Psychology Group provided everything one would want from the wider lab family: Tough questions, random conversations, and countless fun things to think about, and I want to thank all former and current members for their support. The same is true for the Lund Memory Lab, whose members have been incredibly insightful and generous with feedback in all of my projects for the better part of six years.

Another manifestation of my incredible luck has been the amount of traveling I was able to do in these 41 months. Lab visits in Aarhus, York, Kraków, and Stockholm have brought me amazing collaboration partners, whose enthusiasm for my side projects fed back into energy for the thesis. I left every meeting with Scott Cole and Clare Rathbone more excited and motivated than I was before, and thanks to them, my inexplicable obsession with British politics became productive. Annette Bohn believed in the Flashbulb Memory project when I did not and her kindness has often made me think that dinner with her should not be an annual, but a more frequent affair (also, her husband's impassioned intervention made me start wearing a helmet on the bike, for which my parents are greatly thankful). Krystian Barzykowski, by common consent the most energetic man in memory research, has been incredibly generous with his appreciation and reminded me of how wonderful it is that our job is thinking about memory. Aidan Horner, whose questions and comments about the topics presented in this thesis were just as eloquent as his explanations why a pint is just the perfect size for a beer, will hopefully turn from short-term host to post-doctoral co-supervisor. There are many more people who I met during research visits and conferences, who have given me the benefit of their opinions, support, and in many cases friendship.

Part of this, and arguably the most important one, is the work that I have been lucky to carry out in Ukraine. Thanks to Anton Kurapov and Karine Malysheva, we could collect data from 400 Ukrainians and run a one-week Winter School with 100 admirably enthusiastic participants. While I am passionate about all of my projects, the Spohad Lab is in its own league, and I am immensely grateful for having been able to start it. My memory of standing on the Maidan for the first time in 2018 is still fresh – it overwhelmed me to see the very place of the Revolution of Dignity, which I observed from the most anti-democratic corner of Europe. I cannot wait to be back.

While Julia Petz has not (yet) been a collaborator or colleague, she has considerably contributed to my work. I know no other person who can switch so seamlessly between explaining what Plato wrote about memory (which, regrettably, did not make it into this thesis) and narrating the coffee-related perils of her brief reluctant supervision of a child. Knowing Julia to be my friend and being on often parallel personal and professional journeys since those autumn days in Riga has been illuminating, entertaining, and reassuring beyond measure.

As much as I love my job and even though it rarely actually feels like working, the last 41 months also set the stage for a thoroughly unpredicted development in my personal life: I started *enjoying* physical exercise. Therefore, I have to thank two activities: running around in the ever-present light rain of the Dortmund-Ems canal, and later in the freezing winds of the Öresund, as well as picking up heavy things from the floor in peculiar manners to loud music while muscular gays shouted at me – oddly enough – kept me sane, motivated, and grounded. Everyone contributing with coaching, motivation, and support is to be thanked for helping me on a long journey that has given me more happiness than I would have imagined at its beginning.

Most importantly, however, my friends and my family, as scattered in the winds as they may be, have never failed to support me when things got hard and to be happy for me when they were good. Whenever I see the wonderful people that have made the long journey since Wolfenbüttel, Perm, Potsdam, Wrocław, Riga, Lund, Münster, and many places in between, time and distance seam utterly powerless. Maybe Tulving was wrong and not only episodic memory can defy the merciless passing of time: maybe friendship does as well. While I am immensely proud to hand in this thesis, it does not compare to the pride I feel for the privilege to call myself a son, grandson, brother, brother-in-law, uncle, and friend.

Table of Contents

SUMMARY 1

LIST OF ORIGINAL PUBLICATIONS 3

1 THEORETICAL AND EMPIRICAL BACKGROUND 4

1.1 Episodic Memory: Representing the Past to Navigate the Future 5

1.1.1 Functions of a Flexible Memory System 5

1.1.2 Episodic Memory in the Brain 7

1.1.3 Integration and Separation 9

1.2 The Predictive Brain 10

1.3 Prediction Errors and Their Effect on Episodic Memory 12

1.3.1 Conditioned Predictions 13

1.3.2 Semantic Predictions 15

1.3.3 Episodic Predictions 16

1.3.4 Episodic Prediction Error Signalling 18

1.4 Prediction Error Size and Its Impact 19

1.4.1 Determinants of Prediction Error Size 20

1.4.2 Theories about the Effects of Prediction Error Size on Memory 22

1.4.3 Empirical Findings concerning Prediction Error Size and Memory 24

1.4.4 Representational Effects of Prediction Errors 27

1.4.5 Beyond Size: Prediction Error Types 28

2 RESEARCH QUESTIONS & DESIGN 30

3 ORIGINAL PAPERS 33

3.1 Study 1: Prediction errors lead to updating of memories for conversations 33

3.2 Study 2: The benefit of being wrong: How prediction error size guides the reshaping of episodic memories 45

3.3 Study 3: To update or to separate: Neural signatures and consequences of latent cause inference in episodic memory 61

4 DISCUSSION 108

4.1 Summary of the Presented Studies 108

4.2 Limitations and Strengths of the Design **110**
4.3 Memory After Prediction Errors of Different Sizes **113**
4.3.1 Small Prediction Errors 113
4.3.2 Moderate Prediction Errors 115
4.3.3 Large Prediction Errors 117
4.3.4 Taking Stock and Open Questions 118
4.4 Existing Explanations for Mnemonic Effects of Prediction Errors **120**
4.4.1 Reconsolidation Theory 120
4.4.2 Event Memory Retrieval and Comparison Theory 121
4.4.3 Dopamine 122
4.4.4 The Need for a Better Explanation 123
4.5 The REPRISE Framework **123**
4.5.1 Theoretical Foundations 124
4.5.2 Proposed Mechanism 125
4.5.3 Testable Hypotheses and Open Questions 128
4.5.4 Strengths of REPRISE 132
4.5.5 Predictive Reinstatement as the Gateway to Prediction Errors 133
4.6 Future Research **136**
4.6.1 Testing REPRISE 136
4.6.2 The Predictive Mix 137
4.6.3 Post-Prediction Error Buffering Factors 138
4.6.4 Bridging the Gap to Future Thinking 140
4.7 Implications **140**
4.7.1 Implications for Research 140
4.7.2 Wider Applications 141

5 CONCLUSION **144**
SUPPLEMENTARY MATERIALS **145**
LIST OF ABBREVIATIONS **148**
REFERENCES **149**
CURRICULUM VITAE **169**
DECLARATIONS **174**

Summary

According to predictive processing accounts, the brain continuously makes predictions about upcoming input. If this input is expected, no additional resources need to be deployed for its processing. However, if input was unpredicted, a prediction error (PE) arises, signalling that the current state of knowledge and experiences is not capable of adequately representing the world (Barron et al., 2020; Friston & Kiebel, 2009). PEs can therefore be a powerful driver of changes in episodic memory and support the flexibility that is needed to fulfil its future-facing functions (Schacter & Addis, 2009).

While consequences for episodic memory can be observed even after predictions based on semantic knowledge or conditioned response-outcome associations, they can also arise after episodic PEs (Bein et al., 2023). In these cases, predictions are based on episodic memories and PEs can lead to the encoding of a new memory but can also trigger the modification of the pre-existing memories that were leveraged for predictions. Two theoretical accounts suggest that the size of the PE can play a decisive role in shaping memory outcomes (Gershman et al., 2017; van Kesteren et al., 2012). Specifically, small PEs should not lead to any mnemonic changes, moderate PEs can lead to the integration of new information into an existing memory trace, and large PEs can trigger the separate encoding of a new memory trace, while the original memory is left unaltered. However, these hypotheses, and therefore the role of PE size, remain largely untested in episodic PEs.

In three studies, it was therefore tested how PE size affects memory for both the original and unpredicted mismatch event, representational outcomes, and univariate brain activity. Participants encoded naturalistic dialogues in an original form, before a statement in the middle of the conversation was unexpectedly altered to induce a PE. PE size was manipulated by using versions with different magnitudes of deviation compared to the original, by varying encoding opportunities for the original version, and by measuring prediction strength using representational similarity analysis. With a delay of at least one day, participants were tested on recognition memory for both the original and mismatch version.

Studies 1 and 2 showed that memory for original and mismatch targets was highest after small and large PEs but reduced after moderate PEs. Additionally, Study 2 found that source memory for the mismatch target was highest and neural

representations of the original event were most stable after large PEs. This supports theoretical accounts of no change after small PEs, integration after moderate PEs, and separate encoding of the mismatch event while the original memory is preserved after large PEs. Study 3 additionally found that moderate PEs were accompanied by activity in Brodmann Area 9 and the ventromedial prefrontal cortex, both of which are typically associated with integration and updating (Bowman & Zeithamova, 2018; Ferstl et al., 2005). The distinct encoding of the mismatch target after large PEs, however, was not associated with additional frontal activity, but instead with the reactivation of the original memory trace (Pronoza et al., 2025).

These results, the relevant theories, and other empirical studies are used to develop a theory of mnemonic consequences of episodic PEs. The Reactivation and Prediction Causing Integration and Separation in Episodic Memory (REPRISE) framework postulates that PE size is correlated with the strength of reinstatement of the original memory trace and contributes directly and indirectly to representational and behavioural outcomes of episodic PEs. The framework can explain previous mixed results and offer methodological implications for future studies. It may also be the basis for understanding the applicability of results about how PEs affect episodic memory to future thinking and societal challenges such as classroom learning, misinformation correction, and eyewitness testimony.

List of Original Publications

The following papers are presented in this doctoral thesis:

Study 1

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

Study 2

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

Study 3

Boeltzig, M., Siestrup, S., Bramão, I., & Schubotz, R. I. (2026). To update or to separate: Neural signatures and consequences of latent cause inference in episodic memory. *NeuroImage (under review in revised form)*.

1 Theoretical and Empirical Background

There are wrong predictions and there are catastrophically wrong predictions. Telling ourselves that "next week, things will quiet down a bit" is an inaccurate prediction. Nevertheless, we seem to learn very little from it and readily repeat the statement, no more right that time, as soon as opportunity permits. Predicting the end of history and the eternal victory of liberal democracy, as Francis Fukuyama did in 1992, was a catastrophically wrong prediction, and – unfortunately for Fukuyama – one that just about everybody remembers. That prediction, however, when its inaccuracy became apparent, was highly informative, revealing false yet widespread assumptions about history's linearity, nationalism, and the appeal of democracy.

We therefore seem to learn from predictions – and possibly more so when they are incorrect. This can happen even without making a prediction explicit, as the brain constantly predicts the immediate future without our conscious awareness (Friston & Kiebel, 2009; Rowe et al., 2020). Incorrect predictions then, so-called prediction errors (PEs), are a substantial driver of the encoding of new memories and the modification of existing ones (Bein et al., 2023). Since prediction error-based learning can happen automatically at any point, it is a powerful manifestation of the mnemonic flexibility needed to represent a dynamic, ever-changing environment.

However, this ubiquity has not yet manifested in a solid understanding of how episodic memories are transformed by PEs. While a majority of studies has tacitly assumed that all PEs are equal, I will argue that the outcomes of PEs crucially depend on PE size. Three studies are presented that support this idea, employing behavioural techniques, and both univariate and multivariate fMRI analyses. Together, they inform a graded and dynamic view of PE-dependent memory outcomes, which will be summarised within a novel framework of episodic PEs. This framework may inform theorising, decisions about study design, and application of episodic PEs to practical domains such as classroom learning and the correction of misinformation.

In the logic of the famous Anna Karenina principle, where all happy families are alike, but all unhappy families are unhappy in their own way, I will attempt to demonstrate that PEs are unhappy families that, making matters even worse, are treated by most like happy ones. This thesis aims to demonstrate that PE size needs to be considered as a crucial factor in future studies, and that it is fundamental for understanding how PEs contribute to the constructive nature of episodic memory.

1.1 Episodic Memory: Representing the Past to Navigate the Future

While it is well-known that Goethe's Werther is not a believer in rational self-control, he also does not believe that memories can be changed easily: "We are so constituted that we believe the most incredible things; and, once they are engraved upon the memory, woe to him who would endeavour to efface them" (von Goethe, 1774) he writes in a letter, when describing how he made up stories for his love interest's young siblings who would complain about every deviation in his re-tellings. While memory can sometimes be rather unmovable, there is ample evidence showing that memories are indeed not engraved, and that "he who endeavours to efface them" may find his mission achievable. Crucially, as will be discussed in the next section, this propensity for memory to change is fundamental for its functionality as a future-facing mental capacity.

1.1.1 Functions of a Flexible Memory System

Episodic memory is a notion first introduced by Endel Tulving, as a counterpart to semantic memory. Semantic memory was defined by Tulving as a "mental thesaurus", storing knowledge, concepts, and their relations (Tulving, 1972). Episodic memory, on the other hand, was introduced as a system storing events (i.e., episodes) with a temporal and spatial reference, and an autobiographical connection (Tulving, 1972). Later, this definition was supplemented by what is still the core element of episodic memory: mental time travel. This capacity allows not just knowing that past events happened, but being able to re-live them and being aware of the event as one's own experience (autonoetic awareness) from the perspective of the self (Tulving, 1983, 1985, 2002).

Many researchers believe that this capacity for mental time travel has not developed for its past-facing direction, but rather for its utility in the present and the future (but see Boyle, 2022). In addition to enabling a coherent sense of self across time and social bonding through the sharing of experiences, episodic memory is believed to be directive, meaning that past experiences guide behaviour in the present (Bluck et al., 2005; Rasmussen et al., 2015). In fact, cognitive tasks such as perception and navigation have been argued to be the very reason for the evolution of mental time travel (Osvath & Johansson, 2024).

Interestingly, the idea that memories can guide current behaviour and thought was already foreshadowed by Ebbinghaus as his third manifestation of memory (in

addition to voluntary and involuntary retrieval; Ebbinghaus, 1885).[1] This idea was later stated prominently by Addis and Schacter. Their work rests on neuroimaging data showing that remembering the past and imagining the future recruit overlapping brain regions, most notably the medial temporal lobe (MTL), including the hippocampus, as well as the medial prefrontal cortex (mPFC; Addis et al., 2007). Their constructive episodic simulation hypothesis postulates that to simulate the future, the brain flexibly recombines elements of past experiences (Schacter et al., 2012; Schacter & Addis, 2007a, 2007b; see also Mullally & Maguire, 2014). This requires a memory system that is indeed flexible and dynamic, allowing an adaptive, goal-oriented reconstruction instead of slavish replay.

Crucially, the constructive nature of memory is not limited to imagining the future: Even retrieving past events is believed to be reconstructive and rely on a record of the episode itself, but also relevant semantic information about the world and the self (Cheng, 2024; Conway & Pleydell-Pearce, 2000; Rugg & Renoult, 2025; Schacter & Addis, 2007a; Werning, 2020), as well as demands of the current situation (Schacter et al., 2023). While this may at first be counterintuitive, the principle that memories should be useful rather than perfectly accurate is necessary for memory to serve its directive function. As the past rarely repeats itself identically, it is necessary to flexibly draw on multiple sources to construct an adequate future scenario, or to inform decision making in the present.

As flexibility is necessary for memory's functionality, memory errors, so discrepancies between how the event really unfolded and how a person remembers it, are often a benefit rather than a flaw (Loftus, 2005; Loftus & Pickrell, 1995; Schacter, 2001, 2021). For instance, systematic memory distortions help to keep self-esteem high (Conway & Pleydell-Pearce, 2000; Schacter et al., 2023), mood positive (Walker et al., 2002), and groups connected (Hirst & Echterhoff, 2012). Even "classical" forgetting (i.e., fading over time) fulfils emotional and directive functions (Fawcett & Hulbert, 2020). In fact, when forgetting and distortions fail, this can have severe consequences, such as in the case of post-traumatic stress disorder, where a highly negative memory is overly persistent (Ehlers & Clark, 2000; Petersdotter et al., 2025).

[1] The language of the original version is something to be enjoyed: "Die Beschäftigung mit einem gewissen Gedankenkreise erleichtert unter Umständen die spätere Beschäftigung mit einem ähnlichen Gedankenkreise, auch wenn jene erste weder in ihrer Methode noch in ihren Resultaten direkt vor die Seele tritt. Das unermeßliche Gebiet der Wirkung angesammelter Erfahrungen gehört hierher." (Ebbinghaus, 1885, p.2)

1.1.2 Episodic Memory in the Brain

If memory is a flexible and constructive system, the question arises how this is implemented in the brain, and how it supports the encoding and retrieval of memories. When perceiving an event, its elements are represented across a wide array of neocortical structures, each processing different aspects of the situation. These structures signal their inputs to the hippocampus, which is believed bind all these elements together, and store an index (called a "summary sketch" by Barron et al., 2020) of this distributed activity (Horner & Doeller, 2017). Retrieval is then constituted by the reinstatement of similar neural patterns, so essentially the replay of brain activity (Rissman & Wagner, 2012; Rugg & Renoult, 2025). This idea implies that neural patterns in the neocortex at encoding and retrieval are correlated, and that the hippocampus (at least for younger memories) drives the process of reinstating a memory based on a retrieval cue, using the stored index (Horner et al., 2015; Joensen et al., 2024). This process, called pattern completion, thus consists of the retrieval of all elements of an event if one of the elements is cued (Horner et al., 2015).

As a methodological aside, it is precisely this reinstatement of neural patterns during retrieval that enables the use of multivariate brain analyses such as representational similarity analysis (RSA; used in Studies 2 and 3). RSA with fMRI data measures correlations between neural patterns at two time points (Dimsdale-Zucker & Ranganath, 2018; Rissman & Wagner, 2012), and can therefore be used to test how strongly a given episode is being reactivated at a certain time point (Studies 2 and 3), or how strongly representations[2] change due to a manipulation (Study 2).

The functions of binding elements together, storing an index, and orchestrating reinstatement place the hippocampus at the centre of episodic memory (Nadel et al., 2012; Nadel & Hardt, 2011; Xue, 2022). The hippocampus is capable of rapidly storing individual events and binding elements of episodes together into a contextualized memory trace[3] (Kumaran et al., 2016). While the hippocampus is responsible for the

[2] The term representation is not uncontroversial, as it is inconsistently connected to claims of correlation, causality, or teology (Baker et al., 2022; Rugg & Renoult, 2025). Here, I refer to it broadly as a systematic neural correlate of an event or stimulus, aligning it with what RSA can actually measure (Dimsdale-Zucker & Ranganath, 2018).

[3] The term memory trace is similarly contentious. Here, it is sufficient to understand a memory trace as a neural network encapsulating memorial content, allowing the reinstatement of neural activity from encoding at retrieval (see De Brigard, 2014 for a philosophical evaluation). The role of the hippocampus in the retrieval of older memories (McClelland et al., 1995), or whether existing traces get modified or new memory traces are encoded at each retrieval (Agren, 2014; Nadel & Moscovitch, 1997; Robin & Moscovitch, 2017; Sutherland et al., 2020), are less critical for the presented work.

storing of individual, separate episodes, it has been argued that the neocortex represents abstract and structured knowledge, acquired more slowly over many instances (Kumaran et al., 2016). An especially important area for this purpose is the medial prefrontal cortex (mPFC), which contributes to the representation of schemas, and exhibits increased coupling with the hippocampus when events are schema-congruent (Giuliano et al., 2021; van Kesteren et al., 2012; Wang & Morris, 2010). However, a clear division of labour between the hippocampus and the prefrontal cortex has been questioned recently, given that even the hippocampus can also support rapid generalization across episodes (Kumaran & McClelland, 2012; Zeithamova & Bowman, 2020).

While the involvement of the hippocampus and mPFC in episodic memory is widely accepted (also see Kolibius et al., 2025), how memories can change over time or during retrieval, introducing the errors and flexibility introduced above, is subject to debate. There are several theories about the neural processes underlying these distortions (Nadel & Hardt, 2011; Nadel & Moscovitch, 1997; Robin & Moscovitch, 2017; Sutherland et al., 2020). For the present purposes, reconsolidation theory is the most important framework, due to its popularity in the field of PE-induced memory processes (Sinclair & Barense, 2019). Reconsolidation theory posits that memories turn labile when they are retrieved, opening a reconsolidation window. While this window is open, memories are vulnerable for change, for instance due to the injection of amnestic agents in animal experiments (Exton-McGuinness et al., 2015; Fernández et al., 2016), or the presentation of interfering information for humans (Hupbach et al., 2007; Scully & Hupbach, 2020). This information may be added to the original memory trace and cause intrusions when participants are tested on the original event (Hupbach et al., 2009; Parks et al., 2022). Without such interference, the reconsolidation window closes again without changes to the memory. Reconsolidation rests on a molecular process of protein synthesis and synaptic plasticity and is therefore time-dependent. The reconsolidation window may open 3-10 minutes after reactivation (Monfils et al., 2009; Sinclair & Barense, 2018) and be closed after six hours (Schiller & Phelps, 2011). While reconsolidation was initially based on animal studies, pharmacological and behavioural techniques have also been employed for humans using list learning tasks (Agren, 2014; Forcato et al., 2007). Nevertheless, strong doubts remain whether there is sufficient evidence for reconsolidation being applicable to humans (Elsey et al., 2018; Gisquet-Verrier & Riccio, 2018, 2019).

1.1.3 Integration and Separation

As discussed in the last section, a clear division of labour between the mPFC supporting generalization and the hippocampus supporting pattern separation has been questioned (Kumaran & McClelland, 2012; Zeithamova & Bowman, 2020). This raises the issue how a given event is stored and how the brain handles similar episodes. This is especially pertinent in the context of PEs, as different versions of an event will often share substantial similarity.

There are two alternative ways for the brain to address similarity and overlap in episodes. The first strategy, memory integration, refers to storing similar or overlapping episodes as a unified representation, emphasizing their shared features. This allows flexibly drawing inferential relationships between elements that were not directly witnessed together (Zeithamova et al., 2012), and the learning of abstract concepts (Mack et al., 2018). On the other hand, the abstraction and generalization may be detrimental to source and detail memory (Carpenter & Schacter, 2017, 2018a, 2018b; but see Boeltzig et al., 2023; de Araujo Sanchez & Zeithamova, 2023), and induce false memories (N. L. Varga et al., 2019). Consistent with this focus on generalization, it has been found that frontal regions such as the ventromedial prefrontal cortex (vmPFC) are active during the integration of episodes (Bowman & Zeithamova, 2018; Spalding et al., 2018; Zeithamova et al., 2012).

The second strategy, pattern separation, consists of separately storing episodes as distinct memory traces. This enables targeted mental time travel to a single episode, as individual events are stored with their unique details and context (Yassa & Stark, 2011). There are different degrees to this process: Memories can be stored distinctly (Kuhl et al., 2010), but also actively pushed apart by exaggerating their differences in processes such as differentiation or repulsion (Chanales et al., 2017; Drascher & Kuhl, 2022; N. L. Varga et al., 2025). Crucially, pattern separation prevents interference. Generally, similar memory traces can compete with each other at retrieval, given that they are activated by the same retrieval cue, and thereby impede or even prevent successful retrieval. The mechanisms that resolve this competition can also lead to forgetting (Anderson, 2003; Anderson & Hulbert, 2020). By making similar experiences less similar, this interference can be reduced, lowering the risk of forgetting (Brunec et al., 2020; Chanales et al., 2017; Drascher & Kuhl, 2022; Favila et al., 2016).

Crucially, it is not a binary matter whether a given event is integrated or

separated. Instead, both processes can be carried out simultaneously by different hippocampal subfields or cortical regions (Bein, Reggev, et al., 2020; Molitor et al., 2020). Additionally, different aspects of the same event can be integrated or separated (Liu et al., 2026). Task demands can also trigger representational change, as initially separated episodes can become integrated if they need to be connected (Carpenter & Schacter, 2017, 2018a, 2018b; de Araujo Sanchez & Zeithamova, 2023; Zeithamova & Bowman, 2020). Importantly, at encoding, both processes can be triggered by the reactivation of an old memory. If a current event is similar to a previously experienced one, it can act as a retrieval cue, meaning that the old memory is activated while new input is processed. This co-activation can lead to integration (Zeithamova et al., 2012), but also distinct encoding (Kuhl et al., 2010), depending on the circumstances.

Together, integration and separation contribute to the flexible and constructive nature of episodic memory introduced in section 1.1.1. The existence of alternative outcomes to address mnemonic similarity in different situations is powerful evidence that episodic memory is an active process that goes far beyond a simple storing of records.

1.2 The Predictive Brain

One way in which memory can fulfil its directive function, guiding behaviour and decisions in the present and future, is predictive processing. Just as memory, the wider cognitive system is constructive and active: The brain is far from being a passive observer, as it may continuously predict the future and strive to decrease the uncertainty with which it does so (Bar, 2009; den Ouden et al., 2012; Schacter & Addis, 2007a). As many aspects of the world are repetitive (e.g., the concept of an apple remains largely stable over time), robust internal models can adequately reflect it (de Lange et al., 2018). Consequently, the repetitive, stable, and therefore predictable features of sensory inputs do not need to be further processed, while the unpredicted, and therefore relevant share needs to be attended to (Huang & Rao, 2011).

This idea is formulated in predictive processing theories (e.g., predictive coding; Friston, 2019; Friston & Kiebel, 2009). They postulate that the brain continuously makes predictions about upcoming sensory input and compares these predictions with unfolding reality. This allows saving resources: If a stimulus was predicted, it offers no information that was not previously known and therefore does not need to be further investigated. However, if predictions and input mismatch, a PE

arises, signalling the need for attention and learning to keep mental models aligned with a changed reality (Clark, 2013; Friston & Kiebel, 2009; Reichardt et al., 2020). Therefore, no resources are spent on redundant and stable information, but new and unpredicted information is processed and used to improve existing models.

PEs can thus be described as the difference between the predicted and actual outcome (Fernández et al., 2016). Consequently, they provide information helping to tune existing models and thereby reduce unpredictedness in the future (den Ouden et al., 2012) when a better prediction can be made (Hohwy, 2020). The updating mechanism can be described by Bayesian principles, where a prior (i.e., the prediction) is updated with evidence (i.e., the input), producing a posterior, which can later become the new prediction (Henson & Gagnepain, 2010; Hohwy, 2020; this will be discussed in more detail in section 1.4.1).

As internal models will never perfectly predict a highly dynamic world, PEs always occur to some degree. However, not all unpredicted details will be relevant for high-level internal models. Consequently, there are two circumstances protecting the brain against being flooded by PEs. First, second-order predictions can be made about how reliable a PE signal is in a given situation. These second-order predictions adjust PE precision, determining how much gain the PE gets and how much it will be used for model updating. For instance, in volatile environments (e.g., a noisy room or a visually busy scene), PEs are less reliable and therefore have less gain (Barron et al., 2020). Second, the predictive process is hierarchical: higher levels generate predictions that are passed on to lower levels, and PEs are passed upwards in the hierarchy as 'newsworthy' information (Barron et al., 2020). For instance, in the visual system, the primary visual cortex would be fed with predictions by the secondary visual cortex, which in turn gets predictions from higher levels. PEs travel in the opposite direction as the primary visual cortex would relay PEs to the secondary visual cortex (Huang & Rao, 2011; K. M. Lee et al., 2021). Each level can therefore explain away the PE with the prediction it was fed, and at each level, the PE is weighted by precision before travelling up the hierarchy if still unexplained and relevant (Clark, 2014).

Foundational work on predictive processing focused mostly on more low-level sensory prediction (Clark, 2014; de Lange et al., 2018; Huang & Rao, 2011), for instance in oddball paradigms (den Ouden et al., 2012). However, predictions are also made about complex social and non-social actions of others (Brown & Brüne, 2012; Schubotz, 2007; Selvan et al., 2024) and in verbal conversations (Ufer & Blank, 2024),

which are used as stimuli in this thesis. Accordingly, PEs have been observed across a wide array of brain regions, cortical and subcortical, and all along the processing hierarchy (den Ouden et al., 2012).

It is a crucial quality of predictive processing that it happens automatically and without conscious awareness. Therefore, PEs are not always accompanied by a feeling of surprise and can go undetected by participants, despite showing a clear neural signature (Rowe et al., 2020). Consequently, the learning that PEs trigger is automatic, making it a highly powerful process with a strong impact on episodic memory.

1.3 Prediction Errors and Their Effect on Episodic Memory

Even before predictive processing theories were formulated, the observation that we remember unusual events better was formulated by Hedwig von Restorff (1933). Research on the von Restorff effect often targeted perceptual distinctiveness (e.g., letters in a list of numbers, or a different colour in a word list; Hunt, 1995), but has also been extended to higher complexity, such as stimuli with a distinct emotion (Schlüter et al., 2019).

Predictive processing can provide an explanation for the von Restorff effect, namely that the stimulus was unpredicted.[4] As discussed above, PEs signal that the internal models need to be tuned to a changing reality. They can therefore be seen as learning signals, transferring newsworthy information that needs to be included in the internal models (Barron et al., 2020; Exton-McGuinness et al., 2015; Fernández et al., 2016). Therefore, changes in memory are a fundamental outcome of PEs (Reichardt et al., 2020) and can take several forms, depending on the context and type of prediction. While existing semantic knowledge and schemas can be updated by PEs (Bein & Niv, 2025), a growing body of research has also focused on changes to episodic memory, so to representations of distinct events, as a result of PEs.

This relatively recent development (Bein et al., 2023; Greve et al., 2017; Nolden et al., 2024) places predictive processes at the core of a flexible and dynamic memory system, providing an explanation of how experiences are kept in tune with a changing environment. Crucially, episodic memory as an outcome can be influenced by many different PEs, distinguished by their source of prediction. These fall into three broad categories (Figure 1), namely conditioned PEs (i.e., predictions about pain or reward),

[4] Note, however, that distinctiveness and unpredictedness are not synonymous: it is possible that distinct stimuli are expected, and that unexpected stimuli are not distinct.

semantic PEs (i.e. predictions based on abstract rules or schemas), and episodic PEs (i.e. predictions based on previously encoded episodic memories). Even though this thesis focuses on episodic PEs, where both the source and the outcome is episodic memory, the other two categories will also be reviewed. As conditioned and semantic predictions can trigger effects on episodic memory, these research areas carry important methodological and theoretical implications for this thesis.

1.3.1 Conditioned Predictions

Instead of anticipating what exactly will happen, predictions can focus on how positive or negative an outcome will be, preparing for the hedonic value of an upcoming stimulus (see upper row in Figure 1). This happens when one is about to open grandma's Christmas letter, knowing how much money she usually attaches to her note, or just before the physiotherapist starts pushing into a spot which was particularly painful during the last session. The actual monetary value or intensity of pain may then deviate from the prediction, causing a PE. In laboratory experiments, participants usually gradually learn via conditioning which classes of stimuli are paired with positive or negative outcomes – often small monetary rewards or unpleasant electroshocks (Loock et al., 2025; Rouhani et al., 2020).

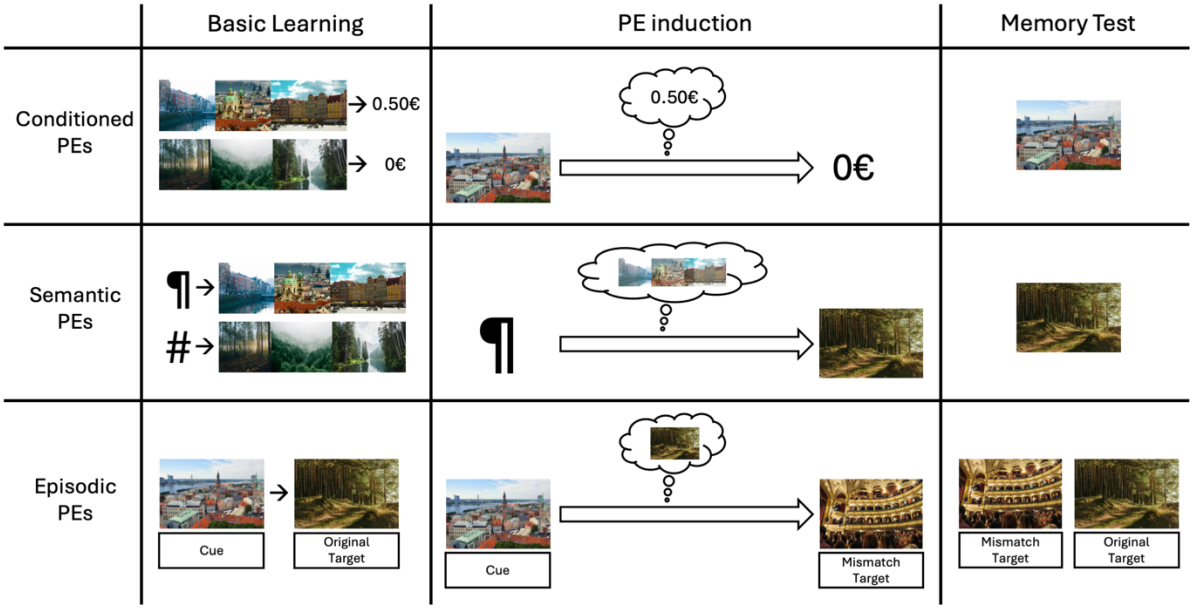
As a coarse conclusion, conditioned PEs generally enhance episodic memory (see Ergo et al., 2020 and Rouhani et al., 2023 for reviews). Therefore, stimuli paired with unexpected reward or pain are remembered better. This memory enhancement of PEs has been found using different retention intervals, encoding instructions, and task characteristics (Pupillo & Bruckner, 2023). Nevertheless, this outcome is not universal across studies (Rouhani et al., 2023). A more detailed analysis of these findings reveals two nuances that are informative for the study of episodic PEs.

First, studies differ in whether they find effects of signed or unsigned PEs. Signed PEs take the direction of mispredictions into account, so whether outcomes are better than expected (positive PE) or worse than expected (negative PE). For instance, if 20€ were predicted to be attached to grandma's letter, this will cause a positive PE if one gets 30€, but a negative PE if it is only 10€. Unsigned PEs, on the other hand, do not make this directional distinction, but only indicate that the outcome differed from the prediction by a magnitude of 10€.

Some studies focused selectively on signed (Calderon et al., 2021; Kalbe & Schwabe, 2021) or unsigned (Kalbe & Schwabe, 2020; Rouhani & Niv, 2019) PEs.

Figure 1

Schematic overview of PE sources



Note. Basic learning refers to what participants encode before a PE is induced, namely associations between stimulus categories and reward or punishment (conditioned PEs), cues and stimulus categories (semantic PEs), or between two unique stimuli (episodic PEs). These learning experiences can then be leveraged to predict an upcoming stimulus (in the thought bubble) upon presentation of the cue. A PE is produced if this predicted outcome fails to materialize and a different stimulus or hedonic outcome is presented. Episodic memory can then be tested only for the cue in conditioned PEs, only for the unpredicted outcome in semantic PEs, and for both (although this is rarely done simultaneously – a shortcoming addressed in the current studies) after episodic PEs. All images are obtained from unsplash.com under a free-to-use license.

Additionally, some investigations have directly compared whether signed or unsigned PEs are more predictive of memory outcomes – with no clear frontrunner across studies (H. Chen et al., 2025; Y. Chen et al., 2025; Loock et al., 2025; Rouhani et al., 2018; Rouhani & Niv, 2021).

An intriguing suggestion why findings are mixed is that effects of signed and unsigned PEs may depend on whether participants actively generate predictions (H. Chen et al., 2025; Ergo et al., 2020). In paradigms where participants need to explicitly cast a prediction (Rouhani et al., 2018), negative PEs are attributable to their own incorrect judgements (thus being their own fault), possibly engaging the noradrenergic system with its response to both punishment and reward. On the other hand, if predictions arise from the task itself (e.g., in paradigms where participants

choose one of several options without any basis for this decision, making winnings a game of chance; Calderon et al., 2021), the focus is on obtaining rewards and negative PEs are not one's own fault. This could engage the dopaminergic system instead (H. Chen et al., 2025). While a direct test of this hypothesis is currently lacking, the idea that active predicting may influence the predictive process will become relevant in section 4.5.5.

Second, timing effects have been studied with conditioned PEs more extensively than in semantic or episodic PEs. Memory enhancements have been found for stimuli preceding (Kalbe & Schwabe, 2020, 2021), co-occurring with (Y. Chen et al., 2025; Jang et al., 2019; Rouhani et al., 2020), or following (Loock et al., 2025) the PE. In other designs, these seem to be mutually exclusive, as for instance Jang et al. (2019) only found effects for co-occurring but not preceding or subsequent stimuli. Depending on the paradigm, with the boundary conditions not yet elucidated, the effects of PEs may therefore be short-lived and highly targeted (Jang et al., 2019), or may open a certain window of increased encoding (Loock et al., 2025). This is relevant in naturalistic settings, such as the design used in this thesis, where stimuli are extended in time beyond the moment of PE induction.

1.3.2 Semantic Predictions

Conditioned predictions specifically target the hedonic value of an upcoming stimulus, but do not encompass predicting the upcoming stimulus itself. This task can be fulfilled by semantic predictions (see middle row in Figure 1), which can for instance entail predicting the purpose of the waiter coming to one's table, or where one would look for the train station in an unfamiliar city. In experiments, semantic PEs have been induced based on such pre-existing real-world schemas and knowledge (Antony et al., 2021, 2023; Ben-Yakov et al., 2022; Brod et al., 2018, 2022; Congleton & Berntsen, 2020a, 2022; de Bruïne et al., 2025; Haeuser & Kray, 2022; Theobald & Brod, 2021; D. K. Varga et al., 2025), or abstract rules learned in the experiment (Brod et al., 2015; Frank et al., 2020, 2022; Greve et al., 2019; Jano et al., 2024; Kafkas, 2021; Kafkas et al., 2025; Kafkas & Montaldi, 2018).

Similarly to conditioned predictions, the broad conclusion of studies using semantic PEs is that they enhance subsequent memory (see Brod et al., 2018; Haeuser & Kray, 2022 for exceptions). One aspect relevant for the present purposes is the reach of semantic PEs. While studies, similarly to conditioned PEs, have been inconclusive

whether PEs also enhance memory for PE-preceding elements (Ben-Yakov et al., 2022; Congleton & Berntsen, 2020a), there is evidence that they can trigger encoding of peripheral elements (Congleton & Berntsen, 2022; Greve et al., 2019; but see Ben-Yakov et al., 2022). These are non-central parts of the episode for which no prediction was made – which are especially prevalent in naturalistic stimuli, which are richer in details than the stimuli used for conditioned predictions.

Another important aspect is the effect of explicit predicting, which has also been discussed in the context of conditioned PEs. Within semantic PEs, Brod and colleagues directly tested whether explicitly making a prediction is necessary for a memory benefit. Comparing explicitly predicting to a condition where predictions are made retrospectively, after knowing the outcome, the authors found that explicit predictions are indeed necessary for PEs to enhance memory (Brod et al., 2018, 2022). However, they also argue that explicitly predicting an outcome may be a compensatory mechanism for an otherwise weak prior (Brod et al., 2022). In other words, in the absence of knowledge that would support a prediction, this lack can be compensated by simply having to make one.

1.3.3 Episodic Predictions

Semantic knowledge may not always be the best source for predictions. If one goes to a relentlessly unconventional restaurant for the second time, it can be more adaptive to make predictions about the waiter's behaviour based on the memory of the first visit than based on a restaurant schema that does not apply. The same is true when no schema is established for a given situation, possibly due to its novelty or because no regularities can be extracted (H. Kim et al., 2019). Furthermore, when people repeat experiences, traveling to the same destination or watching the same movie (O'Brien, 2019), episodic memories may yield the most effective predictions. Lastly, when listening to someone sharing their memory of an event that one also witnessed, tapping the episodic memory of this event to predict what they are going to say next is more adaptive than using an overall schema.

Consequently, a line of research has investigated individual experiences, stored as episodic memories, as the basis for prediction. Paradigms typically involve participants encoding episodes broadly following a cue-target structure (see bottom row in Figure 1). Later, the same cue is presented again, allowing the prediction of the

original target, but a mismatch target is presented instead, inducing a PE.[5] The use of episodic memories for prediction in these paradigms has been corroborated by measures of neural reinstatement (Bein, Duncan, et al., 2020; G. Kim et al., 2014; H. Kim et al., 2019; Long et al., 2016), hippocampal temporal correlation (Sinclair et al., 2021), response times (Greve et al., 2017), and predictive looking (Eisenberg et al., 2018; Wahlheim et al., 2021).

Studies differ widely in what constitutes these cue-target associations. In the simplest case, they are arbitrary associations between static images (Bein et al., 2021; Greve et al., 2017; G. Kim et al., 2014, 2017; H. Kim et al., 2019; Long et al., 2016) or words (Wahlheim et al., 2019), allowing high experimental control. However, many studies have increased ecological validity by using more complex and naturalistic stimuli such as verbal reports (Vlasceanu et al., 2018), narratives played out with toy figures and objects (Jainta et al., 2022, 2024; Siestrup et al., 2022, 2023; Siestrup & Schubotz, 2023), or videos depicting everyday actions (Stawarczyk et al., 2020; D. K. Varga et al., 2025; Wahlheim et al., 2021; Wahlheim & Zacks, 2019). Additionally, as a relative outlier, Sinclair and colleagues (Sinclair et al., 2021; Sinclair & Barense, 2018) induced PEs by unexpectedly interrupting the replay of a familiar video. The share of paradigms using such ecologically valid material seems higher than in other fields of memory research, possibly to adequately capture the ongoing, temporally structured nature of predictions.

After episodic PEs, testing for episodic memory outcomes becomes less straight-forward, as there are two potential memories involved: The original target that was used for the prediction may be modified by the PE, and additionally, a new memory trace for the unexpected mismatch target, which evoked the PE, may have been formed (see Figure 1). Any combinations between modifications to the original and mismatch encoding, as well as the formation of connections between the two events are also possible. Evidence for all these processes has been found in different studies.

First, and most closely related to studies of semantic predictions, new information is often learned better after a PE (Greve et al., 2017; Long et al., 2016; Stawarczyk et al., 2020; Wahlheim & Zacks, 2019). Consequently, the unpredicted

[5] Note that these terms will be used throughout, regardless of the stimulus material. Original targets refer to the originally encoded information that can be predicted upon re-exposure of the cue. The mismatch target is the information that is then presented to evoke a PE (also see Figure 1).

mismatch target usually is encoded better than control events, even though this is not universal across designs and tests (e.g., D. K. Varga et al., 2025).

Second, PEs can influence the original episode, which was leveraged for the (unsuccessful) prediction. Several studies have observed that the original can be weakened by PEs, making it less likely to be recognized later (G. Kim et al., 2014; H. Kim et al., 2019; Vlasceanu et al., 2018). This also reflects the purpose of PE-driven learning: Unexpected information needs to be encoded, but potentially outdated information, no longer reflecting the state of the environment, may also have to be pruned from memory. However, the very opposite effect where originals are strengthened by PEs has also occasionally been observed (Sinclair et al., 2021; T. A. Smith et al., 2013).

Third, the original and mismatch targets may get interwoven in different ways. After PEs, participants often erroneously report details from the mismatching episode when trying to recall the original (Sinclair et al., 2021; Sinclair & Barense, 2018) or struggle to differentiate the two versions, resulting in source memory errors about which episode was the original one (Jainta et al., 2022, 2024; Siestrup et al., 2022; Siestrup & Schubotz, 2023).

It is a methodological problem that the evidence for these processes comes from separate studies, as most experiments tested memory either for original or mismatch targets (Bein et al., 2021 is an exception, but provided no independent analysis of the two). However, as discussed in section 1.1.3, there are different ways of addressing the similarity inherent in original and mismatch targets as they can become integrated or separated, each outcome with unique behavioural consequences. Which outcome is favoured under which circumstances is therefore a crucial question that can only be addressed if both original and mismatch memory are tested within the same experiment – a critical condition observed in all three studies presented here (see Chapter 2).

1.3.4 Episodic Prediction Error Signalling

A functional localization of episodic PE responses was not the main aim of this series of studies (for this, see Liedtke et al., 2025, using data from Study 2). Nevertheless, two regions with high relevance for Studies 2 and 3 should be briefly introduced.

First, the hippocampus has been argued to be placed at the top of the processing

hierarchy in the brain (Barron et al., 2020). This is a consequence of its microstructural architecture (Felleman & Essen, 1991) and converges with its prominent role in encoding and retrieval (see section 1.1.2) and the importance of memory in predicting (Reichardt et al., 2020). The hippocampus may switch between encoding and retrieval states (Long & Kuhl, 2019; B. E. Sherman & Turk-Browne, 2020) and PEs can interrupt ongoing prediction (Sinclair et al., 2021), biasing the hippocampus towards encoding (Bein, Duncan, et al., 2020). Additionally, the hippocampus itself acts as a mismatch detector as subfield CA1 compares predictions from CA3 with sensory input from the neocortex (Barron et al., 2020; J. Chen et al., 2015). The hippocampus is therefore activated by PEs (Groves et al., 2025; Shing et al., 2023). Importantly, however, a recent study suggests that this is mainly the case for episodic PEs, where input mismatches previous experiences, but not for semantic PEs, which violate existing schemas (D. K. Varga et al., 2025).

Second, the inferior frontal gyrus (IFG) robustly scales with PE relevance or informativeness across domains. It has been found to respond to PEs in perceptual tasks (Fujitani et al., 2024; M. T. Sherman et al., 2016; Takasago et al., 2024), stimulus sequences (Schliephake et al., 2021), action observation (El-Sourani et al., 2020; Wurm & Schubotz, 2012), and, crucially, also in predictions based on semantic (Kafkas & Montaldi, 2015) and episodic memory (Jainta et al., 2024; Siestrup et al., 2023; Siestrup & Schubotz, 2023). Specifically, the IFG has been suggested to be involved in generating and maintaining higher-order predictions, explaining its relevance across domains (Fujitani et al., 2024). As these higher-order predictions reflect highly idiosyncratic knowledge and experiences, IFG activation after PEs is moderated by relevant expertise: Musicians show stronger IFG responses to PEs in melodies than non-musicians (Bonetti et al., 2022), and the IFG responds to magic tricks in naïve participants, but not in magicians, for whom these tricks do not involve unpredictedness (Danek et al., 2015).

1.4 Prediction Error Size and Its Impact

Even though prior research has been conducted with different PE types (conditioned, semantic, and episodic) and using a variety of methodological approaches within episodic PEs, I have foreshadowed at the beginning of this thesis that previous research has often treated PEs as happy families within the Anna Karenina principle, assuming they are all alike. This is reflected in the literature in two

ways: First, many studies, especially within episodic PEs, assumed that their experimental manipulations always yield a comparable PE, identical across different trials. Second, studies have been compared with each other without acknowledging that the PEs they evoked may differ.

It has consequently been overlooked that the size of the PE is one crucial dimension (albeit not the only one; see section 1.4.5 and Bubic et al., 2010) along which PEs may differ, and which directly influences memory outcomes. This is in line with theorising (Henson & Gagnepain, 2010; Quent et al., 2021), which will be reviewed in the following. However, it can also be intuitively understood: If the doorbell rings and one predicts that it is the mailman, should it not make a difference for later memory whether it is actually a fundraiser (small PE), or the pope (large PE, unless one lives in the Vatican)? The following sections will introduce how PE size is determined, review how it can affect mnemonic outcomes, and discuss first empirical evidence for this.

1.4.1 Determinants of Prediction Error Size

The Predictive, Interactive Multiple Memory Systems framework (Henson & Gagnepain, 2010), PIMMS for short, lays out how PEs differ from each other in size, and how PE size can be manipulated in experiments. In Bayesian terms, PIMMS distinguishes between a prior, so the prediction, and the evidence, so the actual input (Figure 2A). There are three factors that can be varied to manipulate PE size.

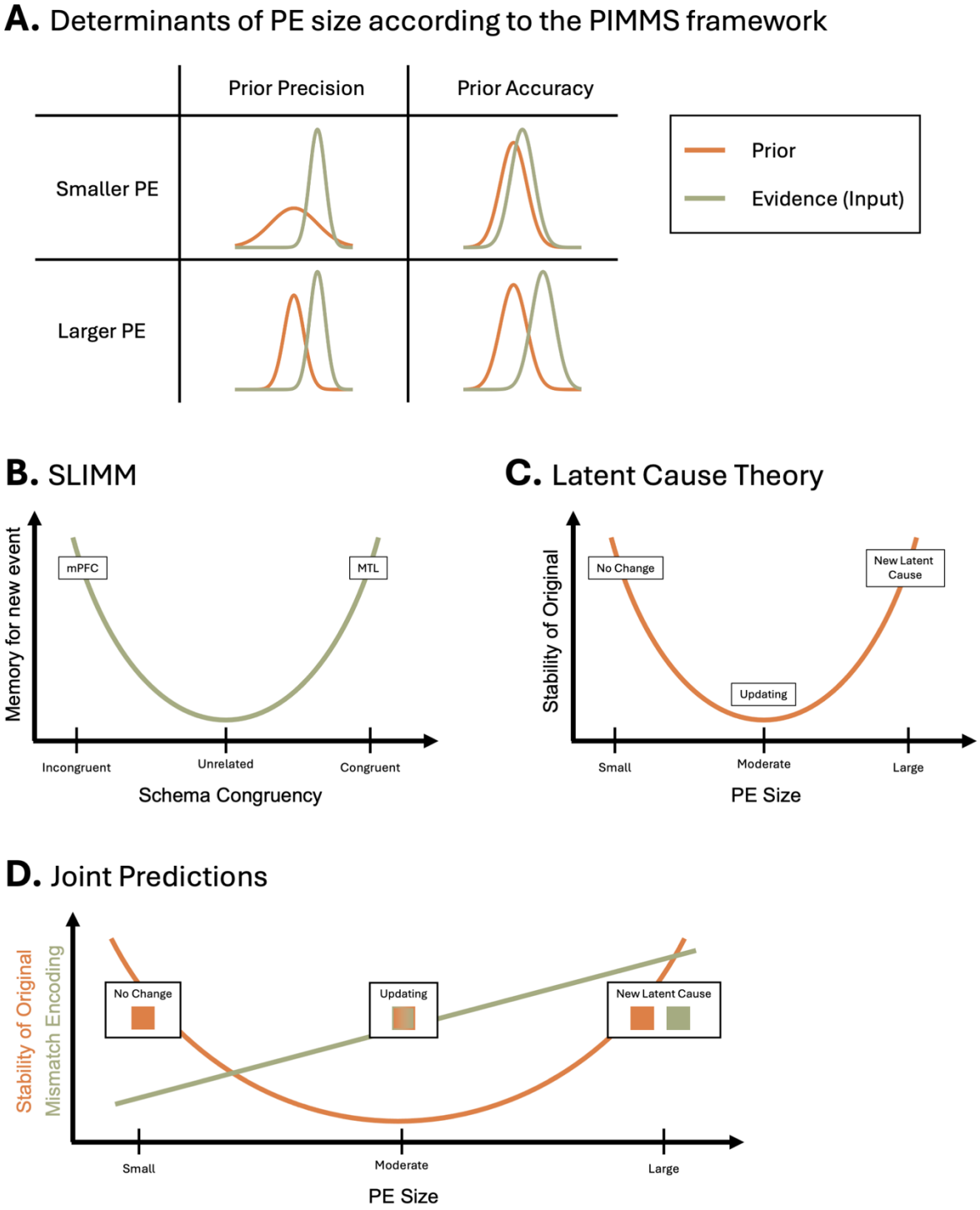
First, prior precision captures that predictions can be precise or vague. The prediction is precise if one is sure that the mailman will be at the door, but vague if one thinks that it will probably be someone from a company. Unspecific predictions cover a larger space of outcomes with some likelihood, while precise predictions are more set on one specific outcome, leaving all other outcomes more unpredicted. Therefore, PEs are generally larger after a highly precise prediction than after a vague one.

Second, prior accuracy refers to the distance between prediction and evidence, so how correct the prediction was. That is similar to the example above: If the mailman is predicted, a fundraiser is less unpredicted than the pope, eliciting a smaller PE.

A third component is evidence precision, referring to how precise the input is. For instance, if an image is blurred, it will lead to a smaller PE, due to the lower reliability evoked by the perceptual ambiguity. However, with one notable exception (Greve et al., 2017), this component has rarely been used in the field and is not employed in the studies presented here either.

Figure 2

Theoretical frameworks of PE size and its consequences



Note. **A.** The PIMMS framework expects smaller PEs when the overlap between prior (i.e., the prediction) and evidence (i.e., the input) is bigger (upper row). For prior precision, this occurs when the prediction is less precise and specific, covering a larger space of possible outcomes. For prior accuracy, smaller PEs are produced by the input being more similar to the prediction. Larger PEs (bottom row) are produced by more specific predictions and bigger differences between

prediction and input. **B.** SLIMM provides hypotheses about memory formation depending on schema congruency. Highly schema-congruent events are assimilated by existing knowledge, driven by the mPFC. Highly incongruent events lead to a pattern-separated encoding, driven by the MTL. Schema-unrelated events promote the lowest memory performance. **C.** Latent Cause Theory (LCT) provides hypotheses about how intact the original association remains after PEs of a given size. Small PEs trigger no change, large PEs trigger the inference of a new latent cause, leading to the encoding of the unpredicted information, but as a new and separate event. Moderate PEs lead to updating of the trace, as new information is integrated into the existing memory. **D.** By combining LCT and SLIMM, one can make predictions about both original and mismatch memory. Small PEs should evoke no change in the original and no new encoding as a separate episode. Moderate PEs should lead to integration of the new information, meaning that more of this new event is being encoded, while the original changes its representation due to the updating. Large PEs should foster a stable original, while the new event is being separately encoded.

Therefore, PEs will become larger with increasing prior precision and decreasing prior accuracy and are at their largest when very specific predictions are highly incorrect. One important additional conclusion from PIMMS is that if no particular prediction is made (i.e., if the prior is essentially flat), there will be only a minimal PE with virtually no impact, no matter the outcome (Henson & Gagnepain, 2010). As will be argued in section 4.5.5, flat priors may be the reason why some experiments do not find PE-based memory effects.

Before applying the notion of PE size to memory, an important caveat should be made. PE size is not an absolute measure, and it does not have standardized units of measurements. Instead, it varies with its context: Depending on predictability of the environment and variance in outcomes, the same event can evoke different PE sizes (Bunzeck et al., 2010; Rouhani et al., 2018). When speaking of small, medium, and large PEs in the upcoming sections, these are relative labels within the context of a given experiment or situation, containing some degree of interpretation.

1.4.2 Theories About the Effects of Prediction Error Size on Memory

While PIMMS (Henson & Gagnepain, 2010) spells out which factors shape PE size, there are two theoretical frameworks targeting the effects of PE size on memory: Schema-Linked Interaction Between Medial Prefrontal and Medial Temporal Regions (SLIMM) by van Kesteren et al. (2012) and Latent Cause Theory (LCT) by Gershman et al. (2017). While SLIMM makes predictions about new encoding of the mismatch target, LCT focuses on stability or change in the memory of the original target.

Together, these two frameworks inform the theoretical assumptions that the three studies of this thesis are based on.

SLIMM postulates a U-shaped relationship between schema congruency and new encoding of mismatch targets (Figure 2B). Better memory is expected for highly schema-congruent events (e.g., seeing a rubber duck in the bathroom), and for schema-incongruent events (e.g., a rubber duck in the bakery). Memory for schema-unrelated events, however (e.g., a rubber duck in the toy shop; examples from van Kesteren et al., 2012), should be weakest. These effects rests on an interplay between the mPFC and the MTL. SLIMM states that when the mPFC detects congruence (called resonance in the framework) with existing schemas, it downregulates the MTL and supports neocortical learning. When congruence is low, the MTL encodes the episode without regulation by the mPFC, producing pattern-separated episodes that are not integrated into the schema. Therefore, unique episodes are only encoded after low schema congruency. The behavioural predictions of SLIMM have been supported in experiments using experimentally established abstract rules (Greve et al., 2019) and in a high-immersion virtual reality using pre-existing schemas (Quent et al., 2022).

Schema congruency is not synonymous with PE size. The left arm of the U, where events are either congruent or unrelated to a schema, would cause a small PE, either because the schema-based predictions were correct, or because no predictions were made. PE size then gradually increases on the right arm of the U as schema congruency decreases. Therefore, SLIMM can be understood to predict a positive and linear relationship between PE size and new MTL-driven encoding of a mismatch target (see green line in Figure 2D).

A different U-shape (Figure 2C) is predicted by LCT (Gershman et al., 2017). This theory is based on computational modelling using animal conditioning studies, where associations between cues and rewards or punishments are learned. LCT focuses on the impact of PEs on pre-existing associations, specifically the extent to which they are updated after PEs. After small PEs, no updating[6] is needed, as the existing model represents the environment adequately. After large PEs, the original association also remains unaltered and a new association, representing the unexpected

[6] Updating within LCT refers to the addition of new information to an existing memory. In the Introduction and Discussion of this thesis, this is how this term will be used – a change in a pre-existing memory trace to better reflect the environment. However, memory updating is often used in the literature as an umbrella term for any changes to memory, sometimes also referring to original weakening or even new encoding.

outcome, is encoded separately. This process, coined new latent cause inference, builds on the assumption that the environment has changed fundamentally and the outcome must have been triggered by changed underlying circumstances. Both the old and the new association remain valid, but in different contexts. The highest amount of memory updating, finally, can be observed after moderate PEs. In these cases, outcomes are not unpredicted enough to assume that the underlying latent cause has changed, but the old model needs to be updated to represent the environment better. New information is therefore integrated into an existing model.

LCT, as a computational theory based on conditioning studies in animals, does not commit to a specific biological implementation. While making claims about the encoding of a new trace after large PEs, it also does not explicitly specify how memory for new events can be supported after small PEs. Instead, its main focus is on the original memory trace and how it is altered by PEs. SLIMM, on the other hand, incorporates neurofunctional hypotheses and focuses on the encoding of new events but does not offer a perspective on how pre-existing associations can be altered by PEs.

Crucially for the present purposes, neither LCT nor SLIMM are focused on a situation where predictions are based on episodic memories. In fact, with the exception of a framework by Wahlheim and Zacks (2025), which does not take PE size into account (see sections 1.4.4 and 4.4.2), no theory about the consequences of episodic PEs exists. In this thesis, LCT and SLIMM will therefore be combined and applied to episodic memory, offering a framework from which testable hypotheses about the effect of episodic PE size on memory can be derived (Figure 2D). Specifically, after small PEs, original memory should be robust, with low mismatch target encoding. Mismatch recognition memory can still be high – but this would be due to the use of a highly similar original trace, not due to the encoding of a separate memory. After moderate PEs, original memory should be low, undergoing maximal change and updating, impeding accurate recognition of the original episode. Mismatch memory would be slightly increased for the same reason. After large PEs, both original and mismatch targets should be well-remembered, as the original trace is intact, alongside a newly formed pattern-separated mismatch target memory.

1.4.3 Empirical Findings Concerning Prediction Error Size and Memory

The idea that memory consequences are determined by PE size has rarely been tested before in the field of episodic PEs. This reflects the tacit assumption that all PE

are equal. Nevertheless, some studies have produced evidence that suggests size-based memory effects, but often without explicitly employing a PE size manipulation or measure. As almost all prior studies separately manipulated either prior precision or prior accuracy in a given experiment, results are organized within each of these determinants of PE size in the following.

Turning first to prior precision, there are several ways in which previous studies have manipulated the strength with which a prediction can be made. A basic approach is varying the encoding conditions of original episodes, reasoning that better encoding should enable a more precise and confident prediction – and therefore bigger PEs when the prediction is incorrect. Accordingly, higher encoding frequency (Siestrup et al., 2023) and more elaborate encoding (Jainta et al., 2022) have led to fewer source confusions between the original and mismatch, as well as better learning of the mismatch target (Greve et al., 2017). Even without a specific manipulation, when originals episodes were remembered compared to forgotten (allowing stronger prediction), mismatch memory was increased (Bein et al., 2021). Similarly, more stable associations between cues and original targets (Greve et al., 2017) led to better encoding of the mismatch targets after PEs. These results are consistent with a separate encoding of the mismatch target after larger PEs, as predicted by LCT and SLIMM.

If predictions are episodic, a stronger reinstatement of the original target before the PE is synonymous with a stronger prediction. Consequently, when participants are reminded of the original episode or asked to actively retrieve it before PE induction, mismatch target learning is increased, as the PE is bigger (Stawarczyk et al., 2020; Wahlheim et al., 2019). A more organic approach not relying on an experimental manipulation (also used in Studies 2 and 3) is to leverage RSA and similar techniques. These techniques can measure how strongly the original target is being reinstated, and therefore predicted, spontaneously during the cue before the PE induction. Most notably, G. Kim and colleagues (2014, 2017) found that stronger original reinstatement led to more weakening of the original memory, and more differentiation between original and mismatch targets (see section 4.3.4).

Lastly, in conditioned predictions, prior precision can be manipulated by using different outcome contingencies (e.g., a reward following a certain cue 10%, 50%, or 90% of the time). Interestingly, and at odds with theoretical accounts, one such study found better memory after moderate PEs (Ortiz-Tudela et al., 2023). The authors

explain this with more information seeking behaviour when uncertainty is high (see section 4.5.5 for a potential alternative explanation). A different study relying instead on participant ratings of outcome certainty found better memory following larger positive PEs (Loock et al., 2025), which is in line with the hypotheses depicted in Figure 2D.

Turning next to prior accuracy, this component of PE size has been shown to affect memory in studies using factual information (i.e., semantic PEs). While one study found a linear effect with more belief updating when predictions were more inaccurate (Vlasceanu et al., 2020), two studies have also found a U-shaped relationship between PE size and memory (Brod et al., 2022; de Bruïne et al., 2025). In these cases, both highly expected and highly unexpected facts were remembered better. Additionally, more inaccurate explicit reward predictions have also led to better episodic memory (Rouhani et al., 2018). Again, these results are consistent with LCT and SLIMM, indicating better memory after larger PEs.

Only one previous study accounted for both prior precision and prior accuracy together. Long et al. (2016) used a continuous neural measure of category prediction and manipulated prior accuracy by using mismatch targets either from the same category as the original target (small deviation) or a different category (large deviation). Results indicated that especially after small deviations, a stronger prediction led to stronger hippocampal activity, taken as a measure of mismatch detection. While this is interpreted as the hippocampus reflecting the need for disambiguation of the similar original and mismatch targets, it is possible that the hippocampal activity also reflects an integration of the mismatch target into the existing model (Kalbe & Schwabe, 2021). High prior accuracy, coupled with high prior precision could be speculated to lead to moderate PEs, triggering updating, while larger deviations lead to larger PEs where new information is encoded separately.

To summarize, there is first evidence that both prior precision and prior accuracy have a PE size-dependent influence on both original and mismatch memory. However, few studies have targeted both prior precision and prior accuracy, and tested memory for both original and mismatch targets. Crucially, no study of episodic PEs has employed a continuous manipulation of PE size. While piecing together bits of evidence from different studies can give indications about the full mnemonic outcomes of episodic PEs of different sizes, this approach is ultimately insufficient. Filling this gap was therefore an overall aim of the presented studies.

1.4.4 Representational Effects of Prediction Errors

Another shortcoming of prior work is that it has selectively focused on the question of what is remembered after PEs, often leaving it open how this is achieved, so how memory representations change in the process. This is relevant as the two memory traces in question, the original and the mismatch targets, will often share substantial similarity: They are both associated with the same cue and, in studies using naturalistic material, are also semantically related. This similarity could be addressed by using either integration or separation, with distinct behavioural consequences (see section 1.1.3). There have been arguments for both representational outcomes being triggered by PEs, but without taking PE size into consideration.

PEs have often been assumed to generally trigger pattern separation, due to their involvement in event segmentation. This is the process of dividing the continuous input of experience into meaningful sub-units, or events (Zacks & Swallow, 2007). One possibility is that event boundaries are induced once PEs accumulate, indicating that the current event model is not suited for explaining the situation and needs replacing (Nolden et al., 2024; Zacks & Swallow, 2007). Even though recent evidence has questioned whether single PEs are sufficient to induce event boundaries (Güler et al., 2025), this backdrop has led to the assumption that PEs generally induce separation in episodic memory (Ben-Yakov et al., 2022; Rouhani et al., 2020, 2023). However, it should be stressed that event segmentation refers to a separation between adjacent elements of the stream of input, so for instance between cue and target (see Study 3). It is therefore not necessarily evidence for a separation between a previously encoded original target and a newly perceived mismatch target – which is a distinction unique for episodic PEs (see Study 2).

Nevertheless, there is further evidence supporting separation from studies using semantic (Frank et al., 2020) and episodic (Bein et al., 2021) PEs. These studies found that PEs promote detailed memories, which were less likely to be confused with perceptually similar lures – a behavioural correlate of pattern separation. Additionally, stronger PEs (measured as prior precision using original reinstatement) have been shown to lead to more differentiation between the cue and the original target upon restudy of the original target in isolation (G. Kim et al., 2017).

However, for the latter study, an alternative explanation has been proposed and supported with computational modelling: G. Kim et al.'s assumption that the cue and original target are differentiated rests on decreased representational similarity after

the PE. Greve et al. (2018) have argued that this decreased similarity can also be explained by the integration of the mismatch target into the original representation. This would likewise reduce representational similarity, but not because of differentiation, but due to integration.

The Event Memory Retrieval and Comparison Theory (EMRC) by Wahlheim and Zacks (2025), which is the first and hitherto only framework for mnemonic consequences of explicitly episodic PEs, argues that PEs trigger integration. They propose that consciously noticed PEs trigger the reactivation of the original episode and that this will lead to the encoding of a recursive representation. This representation includes both original and mismatch targets, their temporal sequence, and the fact that a PE was experienced (Wahlheim & Zacks, 2025). The researchers' own studies serve as evidence, where participants are tested for mismatch target memory first, and then asked whether they remember any change in this item. If they do, they are also asked to recall the original target. Recollecting that there was a change and remembering the original are associated with higher mismatch memory, which is interpreted as evidence for integration (Hermann et al., 2021; Stawarczyk et al., 2020; Wahlheim et al., 2019, 2021; Wahlheim & Zacks, 2019). A major problem in this framework (see section 4.4.2 for a more extensive discussion) is that all PEs are thought to be equal, leaving no moderating role for PE size. In this framework, all PEs, as long as they are consciously detected, will trigger integration.

Despite some studies finding integration and others finding separation after PEs, there have been few attempts to reconcile these findings. However, it is possible that both outcomes can occur after PEs. Importantly, LCT makes predictions about representations dependent on PE size. Moderate PEs are thought to trigger integration, as an existing memory is supplemented with new information. Large PEs trigger pattern separation, as the original memory is preserved, and a new memory (reflecting a different latent cause) is encoded independently. A direct test of these size-dependent representational outcomes within episodic PEs had been lacking and is carried out in Studies 2 and 3.

1.4.5 Beyond Size: Prediction Error Types

While the preceding sections showed how PE size may affect memory performance and neural representations, PEs do not only differ in size. As the distinction between conditioned, semantic, and episodic predictions shows, there are

qualitatively different types of PEs as well (Bubic et al., 2010). Several previous studies have shown a further qualitative differentiation within the class of episodic PEs. Siestrup et al. (2022) used toy figurine-based action sequences to induce structural PEs (swapping the order of two consecutive steps) and content PEs (replacing an object). Structural changes caused activation in areas associated with the temporal organization of episodes and goal-directed action (e.g., superior temporal sulcus and precuneus), while surface changes activated regions relevant for object characteristics in the context of actions (e.g., the intraparietal sulcus and occipitotemporal cortex). In addition to this qualitative difference in brain processing, it was mainly structural changes that led to impaired source memory (Siestrup et al., 2022).

In a follow-up study, Siestrup & Schubotz (2023) always replaced one object with another, but this change could either leave the storyline intact (surface modifications) or alter it (gist modifications). As before, there were qualitative differences in brain activation, with gist modifications again eliciting stronger activity in the superior temporal sulcus and precuneus. Interestingly, the surface changes were more likely to negatively affect source memory. The authors explain this with gist changes leading to the encoding of a new and separate memory trace, without the danger of confusions with the original episode (Siestrup & Schubotz, 2023). This explanation is akin to the inference of a new underlying latent cause and therefore distinct encoding of the new information.

Beyond these studies, little work has been conducted on the influence of PE types on memory outcomes. It is possible that surface-level changes, which may be predicted only at lower levels of the predicting hierarchy (Barron et al., 2020; Clark, 2014), have not been deemed relevant for episodic memory, given that the hippocampus is placed at the top of the hierarchy (Barron et al., 2020; Reichardt et al., 2020). Furthermore, especially in language, memory for the surface form of episodes may be rather weak (Brown-Schmidt & Benjamin, 2018; Reyna et al., 2016), as memories may become more semanticized and gist-like as they get older (Reagh & Yassa, 2014; Yassa & Reagh, 2013). Therefore, it was a subsidiary goal in Study 1 to test for differences in memory outcomes between surface and gist PEs, given that little is known about how PE type affects episodic memory.

2 Research Questions & Design

So far, I have argued that episodic memory flexibly represents a dynamic world and constructively enables forming past and future scenarios. Memory is leveraged by the brain to predict upcoming input and resulting PEs are a prime driver of memory change. However, the precise circumstances and mechanisms of memory outcomes of episodic PEs are still not sufficiently understood, due to a widespread tacit assumption that all PEs are the same. While theoretical frameworks suggest that PE size may be the primary driver in moderating between outcomes (Gershman et al., 2017; Henson & Gagnepain, 2010), this remains largely untested in episodic PEs, as no investigation had continuously manipulated PE size and tested for effects on both original and mismatch memory. This gap in the literature has impeded the formulation of frameworks specifically for episodic PEs and may have hindered a more widespread connection of episodic PEs to applicational problems.

To address these problems, three studies were carried out to investigate the role of PE size in episodic memory outcomes. They were designed to investigate three overarching research questions (RQs).

RQ1: How is recognition memory of the original and mismatch target affected by PEs of different sizes?
RQ2: What are the representational consequences of PEs of different sizes, and how do they relate to behavioural outcomes?
RQ3: What are the neural correlates of memory updating compared to new latent cause inference?

To address these questions, building on strengths as well as shortcomings of prior studies, the study design had to fulfil the following four design criteria:

Criterion 1: Equal focus should be put on original and mismatch target memory, with independent memory tests of both.
Criterion 2: PE size should be manipulated as continuously as possible, targeting both prior precision and prior accuracy.
Criterion 3: The employed material should have high ecological validity, inherent relevance for participants, and be temporally unfolding to allow naturalistic

prediction.

Criterion 4: Predicting or change detection should not be the explicit task of participants, as to not interfere with naturalistic prediction.

Based on these RQs and methodological criteria, an overall design was developed and adapted for all studies. A canon of naturalistic dialogues was written, professionally recorded by 20 voice actors, and paired with unique background sounds. Each dialogue unfolded between two speakers and focused on situations common in the life of young adults. Participants listened to these dialogues in an original form multiple times to allow solid original encoding. At a later exposure, and without warning, the dialogues were modified: One statement in the middle of the dialogue was replaced by another, constituting a substitution of the original target by the mismatch target. This should induce a PE: while the original target was predicted, the input deviated from this. Importantly, by using different versions, with changes that were smaller or larger, we could directly manipulate PE size. Later, in a separate session, participants' original and mismatch target memory was tested.

The three presented studies emerge from independent data collections and utilize behavioural, univariate, and multivariate fMRI analyses. Study 1 was the first test of size-dependent episodic PE outcomes, where PE size was manipulated via encoding frequency (prior precision; Bein, Duncan, et al., 2020) and measured as the perceived difference between original and mismatch dialogue versions (prior accuracy). This was done in a behavioural design, focusing on memory outcomes of PEs of different sizes (RQ1). We expected a U-shaped relationship between PE size and original recognition, and a linear relationship between PE size and mismatch encoding (Figure 2D).

Study 2 employed fMRI to measure prior precision continuously as the neural reinstatement of the original target before PE induction. Prior accuracy was measured via difference ratings as in Study 1, but with ratings obtained in an independent sample. In addition to testing the behavioural effects of PE size (RQ1), the use of fMRI also allowed us to assess representational outcomes using RSA (RQ2). Specifically, it was hypothesized that original representations would be less altered after larger PEs, due to less updating of the original trace. Furthermore, a measure of source memory was introduced in Study 2, and higher source memory was expected after larger PEs, where a new memory trace is separately encoded.

While Studies 1 and 2 created PEs across the whole continuum, Study 3 zoomed in on moderate to large PEs. Specifically, it investigated brain signatures of updating an existing memory, compared to inferring a new latent cause and encoding a new memory trace separately (RQ3). According to LCT, and resonating with hypotheses from SLIMM about schema-congruent events, moderate PEs should be accompanied by higher integration-related frontal activation (Bowman & Zeithamova, 2018; Ferstl & von Cramon, 2001; Spalding et al., 2018) as new information can be integrated into an existing model.

Taken together, these three studies therefore were the first comprehensive research about how episodic PE size shapes memory outcomes (RQ1) and how this is reflected in memory representations (RQ2) and brain activity (RQ3). The results also laid the foundation for a more thorough understanding of episodic PEs, which will be used in section 4.5 to develop an overall framework of how episodic PEs affect memory.

3 Original Papers

3.1 Study 1: Prediction errors lead to updating of memories for conversations

2025

Boeltzig, M., Liedtke, N., Schubotz, R.I.

Memory, 33(1), 73-83

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

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

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.
© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

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

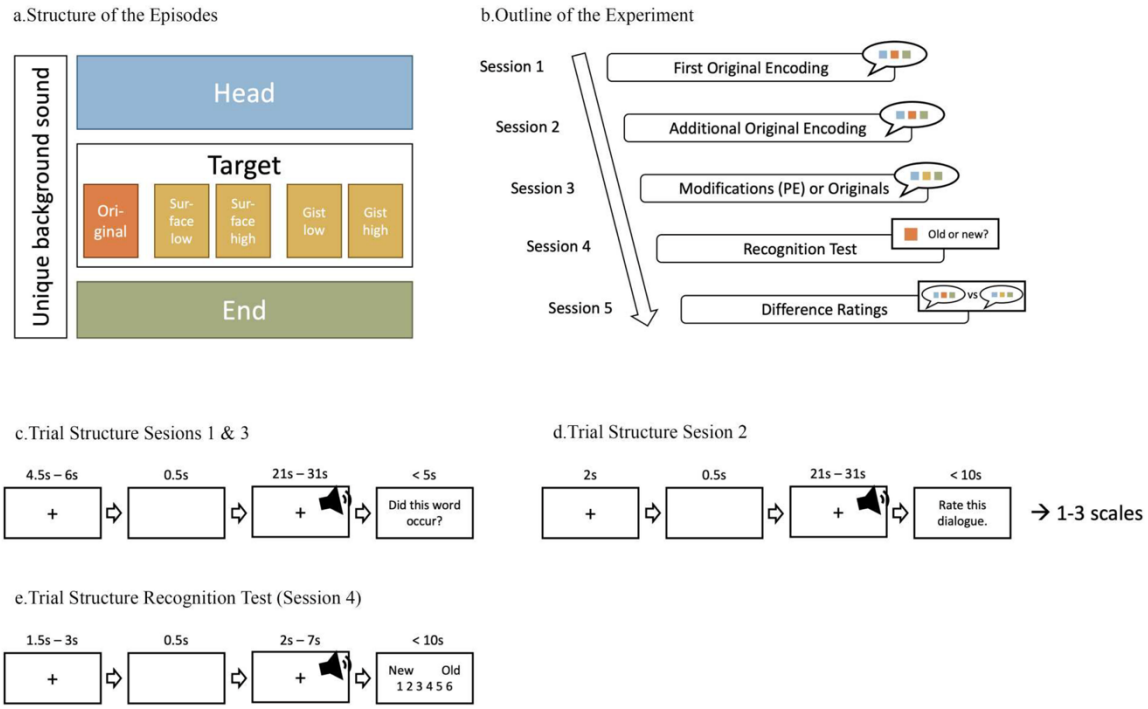


Figure 1. Method of the experiment. (a) The stimuli were dialogues covertly consisting of heads, targets, and ends together with unique background sounds. All dialogues were first presented with the original target and some dialogues were later modified by using one of the four alternative targets, which 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), *p*s < .001, *d*s > 0.84, and new probes were rejected with above-chance accuracy (lures: M = .78, SD = .07; novels: M = .99, SD = .01), *p*s < .001, *d*s > 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), *p*s < .001, *d*s > 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: $p = .182$).

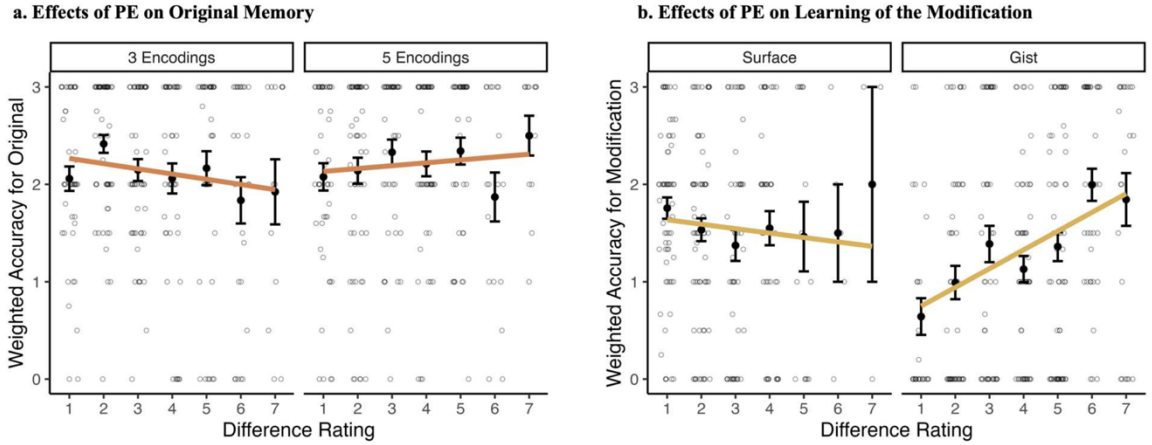


Figure 2. Memory outcomes for original and modification. (a) Weighted accuracy ratings for the original version of the statement (as encoded in Sessions 1 and 2) are plotted against difference rating (i.e., prior accuracy) and encoding frequency (i.e., prior precision). Bold dots represent means for 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

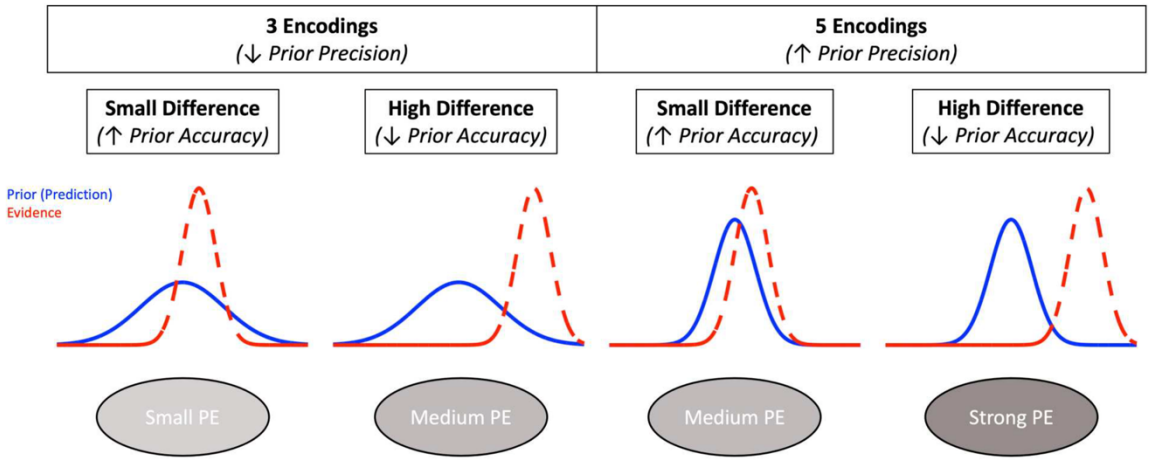


Figure 3. Explanation of the results within the PIMMS framework. Interpretation of our results in light of the PIMMS framework (Henson & Gagnepain, 2010). The prior is generally less precise, meaning that expectations are broader after three than after five encodings (hence the flatter blue line on the 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

3.2 Study 2: The benefit of being wrong: How prediction error size guides the reshaping of episodic memories

2025
Boeltzig, M., Liedtke, N., Siestrup, S., Mecklenbrauck, F., Wurm, M.F., Bramão, I., Schubotz, R.I.
NeuroImage, 317, Article 121375

NeuroImage 317 (2025) 121375

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
1053-8119/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

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 pseudo-randomized. 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

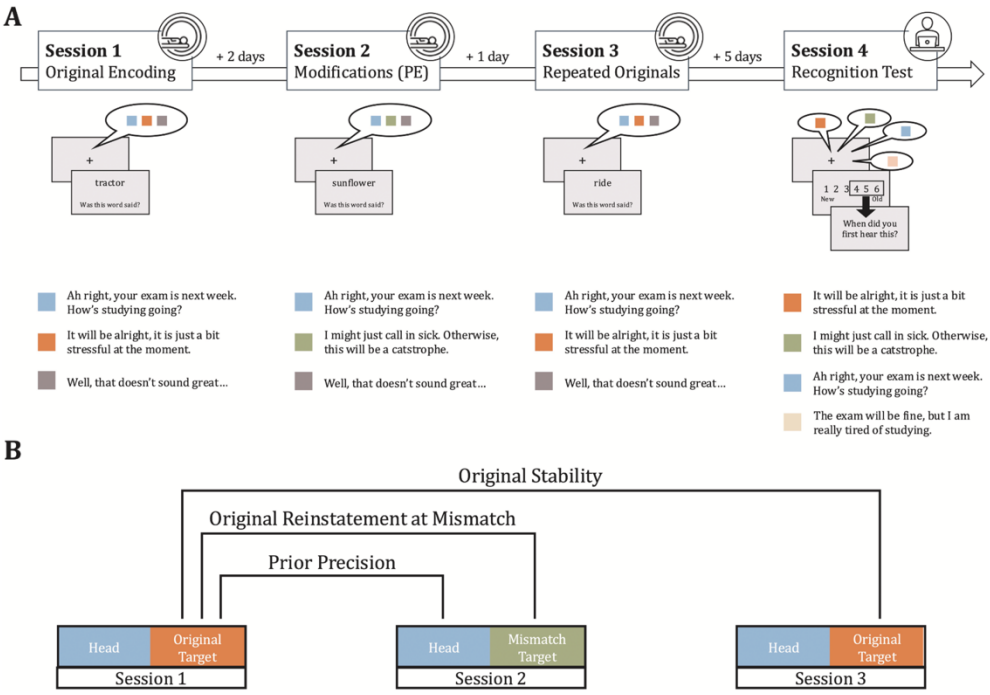


Fig. 1. Experimental Procedure & Representational Similarities of Interest.
Note. **A.** The experiment took place in four sessions. In Session 1, participants were made familiar with the original versions of the dialogues, played once in the scanner, and another time outside of the scanner. In Session 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

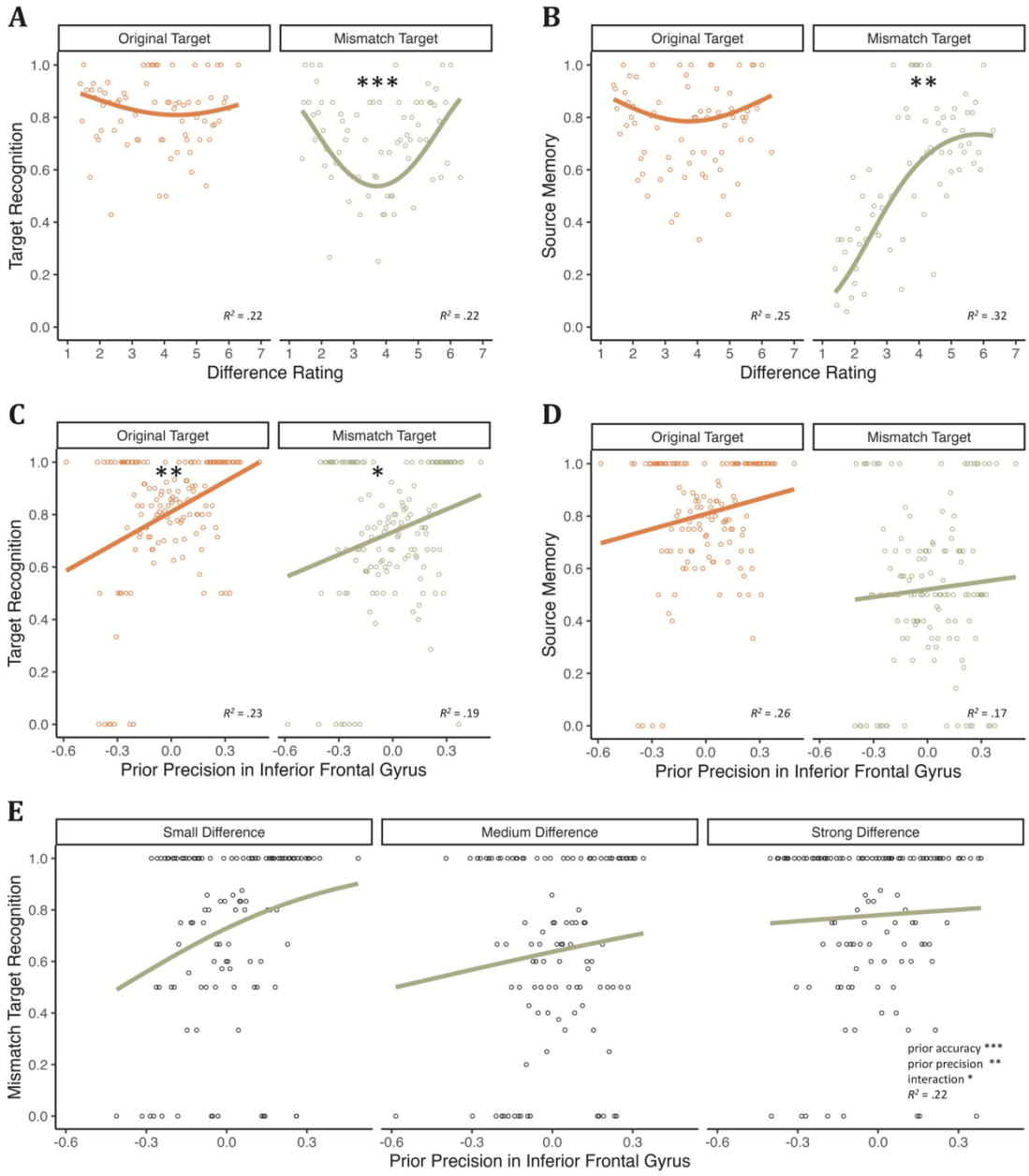


Fig. 2. Influence of Prediction Error Size on Memory Outcomes.

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

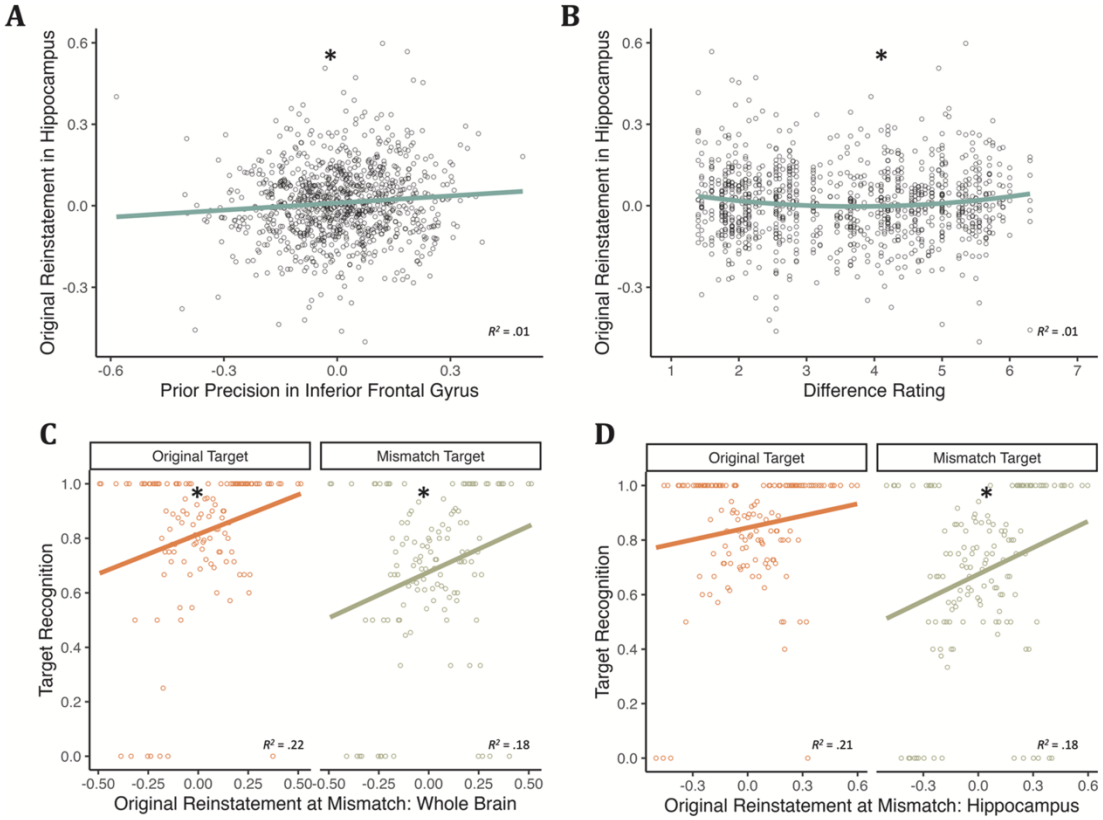


Fig. 3. Predictors and Consequences of Original Reinstatement at Mismatch.
Note. The first two plots show how prediction error size influences how much the original target is reinstated while the mismatch target is played in Session 2. Higher values indicate stronger original reinstatement. Each dot 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.

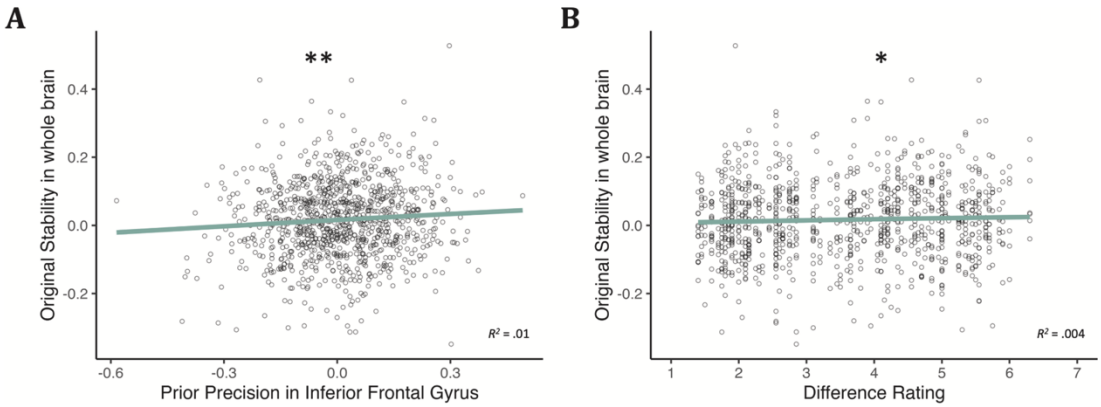


Fig. 4. Influence of Prediction Error Size on the Stability of the Original.
Note. Original stability was measured through representational similarity of the original targets before and after the prediction error. Higher values indicate higher similarity, and therefore less change. For both components 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

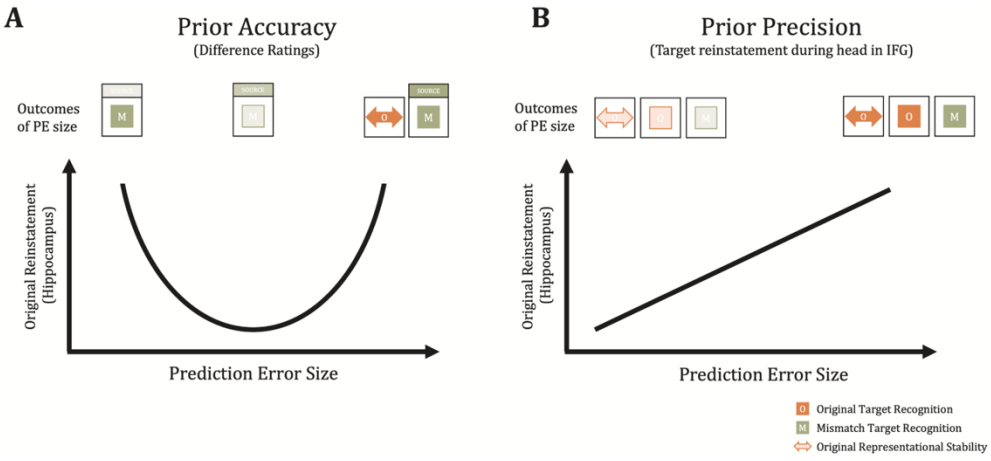


Fig. 5. Overview of Results.
Note. This schematic overview summarises the main findings of the study. It plots reactivation of the original in the hippocampus against prediction error size separately for both prior accuracy (A) and prior precision (B). Additionally, the outcomes of PE sizes are shown 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., Reße, 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.

3.3 Study 3: To update or to separate: Neural signatures and consequences of latent cause inference in episodic memory

2026

Boeltzig, M., Siestrup, S., Bramão, I., Schubotz, R.I.

NeuroImage (under review in revised form)

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

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

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

Author Note

Marius Boeltzig https://orcid.org/0009-0007-8930-0171, marius.boeltzig@uni-muenster.de
Sophie Siestrup https://orcid.org/0000-0002-6365-3055, s.siestrup@uni-muenster.de
Inês Bramão https://orcid.org/0000-0003-2635-2381, ines.bramao@psy.lu.se
Ricarda I. Schubotz https://orcid.org/0000-0001-5802-6869, rschubotz@uni-muenster.de

Correspondence concerning this article should be addressed to Marius Boeltzig, Department of Psychology, Fliednerstraße 21, 48149 Münster, Germany, marius.boeltzig@uni-muenster.de

Abstract

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

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

Highlights

- Prediction errors can cause episodic memory updating or new latent cause inference
- Updating and latent cause inference differ in neural signatures
- Frontomedian BA9 and vmPFC support the updating of memories with new information
- New latent cause inference produces representational separation
- Prediction precision biases towards top-down reinstatement or new evidence sampling

1. Introduction

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

According to Latent Cause Theory (Gershman et al., 2017), the exact impact of PEs on memory depends on PE size. Specifically, moderate PEs lead to integration of the unpredicted event into the current model (which can be fed by episodic memories). Larger PEs, in contrast, lead to the inference of a new latent cause: Instead of integrating new information into the existing model, they prompt a switch to a different model capable of explaining the unexpected event better. If no such alternative model is available, a new one must be formed. Thereby, the brain structures experience into latent causes by weighing when to integrate versus separate events (Bein et al., 2023), producing an optimally structured model of the environment. In the current study, we aimed to investigate how updating via integration after moderate episodic PEs and latent cause separation after large episodic PEs differ from each other in neural signatures as well as representational and mnemonic consequences.

Latent Cause Theory (Gershman et al., 2017) is based on modelling work of Pavlovian conditioning paradigms. In these cases, the models used for predictions are stimulus-outcome mappings, such as a conditioned relation between a tone and a shock. However, internal models are supposed to allow reliable predictions for any sensory input and thus need to reflect more complex regularities and experiences as

well (Brown & Brüne, 2012; Schubotz, 2007). Consequently, prediction has been suggested to be strongly based on memory (Reichardt et al., 2020), and both semantic (Frank et al., 2018; Greve et al., 2019; van Kesteren et al., 2012) and episodic memory (Kim et al., 2014, 2017; Long et al., 2016; Varga et al., 2025; Wahlheim et al., 2021) has been demonstrated to be used for predicting. Basing predictions on episodic memories may be the most effective solution in highly repetitive situations, in unstable contexts where no regularities can be extracted (Kim et al., 2019), or in novel situations where no schema has been formed yet.

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

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

consistent with the inference of a new underlying latent cause.

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

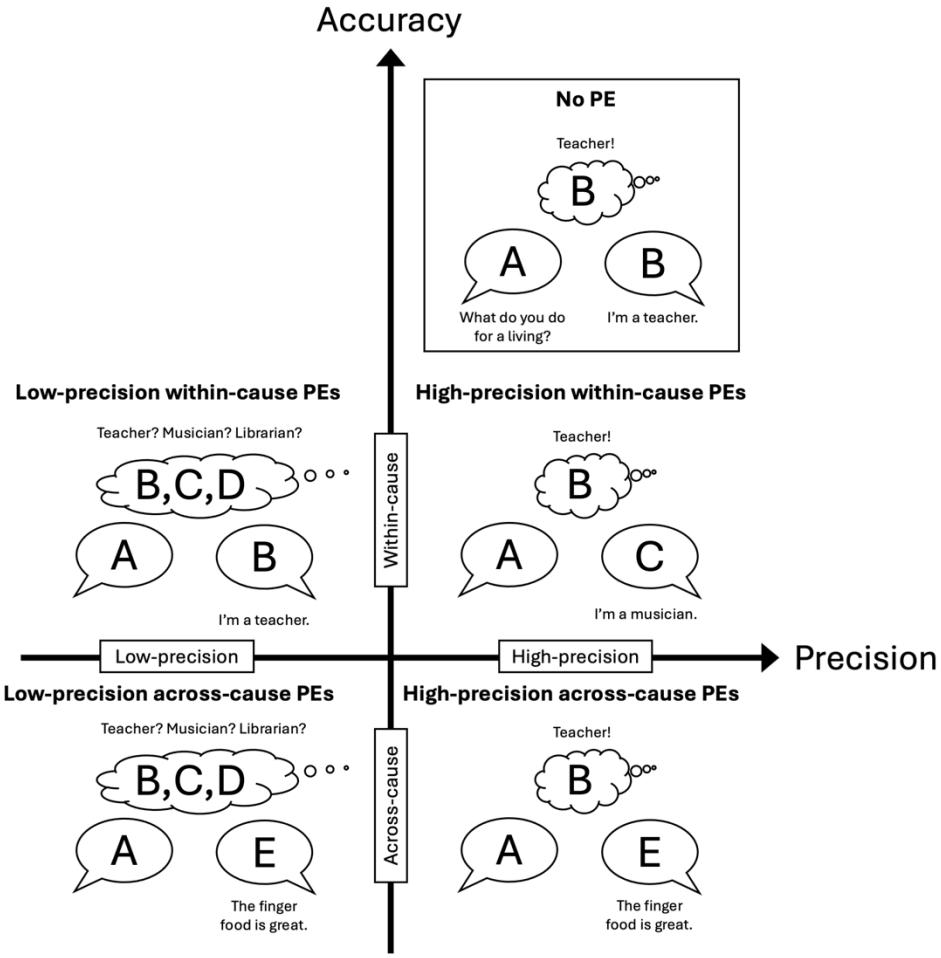
Latent Cause Theory focuses on the *accuracy* of the prediction as the key determinant of PE size, and thereby of whether the current model is updated or a new latent cause is inferred. However, predictive coding accounts emphasize that both accuracy and the *precision* of the prediction jointly determine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

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

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

Figure 1

Schematic depiction of the PEs under study



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

within-cause PE. However, if the response was that the finger food is excellent, this would not have been even vaguely predicted: the person might have changed topics or talked to somebody else. This *low-precision across-cause PE* would trigger new latent

cause inference, and thus shift models (e.g., from professions to food). To examine new latent cause inference after low-precision predictions, we contrast these smaller within-cause (stemming from the same predictive model) and larger across-cause (inducing a change in models) PEs with each other in the analysis.

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

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

Turning to the consequences of the PEs, we predicted better memory for novel information, as well as less original forgetting after larger PEs, so after across-cause compared to within-cause PEs (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025; Liedtke et al., 2025). Additionally, the latent cause separation triggered by low-precision across-cause PEs should lead to the parts of these stimuli being represented more separately, as they are deemed less connected and therefore

less relevant to each other, compared to low-precision within-cause PEs. This is akin to an event segmentation process, where continuous experience is parsed into meaningful units (Nolden et al., 2024; Zacks & Swallow, 2007), which can be triggered by new latent cause inference (Gershman et al., 2017). Lastly, we assessed reactive reinstatement, reflecting the retrieval of previously encoded original versions of a stimulus while a new mismatching one is processed. We predicted that during high-precision across-cause PEs, such reinstatement is stronger than during high-precision within-cause PEs, reflecting more top-down activity.

2. Methods

2.1 Participants

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

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

2.2 Design

We manipulated two orthogonal dimensions to produce the four situations described in the Introduction and depicted in Figure 1. *PE precision* was varied through familiarity to the material: When a given stimulus was presented for the first time, participants could make only low-precision predictions, and when the stimulus was already familiar (at third exposure) predictions were high-precision. To induce truly high-precision predictions, dialogues were played in their original form twice to allow solid encoding. The second exposure to the dialogues was not investigated in the

comparison of neural signatures of the PEs in the manuscript, but is reported in the Supplement. *PE accuracy* was manipulated within the stimulus material, with dialogues being presented either in a coherent version, where two statements were logically related, or an incoherent version where they were not (see 2.3 Material). Additionally, there was a no PE condition with stimuli that were never altered.

2.3 Material

For this study, we adapted 60 previously used naturalistic dialogues between two speakers, discussing topics and events likely to resonate with a student sample (Boeltzig, Liedtke, & Schubotz, 2025; Boeltzig, Liedtke, Siestrup, et al., 2025; Liedtke et al., 2025). One example is provided in Figure 1A, four additional examples can be found in the Supplementary Materials S1, and the full transcribed material is available on OSF, along with one sample of the auditory stimuli that participants were exposed to (https://osf.io/mgu49/overview?view_only=1af278283f49433d8de09f566a2dff97).

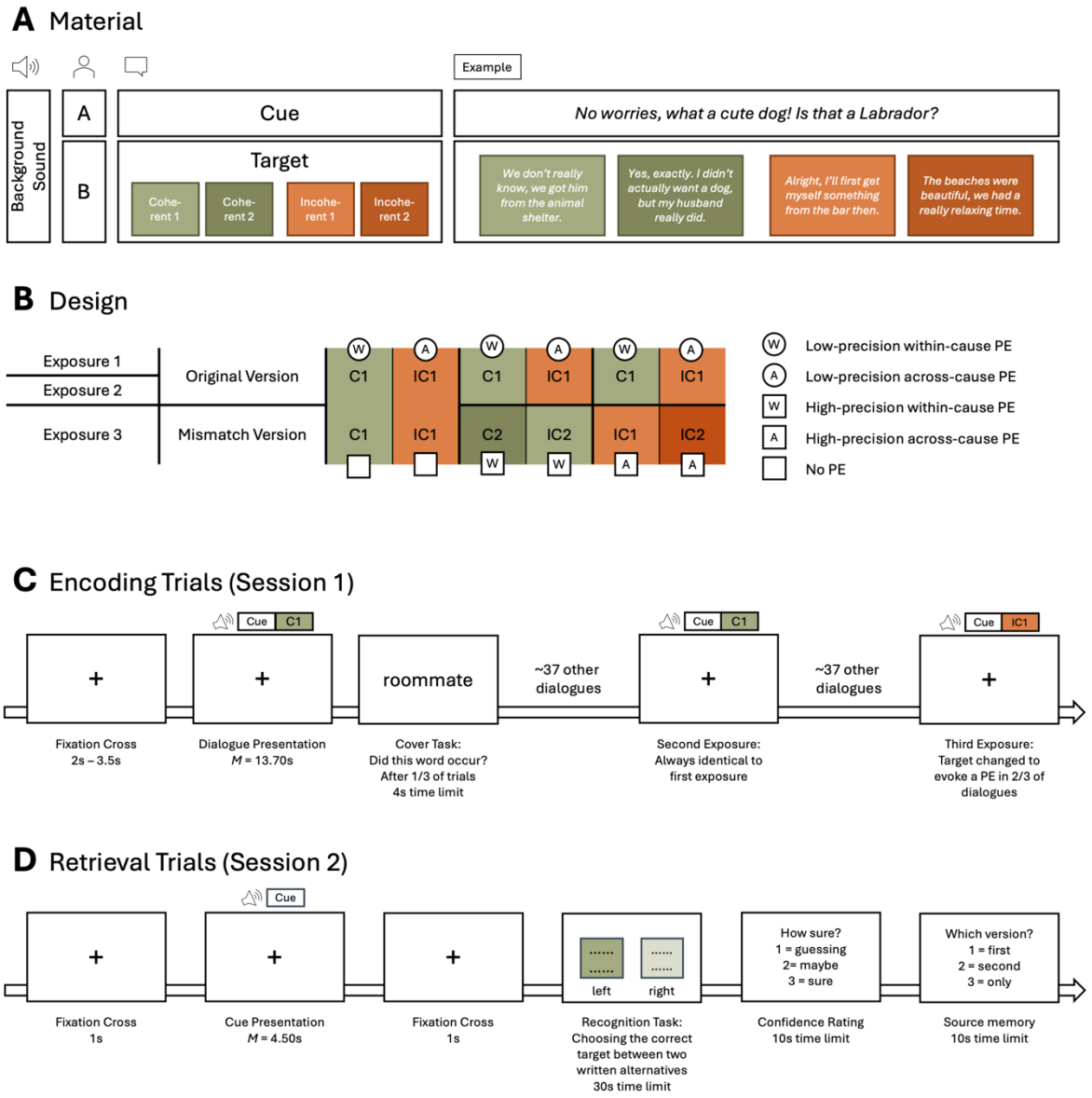
The dialogues were recorded by 20 professional speakers and paired with a unique background sound. For this study, the dialogues were shortened so that they consisted of two parts. The cue consisted of one or more sentences by one speaker, and the target comprised one or more sentences by another. Crucially, while the cue always stayed the same, the targets could be exchanged, producing four dialogue versions. All four targets matched to a given cue involved the same speaker so that a mismatch would not be produced based on a different voice.

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

One third of the dialogues was never changed during the experiment and thus played in the same version three times. The other two thirds of dialogues were presented in an original form twice, and then changed upon the third exposure, meaning that a different target was played in response to the same cue. Half of the dialogues initially had a coherent original, the other half had an incoherent original.

Figure 2

Experimental Procedure



Note. **A.** Dialogues were comprised of two statements, one by each speaker. The statement of the first speaker, the cue, was always the same, but there were four alternative targets. Two of these targets were coherent with the cue, directly responding to it. Two targets were incoherent, not directly responding to the cue, inducing new latent cause inference. In addition to the provided translated sample dialogue, there are more examples in the Supplementary Materials S1 and on OSF. **B.** There were six types of dialogue histories throughout the experiment. Dialogues were always played twice in an original version that could be either coherent or incoherent. Originally coherent dialogues were believed to induce low-precision within-cause PEs, and originally incoherent dialogues induced low-precision across-cause PEs, and therefore new latent cause inference. A third of all dialogues (two leftmost columns) were not changed, but instead played again in their original form at the third exposure, so that the original and mismatch versions were

identical. The other two thirds of dialogues were changed, again either being coherent or incoherent. Thus, at third exposure, high-precision within-cause PEs were evoked by a change to a novel but coherent dialogue version, and high-precision across-cause PEs were induced by a novel incoherent version. **C.** Encoding trials started with a jittered fixation cross, followed by the presentation of a dialogue, during which a fixation cross stayed on the screen. After one third of trials (once for each dialogue), a word was displayed and participants had to respond whether the word was said in the dialogue. Each dialogue was played three times, interleaved with a variable number of other dialogues, and two thirds of dialogues were modified upon third exposure. **D.** The memory test comprised one trial for the unchanging dialogues, and two trials for the two versions of the changing dialogues. After a fixation cross, the cue was presented auditorily. After another fixation cross, two written options were presented, and participants had to choose which one was actually said in response to the cue. The other option was a similar lure. Participants also rated confidence in their choice and indicated whether the just-probed version was the first, second, or only one they had heard of that specific dialogue.

Within the changing dialogues, half of each group then had a coherent mismatch (i.e., new version), and the other half an incoherent mismatch. There were therefore six dialogue types (Figure 2B): (1) unchanging coherent, (2) unchanging incoherent, (3) coherent original, coherent (but new) mismatch, (4) incoherent original, coherent mismatch, (5) coherent original, incoherent mismatch, (6) incoherent original, incoherent (but new) mismatch. There were ten dialogues in each of these conditions. Note that which of the two coherent or incoherent targets would be the original and which one the mismatch version was counterbalanced across participants. Assignment of dialogues to conditions was also counterbalanced.

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

2.4 Procedure

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

memory test. The sessions lasted around 75 and 45 minutes, respectively.

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

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

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

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

In each trial, after a fixation cross (2s), participants listened to the cue of the dialogue, followed by a fixation cross (1s). They then saw two written statements on the screen, one was the original target, and the other was a similar lure that had not been said. Participants chose the correct version using the arrows on the keyboard (30s

time limit) and location of the correct answer was pseudo-randomized. After their response, they indicated confidence using numbers (1 = guessing, 2 = maybe, 3 = sure; 10s time limit). Finally, as a measure of source memory, they were asked to indicate whether this dialogue was the original version which they had heard first, the second version which they only heard later, or the only version of that specific dialogue (10s time limit). This was illustrated with an example and the experimenter made sure that participants understood this task.

2.5 MRI data acquisition and preprocessing

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

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

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

2.6 Behavioural data analysis

First, we tested whether any covariates arising from the procedure influenced recognition performance. Age, gender, sleep duration, guessing the intention of the experiment, serial position of dialogues, cover task performance, test order (original vs mismatch first), semantic similarity between target and lure, and duration of the

cue did not significantly predict recognition performance (see Supplementary Materials S2) and where therefore not controlled for in further analyses.

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

2.7 Univariate MRI analysis

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

We ran two models, which were different in that only the first model always accounted for dialogue history, so the coherence at first exposure, even when modelling responses to induced PEs at the third exposure. In addition to button presses in response to the cover task, we modelled (1) low-precision within-cause PE (i.e., first exposure to coherent dialogues), (2) low-precision across-cause PE (i.e., first exposure to incoherent dialogues), (3) second exposure to coherent dialogues, (4) second exposure to incoherent dialogues, (5) high-precision within-cause PE (i.e., third exposure to a dialogue with a novel coherent target) after originals had been coherent, (6) high-precision within-cause PE (i.e., third exposure to a dialogue with a novel coherent target) after originals had been incoherent, (7) high-precision across-cause PE (i.e., third exposure to a dialogue with a novel incoherent target) after originals had been coherent, (8) high-precision across-cause PE (i.e., third exposure to a dialogue with a novel incoherent target) after originals had been incoherent, (9)

no PE (i.e., third exposure to unchanging dialogue) with coherent originals, (10) no PE (i.e., third exposure to unchanging dialogue) with incoherent originals.

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

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

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

2.8 Representational similarity analysis

In addition to the univariate analysis, we conducted Representational Similarity Analysis (RSA). To that end, we specified a GLM in which each cue and each

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

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

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

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

Third, while the previous measure reflects the predictive reinstatement of a dialogue target, we also analysed original reactivation at mismatch, reflecting a

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

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

3. Results

3.1 Manipulation check and overall performance

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

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

3.2 Processing of PEs

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

3.2.1 PEs after low-precision predictions

Firstly, we contrasted the two low-precision PEs with each other by comparing activity at the first exposure to incoherent dialogues (low-precision across-cause PEs) with activity at the first exposure to coherent dialogues (low-precision within-cause PEs). Given that the dialogues were novel at this time point, predictions were low-precision. After across-cause PEs (incoherent dialogues), new latent cause inference

should take place, while within-cause PEs should support the vague prediction and where therefore used as the baseline here.

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

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

3.2.2 High-precision across-cause PEs

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

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

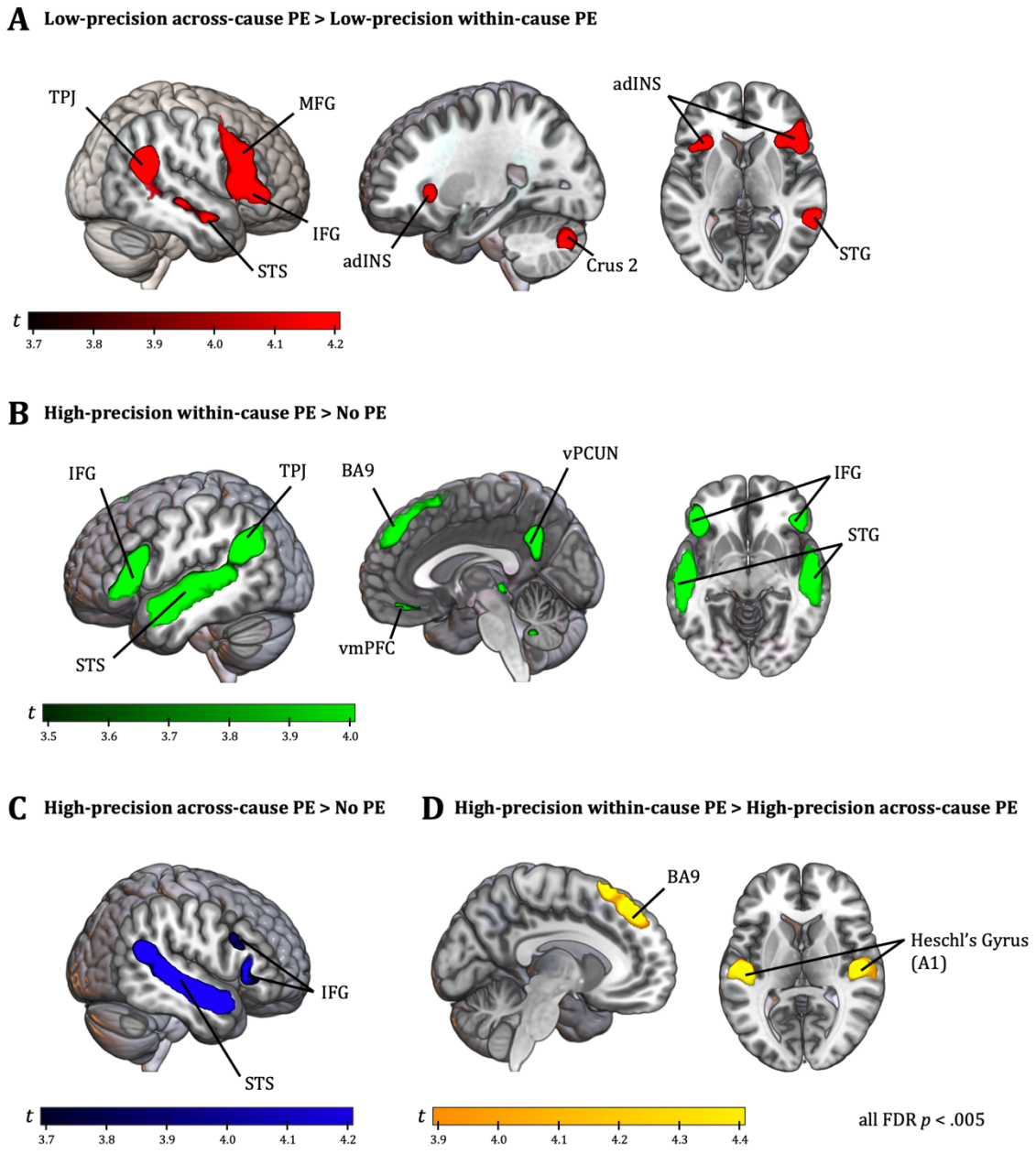
Table 1

Low-precision across-cause PE > High-precision across-cause PE

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

Figure 3

Whole-brain activation for direct contrasts of all PEs



Note. **A.** Low-precision across-cause and low-precision within-cause PEs were compared by contrasting incoherent and coherent dialogues at first exposure. **B.** To test high-precision within-cause PEs, dialogues that were modified at third exposure with a coherent target were compared to unchanging dialogues at third exposure. **C.** Dialogues that were modified at third exposure using an incoherent target were compared to unchanging dialogues at the third exposure to assess high-precision across-cause PEs. **D.** High-precision within-cause PEs and high-precision across-cause PEs were directly compared as well. The opposite contrast yielded no significant clusters. TPJ = temporoparietal junction, MFG = middle frontal gyrus, IFG = inferior frontal gyrus, STS = superior

temporal sulcus, adINS = anterior dorsal insula, STG = superior temporal gyrus, vPCUN = ventral precuneus, vmPFC = ventromedial prefrontal cortex.

Table 2

High-precision across-cause PE > No PE

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

3.2.3 High-precision within-cause PEs

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

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

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

We also compared the two PEs at third exposure after high-precision

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

3.3 Consequences of PEs

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

3.3.1 The effect of low-precision PEs on integrated representations

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

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

3.3.2 Original Reactivation during high-precision PEs

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

Table 3

High-precision within-cause PEs

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

hippocampus.

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

3.3.3 Effects of PEs on memory

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

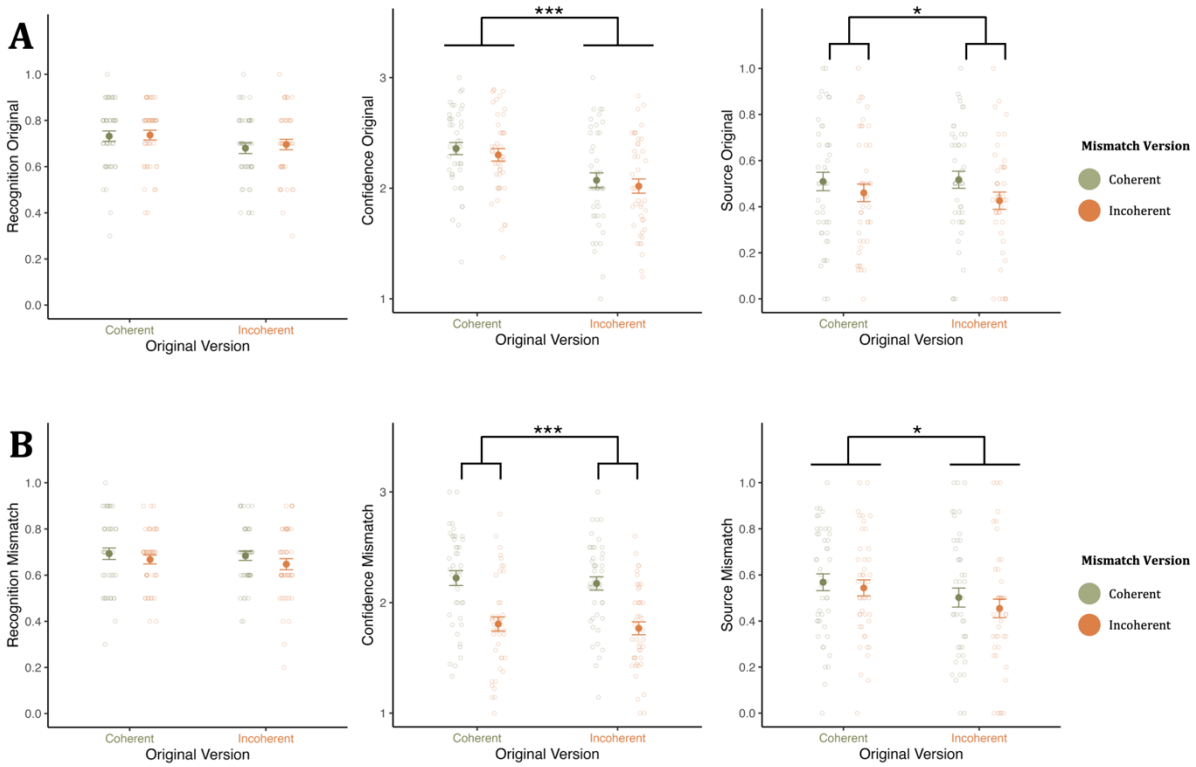
To first establish whether low-precision PEs at first exposure generally influenced memory, we focused on those dialogues that were never changed. Interestingly, there was no evidence that dialogues with and without a low-precision PE at first exposure were remembered differently well (β = .31, *SE*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

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

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

Figure 4

Memory outcomes



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

regardless of whether the original itself was coherent or not and with no interaction (*p*s > .451).

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

Additionally, source memory for the mismatch version was higher when the

original version had been coherent, so after low-precision within-cause PEs compared to low-precision across-cause PEs (β = .37, SE = .18, p = .035), regardless of whether the mismatch version itself was coherent or not and with no interaction (ps > .156).

4. Discussion

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

4.1 General PE response

Before discussing differences between the PEs evoked in the current study, we focus on their commonalities as three regions robustly responded to all PEs, namely IFG, STG/STS, and Crus 2. First, the IFG is involved in PE processing across tasks and contexts (El-Sourani et al., 2020; Li et al., 2019; Schliephake et al., 2021; Wurm & Schubotz, 2012), including episodic PEs (Jainta et al., 2024; Liedtke et al., 2025; Siestrup et al., 2023). Specifically, the IFG may hold the predictive model that will then be compared to reality (Boeltzig, Liedtke, Siestrup, et al., 2025; Fujitani et al., 2024; Liedtke et al., 2025; Sherman et al., 2016). Second, the STG and STS are highly relevant for language processing (Bhaya-Grossman & Chang, 2021) and auditory PEs (Schlossmacher et al., 2022) and may therefore especially have been recruited in the

current design that used conversations as stimuli. However, the STS and STG also process socially relevant information (Masson & Isik, 2021), including speaker familiarity (Cordero et al., 2025), goals of actors (Shultz et al., 2011), and theory of mind (Carrington & Bailey, 2009). This makes this region a key hub in the updating of person impressions (Mende-Siedlecki et al., 2013). Finally, the left Crus 2 in the cerebellum has been shown to be involved in mentalizing (Overwalle et al., 2020), social preference predictions (Haihambo et al., 2023), and responds to PEs in sequences performed by others (Li et al., 2024).

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

4.2 New latent cause inference after low-precision predictions

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

This suggests that when low-precision predictions in novel and unfamiliar situations are violated, across-cause PEs are highly meaningful, informing socially

relevant person models. For instance, such across-cause PEs may signal that the second speaker has misunderstood the first, prompting an expectation of clarification, or that the exchange involves an unobserved third party, leading the listener to infer a different social configuration and informing social impressions.

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

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

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

4.3 New latent cause inference after high-precision predictions

While in the absence of specific predictions, new latent cause inference involved an extended network of brain regions implicated in social processing, fewer regions

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

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

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

4.4 Model updating after high-precision predictions

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

Brain activity was strongly in line with this prediction. Again, PE-general areas

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

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

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

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

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

4.5 Implications for latent cause theory in episodic predictions

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

Second, neural processes at new latent cause inference differ depending on whether predictions are low-precision or high-precision, with fewer brain regions responding when high-precision predictions can be made. At first glance, this might be inconsistent with the idea that high-precision predictions generally trigger stronger PEs (Greve et al., 2017; Henson & Gagnepain, 2010). However, a more distributed

response to PEs is not synonymous with a stronger PE but may rather indicate broad information gathering when little is known about the novel situation. Additionally, the originally encoded memories were more strongly reactively reinstated when an across-cause PE compared to a within-cause PE was induced. This raises the intriguing possibility that these across-cause PEs may have biased the system from bottom-up processing of the unexpected input towards a top-down reinstatement of a previously encoded version of the same episode. This very reinstatement may then have constrained the PE response.

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

Fourth, the results underscore that predictive models can have several sources in any given situation. For instance, when participants first listen to a novel dialogue, it is unclear what they predict and where these predictions stem from. Later, familiar dialogues could be predicted based on a specific episodic memory, as the RSA suggests, and prior research corroborates (Boeltzig, Liedtke, Siestrup, et al., 2025; Haque et al., 2020; Kim et al., 2014; Stawarczyk et al., 2020). However, after low-precision across-

cause PEs, the previously encoded dialogue continuations were less predicted upon future exposures – leaving it open what was predicted instead. Previous studies have usually either induced episodic PEs by modifying previously encoded episodes (Kim et al., 2014; Nolden et al., 2025; Stawarczyk et al., 2020), or semantic PEs by violating abstract rules or schemas (Frank et al., 2020; Kalbe & Schwabe, 2020; Pupillo et al., 2023). In real-life prediction, it is conceivable that both sources are tapped in any given context (Cheng et al., 2016), and that one or more relevant episodic memories are leveraged for prediction in novel situations (Liedtke et al., 2026). As PEs based on previous episodic experiences differ in neural correlates from PEs based on semantic knowledge (Varga et al., 2025), it is an interesting avenue for future research to focus on identifying the sources for predictions in naturalistic viewing and investigate whether they modulate mnemonic and representational outcomes.

4.6 Limitations

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

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

It is a well-known challenge in the field that comparing findings between studies is inherently difficult. Context can moderate the effects of novelty and surprise (Barto et al., 2013; Kafkas & Montaldi, 2018; Quent et al., 2021; Reichardt et al., 2020), and the stability of the environment can influence how strongly the system is tuned to PEs (Rouhani et al., 2018) or whether predictions are made on a category- or item-

level (Kim et al., 2019). In the current study, uncertainty was probably generally higher than in previous studies of PE size, given that new material was continuously introduced and modified in a single session, with more obvious PEs evoked by the incoherent versions. This uncertainty may influence the strength of predictions and render PEs generally smaller (Henson & Gagnepain, 2010). To account for this and other factors which make an exact PE size determination difficult, future studies should focus on modelling the full space of possible PE sizes and consider utilising physiological measures of PE (Brod et al., 2022; Loock et al., 2025).

More broadly, experiments creating episodic PEs rely on participants encoding episodes, and then re-playing these episodes in an identical manner before an unexpected mismatch is induced (Greve et al., 2017; Kim et al., 2014; Varga et al., 2025; Wahlheim & Zacks, 2019). This naturally simplifies naturalistic predicting. Even though we are often exposed to retellings of the same stories and highly repetitive event, and watch movies, read books, and listen to audiobooks that we already know (O'Brien, 2019), the situation including its context will not directly repeat itself. While it can still be highly effective to leverage episodic memories in these situations, future research should also create conditions where the new situation is similar, but not completely identical, and test whether this modulates signalling and consequences of episodic PEs.

5. Conclusions

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

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

Acknowledgements

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

Author Contributions

Boeltzig, M.: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Project Administration

Siestrup, S.: Conceptualization, Software, Formal Analysis, Writing – Review & Editing

Bramão, I.: Conceptualization, Methodology, Writing – Review & Editing

Schubotz, R.I.: Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Funding Acquisition

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.

Competing Interests

The authors have no conflicts of interest to declare.

Data Availability

The data needed to recreate behavioural and representational similarity analysis, the *t*-maps for univariate contrasts, and the stimulus material are publicly available on OSF:

https://osf.io/mgu49/overview?view_only=1af278283f49433d8de09f566a2dff97

References

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

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

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

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

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

Bein, O., Gasser, C., Amer, T., Maril, A., & Davachi, L. (2023). Predictions transform memories: How expected versus unexpected events are integrated or separated in memory. *Neuroscience & Biobehavioral Reviews*, *153*, 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

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

Bhaya-Grossman, I., & Chang, E. F. (2021). Speech computations of the human superior temporal gyrus. *Annual Review of Psychology*, *73*(1), 1–24. https://doi.org/10.1146/annurev-psych-022321-035256

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Greve, A., Abdulrahman, H., & Henson, R. N. (2018). Neural differentiation of incorrectly predicted memories. *Frontiers in Human Neuroscience*, *12*, 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

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

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

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

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

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

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

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

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

Hertrich, I., Dietrich, S., & Ackermann, H. (2016). The role of the supplementary motor area for speech and language processing. *Neuroscience & Biobehavioral Reviews*, *68*, 602–610. https://doi.org/10.1016/j.neubiorev.2016.06.030

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

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

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

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

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

Kemp, P. L., Sinclair, A. H., Adcock, R. A., & Wahlheim, C. N. (2024). Memory and belief updating following complete and partial reminders of fake news. *Cognitive Research: Principles and Implications*, *9*(1), 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

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

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

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

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

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

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

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

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

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

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

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

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

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

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *The Journal of Neuroscience*, *36*(50), 12677–12687. https://doi.org/10.1523/jneurosci.1850-16.2016

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

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

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

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

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

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

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

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

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

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

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

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*, Article 27. https://doi.org/10.3389/fninf.2016.00027

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

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

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

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

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

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

Rehbein, M. A., Kroker, T., Winker, C., Ziehfreund, L., Reschke, A., Bölte, J., Wyczesany, M., Roesmann, K., Wessing, I., & Junghöfer, M. (2023). Non-invasive stimulation reveals ventromedial prefrontal cortex function in reward prediction and reward processing.

Frontiers in Neuroscience, *17*, Article 1219029. https://doi.org/10.3389/fnins.2023.1219029

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

4 Discussion

4.1 Summary of the Presented Studies

The three presented studies substantially support the idea that mnemonic and representational outcomes of episodic PEs strongly depend on PE size. Behavioural, univariate fMRI, and multivariate fMRI analyses supported the hypotheses derived from LCT (Gershman et al., 2017) and SLIMM (van Kesteren et al., 2012). Specifically, small PEs leave the original target memory unaltered. Moderate PEs lead to the integration of the mismatch target into the original target representation, associated with increased frontal engagement and reduced recognition memory. Large PEs lead to the inference of a new latent cause, characterised by little change of the original memory and the distinct encoding of a new mismatch memory trace. The cumulative evidence from the three studies was therefore able to respond to the three research questions, pertaining to mnemonic (RQ1) and representational (RQ2) consequences of PEs of different size, as well as differences in neural signalling (RQ3).

Study 1 behaviourally tested the effect of PE size on memory performance (RQ1). Participants encoded original dialogues three or five times to manipulate prior precision and rated the perceived difference between original and mismatch dialogues at the end of the experiment to measure prior accuracy. Results were in line with the prediction of a U-shaped relationship between PE size and original stability. The original was recognized better after small and large PEs, and worse after moderate PEs, with PE size being produced by the interaction between prior precision and prior accuracy. The experiment also distinguished whether modifications were made on a surface level (i.e., changing only the phrasing) or gist level (i.e., changing the meaning) to characterise the effect of different PE types. While this did not affect memory for the original target, PE type interacted with prior accuracy for mismatch target memory. Larger gist changes (i.e., lower prior accuracy) led to better learning than smaller gist changes, while surface changes led to high memory regardless of the amount of change. We speculated, in line with SLIMM's predictions concerning schema-congruent information, that participants responded correctly to the mismatch targets based on their memory of the highly similar original targets, not because a new memory trace was encoded.

Study 2 extended these findings, replacing the discrete manipulation of prior precision with a neural measure, namely the reinstatement of the original target

during cue re-exposure just before the PE. Prior accuracy was measured behaviourally as a difference rating in an independent sample. Again, recognition test results supported the hypotheses about the effects of PE size on memory (RQ1). Additionally, it was found that larger PEs left original representations more intact than moderate PEs, which led to more change in the memory trace (RQ2). Larger PEs also led to high original and mismatch target recognition memory, but also high source memory for the mismatch version, consistent with the distinct and separated encoding of a new memory trace. Furthermore, there was evidence that larger PEs led to more reinstatement of the original memory trace while the mismatching target was presented. This increased original reinstatement during mismatching input was associated with better memory for both original and mismatch targets. In summary, Studies 1 and 2 found evidence for integration after moderate PEs and distinct encoding after larger PEs.

Studies 1 and 2 created a broad continuum of PE sizes to assess their behavioural and representational consequences. The distinct encoding after larger PEs is in line with LCT (Gershman et al., 2017), which postulates that large PEs lead to the inference of a new latent cause, where the system assumes that the old memory is valid under some circumstances, but that the context has changed so substantially that it does not currently apply. Even though the evidence from Studies 1 and 2 was consistent with this idea, the continuous manipulation of PE size precluded a clear distinction between trials in which this latent cause inference happens and in which an old trace is updated with new information instead. Consequently, we could not assess the neural correlates of updating compared to latent cause inference (RQ3).

Study 3 addressed this and created circumstances where mismatching input clearly did or did not belong to the same underlying latent cause as the original target. A univariate analysis contrasting these cases showed that integrating new information into an existing memory trace (moderate PEs) engaged frontal regions, above all Brodmann area 9 (BA9), which supports the reconciliation of new information with a current situation model (Ferstl et al., 2005; Ferstl & von Cramon, 2001; Yarkoni et al., 2008). When mismatch targets evoked the inference of a new underlying latent cause (large PEs), no additional frontal activation was observed. Furthermore, consistent with LCT, new latent cause inference during novel stimuli induced an event boundary within these episodes (RQ2), leading to less predictive use of the target upon future exposures to the cue.

Together, the findings provide strong support for PE size-dependent episodic memory processes, in line with LCT and SLIMM. Note that further evidence was also provided in a parallel paper based on the same data as Study 2. Highly relevant to the present purposes, it was found that small gist changes, which may engender moderate PEs, triggered increased parahippocampal activity. This is consistent with the updating of an existing memory (Liedtke et al., 2025), similarly to BA9 and vmPFC activity in Study 3 (see Kalbe & Schwabe, 2021 for a similar argument).

4.2 Limitations and Strengths of the Design

Before situating the findings within previous theoretical and empirical work, it is worth evaluating which conclusions can and cannot be drawn based on the design of the current set of studies. While more specific points are discussed in the individual studies, I will point out four overarching limitations before briefly highlighting strengths of the employed stimulus material.

First, we did not employ a direct measure of PE size. Both the behavioural difference ratings (Studies 1 and 2) and the strength of predictive reinstatement (Studies 2 and 3) are measures of only one of the components influencing PE size (see section 1.4.1). Future studies could use a compound measure of PE size, both to validate the experimental manipulations, and to directly predict memory outcomes. Candidates are univariate IFG activation strength (Liedtke et al., 2026), skin conductance (Kalbe & Schwabe, 2020, 2021; Loock et al., 2025), pupil dilation (Brod et al., 2018, 2022; Kafkas & Montaldi, 2015; Theobald & Brod, 2021), or reading times (Haeuser & Kray, 2021). However, the latter three measures have not yet been used for episodic PEs (and see Stemerding et al., 2022 for mixed results of their validity in conditioned PEs). Any use within episodic predictions therefore needs to be validated first, ideally against both neural and behavioural correlates of PEs. This would be a crucial step for the field, allowing to measure trial-to-trial differences in PE size without resource-intensive neuroimaging techniques.

Second, we chose a recognition memory test in Studies 1 and 2, and a similar lure forced choice in Study 3. This constrains the conclusions that can be made about memory outcomes: Studies 1 and 2 did not offer a behavioural test of pattern separation, which could be achieved with a mnemonic similarity task where enhanced similar lure rejection signifies more distinct memories (Bein et al., 2021). None of the studies directly tested for behavioural correlates of memory integration, for instance

by testing whether the two alternative targets are associated with each other (Liu et al., 2024), or for associative memory between the cue and the targets. As the processes of integration and separation play a key role in previous accounts (see section 1.4.4), as well as the framework I will propose (section 4.5), demonstrating effects of PE size on behavioural outcomes would further strengthen these conclusions.

Third, a related point concerns the balance between behavioural and neural measures. In Study 2, to test how representations changed due to the PE, we replayed the original dialogues after PE induction. However, this replay may have clouded behavioural effects, affording an additional learning opportunity for the originals. This may for instance explain why the descriptively U-shaped patterns of original target recognition and source memory did not gain statistical significance. To avoid this replay, one could use the memory test as a post-PE representational measurement, but this would be limited to recognition tests, and due to decodable neural signatures of retrieval states (Kerrén et al., 2021; Long & Kuhl, 2019), this approach may be less reliable. In Study 3, to maximize the odds for latent cause inference, we used incoherent dialogues, which may generally be remembered less well (Arslan & Kominsky, 2026; Cohn-Sheehy et al., 2022), possibly obscuring potential behavioural effects of PEs. Future studies may therefore choose to rebalance different aspects of the design, depending on the specific research question.

Fourth, previous literature (see sections 1.3.1 and 1.3.2) has tested whether the effects of PEs are temporally and conceptually localized, or if events before and after, as well as peripheral details are also enhanced by PEs. Except for testing memory for the cue, our studies do not offer a perspective on this question. The fact that the targets continued for a variable amount of time after the PE was induced may also have clouded some of the effects if they are more temporally constrained. The reach of PEs is an important question for their application and should therefore get more attention in the future.

These limitations can be remedied by shifting focus in future experiments. Conversely, there are strengths of the design that should be pointed out. As memory research increasingly embraces more naturalistic stimulus material (Maguire, 2022), episodic PE studies are somewhat ahead of the trend, often using naturalistic narrative videos instead of arbitrary associations between items (Siestrup et al., 2022; Sinclair & Barense, 2018; D. K. Varga et al., 2025; Wahlheim et al., 2021). In studies of PEs, these naturalistic episodes have the decisive advantage that they entail a coherent and

causal temporal structure, allowing predictions to be updated continuously, as they would in naturalistic predicting. The use of static images, while having other advantages, does not share this quality, as new information is only added every couple of seconds. How these breaks may affect prediction and whether the prior that is formed after cue onset is maintained over time is currently unclear.

The use of dialogues, however, was unprecedented. As a potential disadvantage, the surface form (i.e., the exact phrasing) of verbal statements is often poorly encoded (Reyna et al., 2016; but see Keenan et al., 1977), possibly limiting the effect of small modifications. Furthermore, the exact moment when a PE is induced is more elusive, as mid-sentence changes can be difficult to implement without altering prosody already at the beginning of the sentence. Lastly, and more pragmatically, producing convincing dialogues is labour- and resource-intensive and their language specificity constrains their use to certain samples.

The advantages of the dialogues, on the other hand, are numerous. The stimuli were interesting and engaging, as post-experimental interviews with participants suggested, resulting in high attention check and memory performance. Using the auditory modality without visual input can be especially valuable if combined with EEG or pupil dilation, as there are no eye movements or luminescence differences that would impact these measures. If this is no concern, the dialogues could also be paired with background images to test effects of episodic context on mnemonic consequences of PEs (Bein et al., 2023; Liu et al., 2024). Lastly, by assigning faces or social groups to the speakers, social influences on PEs can be tested (Boeltzig et al., 2023; Hirst & Echterhoff, 2012).

More broadly, the use of auditory stimuli is rather uncommon in episodic memory research. The few studies directly investigating memory for conversations date back several decades (Bates et al., 1978; Keenan et al., 1977; Kintsch & Bates, 1977; Neisser, 1981; Pezdek & Prull, 1993), and progress has largely stalled since (see Brown-Schmidt & Benjamin, 2018 for a review). Both within and beyond episodic PEs, understanding memory processes triggered by conversations is crucial, given that memories frequently change when we talk about them with others (Granhag et al., 2012; Hirst & Echterhoff, 2012). Furthermore, when listening to lectures or podcasts, attending work meetings or conferences, when gossiping with friends, or talking to teachers, doctors, or coaches, information is presented auditorily. Nevertheless, episodic memory experiments often present information visually, leaving it unclear

whether the various effects found in that domain generalize to auditory stimuli. It may therefore be time to leverage the technological advances since the 1990s and have a fresh look at memory for conversations. The stimulus set resulting from this project is well-placed for further investigation of episodic PEs and episodic memory in general.

4.3 Memory After Prediction Errors of Different Sizes

This series of experiments establishes PE size as a central pillar for determining episodic memory outcomes. This is a major advancement in the field of episodic PEs, where most previous studies tacitly assumed that each trial would induce essentially the same PE. Additionally, studies have been compared with each other without explicitly acknowledging that results can diverge because the PEs they elicit are different. While we intentionally created variance in PE size, it is likely that previous studies inadvertently did so as well, as spontaneously arising trial-to-trial differences in prior precision (i.e., natural differences in original reactivation that are not experimentally induced) testify (Study 2; G. Kim et al., 2014, 2017; Stawarczyk et al., 2020). Ignoring the influence of PE size therefore poses a major challenge to validity, reliability, and comparability of studies.

In the following, I will therefore outline in detail which effects of small, moderate, and large PEs were observed on memory and representations, and how they were reflected in brain activity. The focus will be on the results from the three presented studies and their connection to the relevant theories. Wherever possible, comparisons to previous literature, including semantic PEs in addition to episodic ones[7], will be made. As pointed out before, given that PEs are relative within their context, these comparisons are conducted with a degree of caution.

4.3.1 Small Prediction Errors

In Studies 1 and 2, following PIMMS (Henson & Gagnepain, 2010), small PEs were evoked using high prior accuracy and low prior precision (note that no small PEs were created in Study 3). High prior accuracy was generated when original targets were minimally modified. While we did attempt to create gist changes that could be smaller than larger surface changes, participants' difference ratings suggested that the smallest PEs were likely caused by surface modifications, altering the phrasing with

[7] Conditioned PEs are not included in the discussion, because their relation to PE size is complicated by the issue of signed or unsigned, and positive or negative PEs, both of which do not immediately map on epsiodic PEs.

little impact on overall meaning (see Supplementary Materials S1). Low prior precision was orthogonally promoted by allowing fewer encoding opportunities (Study 1) or measured with RSA as the reinstatement of the original target (Study 2).

The resulting small PEs were hypothesized, in line with LCT (Gershman et al., 2017), to trigger no change in original memories. Additionally, following SLIMM (van Kesteren et al., 2012), no encoding of a mismatch memory trace was expected. Both processes would be in line with the idea that the current base of experiences can explain the state of the environment and no adjustment is needed.

Accordingly, we found original memory to be intact after small PEs (Studies 1 and 2), supporting that no integration or weakening took place. Even though mismatch target memory performance was high, small PEs were also associated with low source memory (i.e., participants frequently thought that the mismatch target was the original version). Additionally, unreported data from a cued recall test suggests that participants were more likely to reproduce only one version of the target after small PEs, and more likely to reproduce both versions after larger PEs (see Supplementary Materials S2 and Pronoza et al., 2025). Both analyses support that no new mismatch memory trace was encoded after small PEs and that participants correctly responded to the mismatch target probe based on their memory for the original target.

Previous studies generally favoured inducing more sizeable episodic PEs so that little precedence for small episodic PEs is available. However, G. Kim et al. (2014) used a subtle PE induction, based on implicit episodic learning of long image sequences. They found that smaller PEs (operationalized by lower prior precision measured as predictive original target reinstatement) were associated with less original weakening, so better memory for the original target after the PE, compared to stronger PEs. A different group replicated this finding in a similar paradigm, while specifying that this effect was only observed when an individual item was predictively reinstated, not if an overall visual category was predicted (H. Kim et al., 2019). These findings of robust, unaltered memory after small PEs therefore corroborate our results.

This initial evidence from episodic PEs is also supported by studies using semantic PEs. If new information is similar to a prediction based on prior knowledge, intuitively, there is little knowledge updating (Brod et al., 2022; de Bruïne et al., 2025; Vlasceanu et al., 2020). Additionally, if an event is highly schema-congruent, causing a minimal PE, memory performance relies on schemas instead of the encoding of a

distinct and detailed memory trace (Greve et al., 2019; Quent et al., 2022).

In summary, small PEs evoke little change in episodic memories. Original memories remain largely intact and no new memories are encoded. While this has been shown before in semantic PEs and formalized in the SLIMM framework (van Kesteren et al., 2012), our studies extend this effect to episodic PEs.

4.3.2 Moderate Prediction Errors

Moderate PEs were evoked by a combination of prior accuracy and prior precision in Studies 1 and 2. In the categorical design of Study 3, moderate PEs were induced by using noticeable mismatches that however were still clearly consistent with the cue and the original target.

Generally, these moderate PEs led to weakened memory, both for the original and for the mismatch target (Study 3 being the exception, see that paper for a discussion). They also were associated with decreased representational stability of the original memory trace, indicating that this trace was modified (Study 2). The univariate activity that accompanied moderate PEs was unique, most notably including BA9 and the vmPFC (Study 3). Additionally, in a separate analysis of data from Study 2, we found increased parahippocampal activity for moderate, compared to large PEs (Liedtke et al., 2025).

This pattern of results is consistent with the notion proposed by Gershman et al. (2017) that moderate PEs lead to the integration of new information into an existing model. In the context of episodic PEs, this amounts to the encoding of an integrated representation between original and mismatch targets (Schlichting & Preston, 2015; Zeithamova & Bowman, 2020), which alters the original trace, manifesting in reduced representational stability. The tendency of these representations to reduce memory for individual event features, especially source and detail memory (Carpenter & Schacter, 2017, 2018b) can explain the lower levels of recognition memory. Additionally, this process has been suggested to require frontal activation which we found in the vmPFC, which is robustly associated with integrative encoding (Bowman & Zeithamova, 2018; Spalding et al., 2018; Zeithamova et al., 2012), and BA9, which supports the updating of an existing model with accommodable information (Ferstl et al., 2005; Ferstl & von Cramon, 2001; Yarkoni et al., 2008). The increased parahippocampal activity observed by Liedtke et al. (2025), as well as a study finding increased hippocampal activity after moderate PEs (Long et al., 2016) could also be seen to be consistent with

updating (also see Kalbe & Schwabe, 2021).

Integrated representations can enable higher mnemonic flexibility, supporting inferences across event boundaries (Bauer et al., 2020; Liu et al., 2024; Schlichting & Preston, 2015). In the context of PEs, this could manifest in enabling connections between the old and new state of the environment (Wahlheim & Zacks, 2025). Additionally, according to LCT (Gershman et al., 2017), updating the existing model with new information after moderate PEs is the most adaptive operation. Moderate PEs indicate that the environment has changed slightly, but not radically, making it efficient to update the existing episode with new information, resulting in only one version of the event that is best placed to guide behaviour effectively in the future.

While our studies were the first to directly support the notion of integration after moderate episodic PEs, some previous studies are consistent with it. Encoding conditions that may have promoted moderate prior precision have led to less robust original source memory (Jainta et al., 2022; Siestrup et al., 2023) and reduced mismatch target learning (Greve et al., 2017) compared to larger PEs. Furthermore, a moderate amount of change in the episode (i.e., moderate prior accuracy) has been associated with weakened memory for original targets (Bein et al., 2021). Direct evidence also comes from semantic prediction, where moderate PEs promoted the lowest level of subsequent memory, compared to small and large PEs (Brod et al., 2022; de Bruïne et al., 2025). While these studies do not provide direct evidence for integration after moderate PEs, it is one possible explanation for the observed reduced memory.

There is directly opposing evidence as well. Ortiz-Tudela et al. (2023) found evidence for an inverted U-shape, where memory was best after moderate PEs. In that paradigm, participants learned arbitrary contingencies between scenes and objects (e.g., a picture of a beach being predictive of musical instruments). To create PEs of different sizes, it was manipulated how strong these contingencies were, allowing either rather certain (e.g., the beach being followed by instruments in 80% of trials, and by household items and fruits in 10% each) or uncertain predictions (e.g., the beach being followed by instruments, household items, and fruits in 33% of trials each). One explanation by the authors for the unexpected result was increased information seeking under the high uncertainty evoked by moderate PEs, giving way to stronger encoding. While this is a valid explanation, a situation where a cue signals high uncertainty is not common in episodic PE studies: Here, despite natural or

induced fluctuations in prior precision, participants will always have some information allowing a prediction, in addition to a rich base of extra-experimental episodic memories and semantic knowledge. Therefore, the high uncertainty, combined with the obligation to make an overt prediction (see section 4.5.5 for a critical evaluation of this) may have created the results in the study by Ortiz-Tudela et al. (2023), while they may not apply to episodic PEs.

4.3.3 Large Prediction Errors

Large PEs were induced by strong priors due to robust encoding (Study 1) or high spontaneous original reactivation measured using RSA (Study 2), by strong deviations from previously encoded versions (Studies 1 and 2), or by using mismatch targets that were incoherent with the cue and the original target (Study 3). According to LCT (Gershman et al., 2017), large PEs should trigger the inference that the underlying latent cause has changed: The observed event is so inconsistent with the current prior that a new prior is needed. Thus, and in line with SLIMM (van Kesteren et al., 2012), the new event is encoded separately, while the original trace remains unmodified, and therefore representationally stable. In other words, both events are seen as valid under different circumstances and therefore preserved in memory.

Accordingly, we consistently found the highest recognition memory for original and mismatch targets after large PEs (Studies 1 and 2). The idea that the new target was encoded distinctly is supported by high source memory after larger PEs and more unchanged original representations, both compared to moderate PEs (Study 2). This is consistent with an account of pattern separation, where similar memories are differentiated in order to reduce interference (Brunec et al., 2020; Chanales et al., 2017; Drascher & Kuhl, 2022; Favila et al., 2016).

Our findings also offer an account of how this distinct encoding was triggered. First, we found that larger PEs evoked a stronger reactivation of the original targets during the incoming mismatch. It has been shown that this reactivation of a memory during the processing of new and similar input can protect against interference and forgetting (Detre et al., 2013; Kuhl et al., 2010; Newman & Norman, 2010; Scully & Hupbach, 2020). Second, while we found that moderate PEs engaged the vmPFC and BA9, the process of distinct encoding after large PEs seems to be implemented without the need for additional frontal activation (Study 3). Instead, reactivation alone seems to determine the representational outcome. This empirical finding strongly supports a

computational model by Pronoza et al. (2025), using behavioural data from Studies 1 and 2. This model was set up without an explicit decision maker determining whether new input should be integrated into existing representations or encoded separately. Instead, internal network dynamics based on similarity between events were sufficient to determine representational consequences, which is in line with our findings.

Our empirical results, along with a proposal for the mechanisms underlying distinct encoding after large PEs are therefore an important advancement in the field. They supplement previous studies, which often found more robust original and mismatch target memory after larger episodic PEs (Bein, Duncan, et al., 2020; Greve et al., 2017; Jainta et al., 2022; Siestrup et al., 2023; Siestrup & Schubotz, 2023; Vlasceanu et al., 2018). Additionally, larger semantic PEs have been found to strongly promote learning (Brod et al., 2022; de Bruïne et al., 2025; Greve et al., 2019; Vlasceanu et al., 2020).

4.3.4 Taking Stock and Open Questions

To summarise, the behavioural and neural results of the three presented studies strongly support an account of no change after small PEs, integration of new information into an existing memory after moderate PEs, and distinct encoding of the mismatch target after large PEs. These findings are in line with LCT (Gershman et al., 2017) and SLIMM (van Kesteren et al., 2012). While previous accounts have often argued for PEs to either generally promote separation (Frank et al., 2020), or generally promote integration (Wahlheim & Zacks, 2025), our results suggest that both outcomes can be promoted by episodic PEs, depending on their size.

Before evaluating the account based on LCT and SLIMM against other explanations of PE effects (section 4.4) and developing a more general framework of PE effects in context (section 4.5), three open questions about PE size and memory outcomes left unanswered in the presented studies should be acknowledged.

First, the mnemonic effects of prior precision (i.e., how strongly the original is being reinstated before the PE) pose a challenge. In Study 2, stronger prior precision (and thereby a larger PE) was associated with better original and mismatch recognition. This is contrary to the effect found by G. Kim et al. (2014), who demonstrated a linear negative effect. Interestingly, an unpublished analysis of Study 3 data (see Supplementary Materials S3) also finds this negative effect for the mismatch memory target. Thus, while comparatively higher prior precision enhanced

memory in Study 2, it weakened it in Study 3 and G. Kim et al.'s study.

This pattern can potentially be explained by how strong predictions were overall in different studies. Study 3 and the study by G. Kim and colleagues employed a continuous stream of short stimuli and intermixed original encoding trials with PE trials, resulting in weaker original encoding, a more uncertain situation, and therefore weaker predictions overall. In Study 2, however, the dialogues were more contextually embedded (Bäuml, 2019; Bramão et al., 2017; Godden & Baddeley, 1975), fewer in numbers, encoded both inside and outside the scanner, so in two spatial contexts (Imundo et al., 2020), and consolidated overnight (Abel et al., 2023). As all these factors can enhance encoding or facilitate retrieval, predictions may have been stronger overall. Therefore, Studies 2 and 3 together may form the full U-shaped continuum from weak to strong priors. Such a U-shaped relation between prior precision and original memory was indeed found in one study (Vlasceanu et al., 2018), and a replication should be possible within the current design.

Second, in Study 3, we replaced the continuous manipulation of PE size by a categorical one, comparing obviously incoherent episodes which elicited latent cause inference, with coherent episodes, which should lead to model updating. While this was a necessary first step to distinguish neural correlates of these two processes, a continuous perspective on univariate brain activity would be valuable. Observing transition points where activity changes can furthermore help to better determine when moderate PEs become large, which can feed back into designing future experiments.

Third, the relative contributions of prior precision and prior accuracy to PE size, as well as mnemonic outcomes cannot yet be determined. It is possible that they do not contribute equally, or that their importance differs between contexts. As few previous studies have manipulated both factors, this remains a challenge for future research.

As an overall point, it is evident that more studies are needed. Our work was novel in manipulating episodic PE size continuously and measuring effects on both original and mismatch memory. This must be extended to other stimuli (including less complex stimuli allowing cleaner representational measurements), designs, and memory tests, including behavioural tests of integration (Preston et al., 2004) and separation (Bein et al., 2021; Frank et al., 2020). These replications and extensions will increase certainty about mnemonic and representational outcomes after PEs.

4.4 Existing Explanations for Mnemonic Effects of Prediction Errors

Despite the remaining challenges, the current series of studies suggests a theoretically grounded picture of size-dependent integration or separation as the basic mechanism of memory outcomes after episodic PEs. In this section, I will discuss previously suggested theoretical accounts of PE-based learning and evaluate whether they can explain the current findings.

4.4.1 Reconsolidation Theory

Two high-impact studies of episodic PEs explain their findings in terms of reconsolidation theory (Sinclair et al., 2021; Sinclair & Barense, 2018). As discussed in section 1.1.2, reconsolidation theory posits that memories are rendered labile when retrieved, making them susceptible to change or disruption during an ensuing reconsolidation window (Exton-McGuinness et al., 2015; Fernández et al., 2016). Accordingly, Sinclair et al. let participants watch previously encoded action videos that were interrupted unexpectedly to induce a PE. Watching a similar video afterwards led to increased intrusions in the recall of the original video, compared to a no PE condition (Sinclair et al., 2021; Sinclair & Barense, 2018).

While there are doubts about the biological evidence for reconsolidation in humans which are beyond the scope of this thesis (Elsey et al., 2018; Gisquet-Verrier & Riccio, 2018, 2019), reconsolidation theory crucially cannot explain the results of the presented studies. The main reason for this is that reconsolidation, as a molecular process, is time-dependent. Specifically, there needs to be a delay of three to ten minutes between reactivation and interfering information – any information before this point should not impact the original memory trace. In our experiments, the PE is evoked by the mismatching information itself, so that reactivation and interfering information are separated not by minutes, but by seconds. Potentially interfering information that could cause intrusions is therefore presented too early to take an effect under reconsolidation theory. Additionally, reconsolidation effects are thought to be triggered by incomplete reminders, consisting of only parts of the original episode (Forcato et al., 2010; Sinclair & Barense, 2018). Here, we presented the full dialogues that, despite including a mismatch target, began and ended in the same way as the original stimuli. Whether this qualifies as an incomplete reminder is doubtful. Lastly, reconsolidation theory would predict that all PEs should lead to roughly the same detrimental effect on the originals as they were all reactivated in the same way.

This, however, is inconsistent with our results of PE size-dependent memory outcomes.

The fact that reconsolidation cannot explain the current set of results is not to say that it is generally false. It may apply in contexts such as conditioned PEs (Ergo et al., 2020), or in animal studies (Fernández et al., 2016). Another possibility is that reconsolidation is limited to negative PEs (see section 1.3.1). If a familiar video is suddenly interrupted (as done in Sinclair's studies), this is akin to a predicted outcome not materialising and thereby fundamentally different from how PEs are usually induced in the field, where an expected outcome is replaced by another. Future studies should build on work characterising neural signatures of positive and negative PEs (Schliephake et al., 2021; Siestrup et al., 2025) and test how these may differentially affect memory.

4.4.2 Event Memory Retrieval and Comparison Theory

EMRC (Hermann et al., 2021; Wahlheim & Zacks, 2019, 2025) is currently the only theory of mnemonic effects of specifically episodic PEs (see section 1.4.4). EMRC argues that consciously detected PEs trigger integration, leading to the encoding of a recursive representation comprising original and mismatch episodes, their temporal sequence, and the fact that a PE was elicited. While there has not been a direct test of this theory that would refute it, there are inconsistencies between the claims of EMRC and the evidence they are based on, as well as with evidence from other studies, including the presented ones. For instance, EMRC postulates that conscious change detection is necessary to avoid interference. This role of consciousness does not align well with predictive processing, which claims that predicting is an automatic and usually unconscious process (Friston & Kiebel, 2009; Rowe et al., 2020). The role of consciousness is also incompatible with data showing memory consequences of PEs in implicit tasks (G. Kim et al., 2014), and is unlikely to hold in our studies.[8]

The most crucial problem against the backdrop of our results, however, is that there is no graded perspective to PEs in EMRC. Instead, it postulates that all noticed

[8] In Studies 2 and 3, we asked participants at the end of the experiment how many of the dialogues they thought were changed. In both studies, 66% of dialogues were actually modified. In Study 2, the mean guess was 49% (*SD* = 17%, range: 3% - 80%). In Study 3, the sample guessed that *M* = 43% (*SD* = 35%, range = 5% - 100%) were modified. While this retrospective measure naturally has its drawbacks, this number is considerably lower than overall memory performance in each study (Study 2: 75%, Study 3: 71%), making it unlikely that change detection was a prerequisite for memory in these studies.

PEs will lead to integration, and all unnoticed PEs will cause interference. At the same time, the evidence for integration, even after consciously noted PE, is limited. It is based on a test strategy where original targets are only tested if participants indicate that they remember a change after having been tested on the mismatch target – leaving it open whether they would remember the original target, but not its connection to the mismatch target. Furthermore, the authors present no neural evidence for integration, only showing that stronger intentional original reactivation before the PE leads to better memory (Stawarczyk et al., 2020). While this is consistent with the idea that stronger prior precision will cause a larger PE (Henson & Gagnepain, 2010), it is no evidence for integration. Additionally, this finding was observed on a between-subject level, with only weak evidence in a trial-to-trial perspective within subjects. Thus, this analysis merely shows that participants who are better at intentionally reactivating the original memory also have better subsequent memory.

I would argue that there is not only a lack of evidence for integration in EMRC, but that the underlying results are even fully consistent with pattern separation. Original reactivation does not always lead to integration, but can also promote distinct encoding or pattern separation (Study 2; Jacques et al., 2013; Kuhl et al., 2010; Zeithamova & Bowman, 2020), and even pattern separated representations can be connected when prompted at test (Carpenter & Schacter, 2017, 2018b; Zeithamova & Bowman, 2020). Furthermore, the (untested) claim of robust sequence memory after PEs may rather align with pattern separation, as source and detail memory are thought to be impaired after memory integration with its purpose of generalisation (Carpenter & Schacter, 2017, 2018a, 2018b; Varga et al., 2019; but see Boeltzig et al., 2023 and de Araujo Sanchez & Zeithamova, 2023). As the changes induced in the studies on which EMRC bases are rather strong, the possibility of pattern separation fits neatly into the account of distinct encoding after large PEs.

Therefore, EMRC's hypotheses currently remain in need for more support and the evidence it builds on is consistent with alternative accounts as well. Against the background of the current findings, LCT (Gershman et al., 2017), and SLIMM (van Kesteren et al., 2012), it is highly likely that integration does happen after PEs, but as one possible outcome, not as the universal one.

4.4.3 Dopamine

One biological explanation why PEs enhance memory is that they activate the

ventral tegmental area, triggering the release of dopamine. Dopamine, in turn, activates the hippocampus, leading to enhanced encoding (Bunzeck et al., 2010; den Ouden et al., 2012; Ergo et al., 2020; Jang et al., 2019; Schiffer et al., 2012). Interestingly, dopamine may also play a role when outcomes are worse than expected, as the locus coeruleus co-releases dopamine and norepinephrine (Groves et al., 2025; Rouhani et al., 2020, 2023; Rouhani & Niv, 2021).

There is initial evidence that dopamine could also play a role in semantic PEs, as at least two studies found the ventral tegmental area to be activated due to violations of schemas (Frank et al., 2022; D. K. Varga et al., 2025). Whether dopaminergic modulation is also relevant for episodic PEs is hitherto unclear, as at least two studies testing for the role of the ventral tegmental area found no effects (Sinclair et al., 2021; D. K. Varga et al., 2025). As the trade-off between different representational outcomes after episodic PEs is associated with the complex reorganization of episodic memories and therefore goes beyond enhanced or reduced encoding, it is currently unclear which role dopamine can play in this process. Nevertheless, it is likely that there is such a role, which should be addressed by future research.

4.4.4 The Need for a Better Explanation

In summary, reconsolidation theory and EMRC cannot comprehensively and parsimoniously explain the results of the three presented papers, and too little is known about dopaminergic modulation after episodic PEs. This also exposes the lack of theorising about specifically episodic PEs, with only one prior suggestion how they can affect memory (Wahlheim & Zacks, 2025). Therefore, attempting to fill this void, in the next section, I will propose a framework for memory consequences of episodic PEs, incorporating and extending the theories tested in the current set of studies.

4.5 The REPRISE Framework

Based on the theoretical background and empirical evidence, I propose that the mnemonic consequences of episodic PEs are best understood as a size-dependent trade-off between integration and separation. The reactivation of a previously encoded episode, both as a prediction, and in reaction to incoming mismatching input, is proposed as a driving factor in this process. Below, I will introduce these ideas as the *Reactivation and Prediction Causing Integration and Separation in Episodic*

Memory framework, REPRISE[9] for short, outlining its theoretical and empirical foundations, explaining the suggested mechanism, and discussing its limitations, open questions, and strengths. I will argue that this novel framework parsimoniously explains previous results, grounds them in existing theories from other types of PEs and merges them with episodic memory processes. REPRISE can also explain why some previous studies have not found the expected memory enhancement of PEs, and can inform decisions about future experimental designs.

4.5.1 Theoretical Foundations

REPRISE draws on the three theoretical foundations discussed in this thesis. First, it incorporates the principles of PE size determination outlined in PIMMS (Henson & Gagnepain, 2010). The three components of PE size, prior precision, prior accuracy, and evidence precision (this factor was not manipulated in the three studies of this thesis) are incorporated, together contributing to PE size.

REPRISE draws strongly on LCT (Gershman et al., 2017), which proposes size-dependent mnemonic and representational effects of PEs. Specifically, small PEs cause no change to existing memories, moderate PEs trigger integration of new information into an existing model, and large PEs leave the original information intact, as the new information is thought to pertain to a different underlying latent cause. LCT mainly focuses on the impact on the original memory trace and was postulated based on conditioning studies and therefore lacks a mechanism of memory modification appropriate for more complex episodic memories.

Hypotheses for the encoding of the mismatch memory can be derived from SLIMM (van Kesteren et al., 2012). This framework proposes more mPFC-driven integration into existing knowledge for schema-congruent information, and more MTL-driven pattern-separated encoding for events incongruent with schemas. However, SLIMM targets only the interaction of new events with pre-existing semantic memory, not with episodic memories used for prediction. It therefore does not offer a view on how the original and mismatch events interact with each other after PEs.

[9] In music, the sonata form (a technique often used in the first movement of a piece from the classical period) consists of the exposition, introducing the melodic material, the development, where this material is elaborated and re-worked, creating something new from it, and the recapitulation. The reprise is part of the exposition: The material is simply repeated, allowing the listener to get acquainted with it to trace motifs during the ensuing development. The name of the model is therefore fitting: Just like a musical reprise, previously encoded memories are being reactivated, preparing their transformation, where they can create something new in interaction with incoming mismatching information.

In addition to these theories, REPRISE builds on basic research concerning memory integration and separation. As described in section 1.1.3, integration refers to the formation of linked representations of similar or overlapping events, while separation describes the formation of distinct episodic memory traces (Brunec et al., 2020; Kumaran et al., 2016; Schlichting & Preston, 2015; Zeithamova & Bowman, 2020). The two processes have different consequences. While integration helps to generalize and make inferences across events (Carpenter & Schacter, 2017, 2018a, 2018b; Zeithamova & Preston, 2010), pattern separation supports detailed memories of discrete events and reduces the risk of interference between them (Brunec et al., 2020; Chanales et al., 2017; Drascher & Kuhl, 2022; Favila et al., 2016). Crucially to REPRISE, both processes have been linked to the reactivation of an older memory while experiencing a similar new event (Kuhl et al., 2010; Zeithamova et al., 2012).

4.5.2 Proposed Mechanism

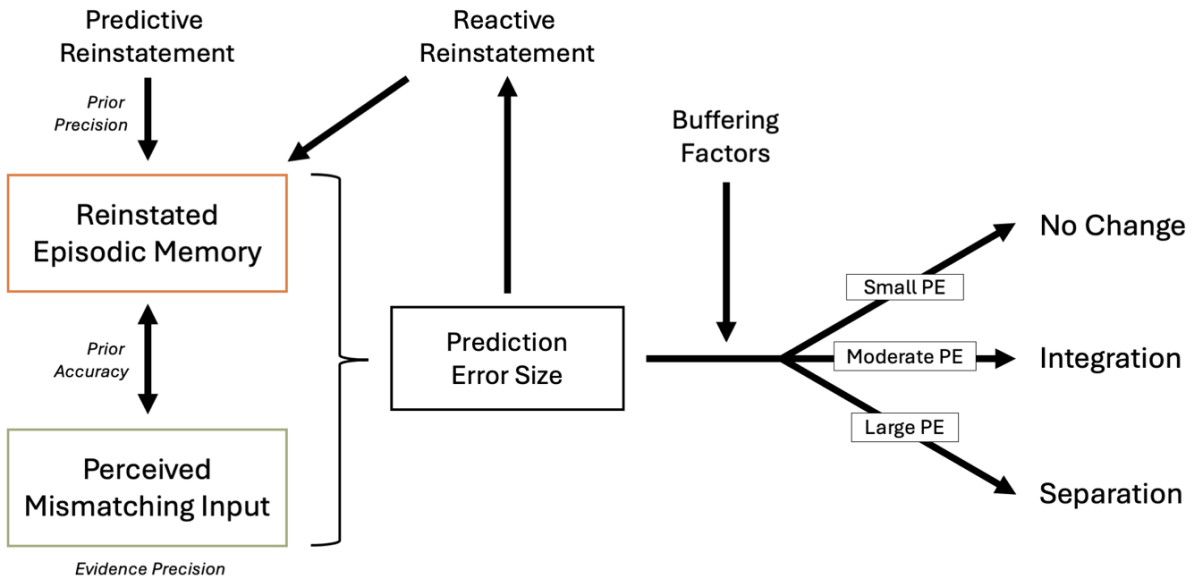
The mechanism suggested by REPRISE is shown in Figure 3. As a first overview, it includes an episodic memory, reinstated predictively and reactively, which is compared to incoming input. These elements jointly contribute to PE size which, both indirectly via reactive reinstatement, and directly through the recruitment of relevant brain regions, biases outcomes towards maintenance without change, integration, or separation. Furthermore, the model includes buffering factors that can influence these outcomes. These elements will now be unpacked and discussed step by step.

The described process necessarily starts with an episodic prediction. Generally, a prior has to be formed for any PE-dependent effects to be triggered, as there is no PE without a prediction (Henson & Gagnepain, 2010; Quent et al., 2021). Additionally, this prediction needs to be sufficiently episodic for REPRISE's hypotheses to apply. That does not mean that semantic knowledge (section 1.3.2) or conditioned associations (section 1.3.1) cannot contribute to predictions alongside specific episodic memories (Cheng et al., 2016). It is likely that there will always be some degree of interplay between these sources of predictions (see section 4.6.2 for a discussion of the "predictive mix" and the selection of episodic memories for prediction).

Episodic memories are reinstated as a prediction. That means that original targets are retrieved when their relevant cue is presented in preparation for the familiar unfolding of the episode (Study 2; G. Kim et al., 2014, 2017; H. Kim et al., 2019). Predictive reinstatement, by definition, takes place before a PE is generated, as

Figure 3

The REPRISE Framework



the brain anticipates upcoming input. It is therefore congruent to prior precision from the PIMMS framework so that stronger predictive reinstatement reflects higher prior precision, meaning that PEs will be larger (Studies 1 and 2; Greve et al., 2017; Henson & Gagnepain, 2010; G. Kim et al., 2014).

The episodic memory that is predictively reinstated will then be met with the evidence, so the actual input. This input can differ in its quality (i.e., evidence precision) and more reliable input will trigger larger PEs than perceptually ambiguous input (e.g., when you are not sure if you correctly understood someone; Greve et al., 2017).

The predictively reinstated episodic memory and the new input will then be compared with each other to establish how unpredicted the input was. This process can be described with prior accuracy (Henson & Gagnepain, 2010). Lower similarity between prediction and input (i.e., a worse prediction) will evoke larger PEs than higher similarity (Studies 1 and 2; Bein, Duncan, et al., 2020; Brod et al., 2022; Long et al., 2016; Vlasceanu et al., 2020).

Together, the strength of reactivation of the episodic memory, the clarity of the input, and the degree of similarity between prediction and input, produce a PE signal (Greve et al., 2017; Henson & Gagnepain, 2010). This PE – and this is the key element to REPRISE – can differ in size, depending on its three components. While it also

depends on the stimulus type which regions signal PEs, it is likely that the IFG scales with PE informativity across domains (El-Sourani et al., 2020; Fujitani et al., 2024; Jainta et al., 2024; Siestrup et al., 2023; Siestrup & Schubotz, 2023). Furthermore, the hippocampus has been shown to respond to contextual novelty (Barbeau et al., 2017; Kafkas & Montaldi, 2015), acting as a mismatch detector when predictions are violated (Bein, Duncan, et al., 2020; Long et al., 2016). It may especially do so based on episodic mismatches, as the hippocampus did not respond to schema violations, but only to deviations from previously encoded events in a recent study (D. K. Varga et al., 2025). This would place the hippocampus at the centre of the comparison process between a predictively reinstated episodic memory and new incoming input.

The hippocampus, IFG, and other domain-specific regions will therefore generate a PE signal, scaling with relevance or size, based on prior precision, prior accuracy, and evidence precision. This signal will bias memory outcomes in two paths, one direct and one indirect.

Starting with the direct path, PE size may directly trigger processes resulting in different representational outcomes. Following LCT (Gershman et al., 2017) and SLIMM (van Kesteren et al., 2012), REPRISE predicts no change after small PEs, integration after moderate PEs, and distinct encoding after large PEs. Integration may directly be triggered by moderate PEs via brain areas such as BA9 (Study 3; Ferstl et al., 2005; Ferstl & Cramon, 2001) and vmPFC (Study 3; Bowman & Zeithamova, 2018; Spalding et al., 2018; Zeithamova et al., 2012), but possibly also by the activity of different hippocampal subfields (Bein, Duncan, et al., 2020). While signalling pathways are yet to be established, it is plausible that the activation of these higher-order regions directly supports the process of integrating new information into an existing memory.

In addition to higher-order regions triggering representational outcomes, the reactivation itself may directly contribute to this, especially after large PEs. A computational model based on behavioural data from Studies 1 and 2 suggests that internal network dynamics based on the similarity between the original and mismatch events is sufficient to model representational consequences and that no explicit decision maker is needed (Pronoza et al., 2025). This tallies with the finding that new latent cause inference was not associated with unique univariate activity (Study 3) and with studies showing that reactivation of the original memory is directly associated with representational outcomes (Kuhl et al., 2010; Scully & Hupbach, 2020;

Zeithamova et al., 2012). Therefore, both original reactivation and additional involvement of brain regions can directly lead to no change, integration, or separation after PEs, while their respective contributions may differ based on PE size.

In addition to this direct path, there is an indirect path, as PE size influences original reinstatement reactively. Studies 2 and 3 showed that in the time window during which the mismatching input was presented, larger PEs were associated with stronger reinstatement of the original. This reinstatement is not predictive, as the time window for prediction is already closed. Instead, it was directly affected by prior accuracy: If mismatch episodes were more different from the original, reinstatement was stronger. How different the mismatch target will be is not known before it is processed and can consequently not affect predictive reinstatement. Therefore, it is suggested in REPRISE that larger PEs lead to increased reactive reinstatement of the original target. This adds to or prolongs the reinstatement that started predictively before the mismatch onset and can therefore further bias representational outcomes (see section 4.5.3 for a more detailed discussion of this indirect path).

Based on the effect of prior precision on PE size and because of the indirect path where PE size influences original reinstatement, REPRISE assumes that PE size and the strength of original reactivation are correlated. Smaller PEs are associated with weaker original reactivation, and larger PEs are associated with stronger original reactivation. Therefore, stronger original reactivation is associated with pattern separation in the model.

However, due to evidence concerning different representational consequences of original reactivation (Long et al., 2016) and studies showing that factors after the PE can influence effects, the mapping of PE size on representational outcomes is not absolute. Instead, REPRISE suggests that there are post-PE buffering factors that can bias representations towards no change, integration, or separation, orthogonal to PE size. These factors have been shown to include plausibility (de Bruïne et al., 2025) and prediction correctness (Pupillo et al., 2023), but have rarely been explored so far (see section 4.6.3 for further discussion).

4.5.3 Testable Hypotheses and Open Questions

An overview of hypotheses that can be derived from REPRISE is provided in Table 1, together with the current level and sources of evidence.

While there is substantial theoretical reasoning, especially from PIMMS, for

Table 1

Testable hypotheses by REPRISE

	Hypothesis	Evidence Level	Current Studies	Other Studies
1	Stronger predictive reinstatement leads to larger PEs	+++	1, 2	Greve et al. (2017)
2	Bigger differences between prediction and input lead to larger PEs	+++	1, 2, 3	Bein, Duncan, et al. (2020)
3	Higher evidence precision leads to larger PEs	+	NA	Greve et al. (2017)
4	Larger PEs lead to stronger reactive reinstatement	+	2, 3	NA
5	Buffering factors can influence mnemonic PE effects	+	NA	de Bruïne et al. (2025)
6	Small PEs cause little or no memory change	++	1, 2	G. Kim et al. (2014)
7	Moderate PEs cause integration	++	1, 2, 3	Vlasceanu et al. (2018)
8	Large PEs cause separate encoding	++	1, 2, 3	Siestrup et al. (2023)

Note. An estimation of the strength of current evidence is indicated in three levels, from low (+) to high (+++). Additionally, it is listed which studies in this thesis support each hypothesis. One prior study is also listed as an example supporting each hypothesis. Entries of NA indicate that direct evidence is lacking.

Hypotheses 1-3 that higher prior precision, lower prior accuracy, and higher evidence precision lead to larger PEs, the current solid level of evidence (Bein, Duncan, et al., 2020; de Bruïne et al., 2025; Greve et al., 2017; Henson & Gagnepain, 2010; G. Kim et al., 2014) can be further strengthened in the field of episodic PEs. Especially evidence precision, to my knowledge, has only been manipulated in a single study (Greve et al., 2017). Further work can also utilise compound measures of PE size such as skin conductance (Kalbe & Schwabe, 2020, 2021; Loock et al., 2025) or pupil dilation (Brod et al., 2018, 2022; Kafkas & Montaldi, 2015; Theobald & Brod, 2021). If these measures are responsive to manipulations of PE components, or correlate with reinstatement, this would further support the relationship between PE size and its manipulations.

The part of REPRISE with the least solid evidence is the effect of PE size on reactive reinstatement (Hypothesis 4), constituting the indirect path of how PE size can bias representational outcomes. In Studies 2 and 3, we found that larger PEs led to stronger reinstatement of the original, but few previous studies have focused on this process.

However, a study by Bein, Duncan, et al. (2020) offers some potential support for this effect. Participants were asked to retrieve a previously encoded visual scene in their mind. Once they had done so, they were shown the scene, but with zero to four alterations. Compared to no changes, two or three alterations reduced representational similarity in hippocampal subfield CA1 between retrieval and mismatch displays. However, interestingly, scenes with one or four alterations did not significantly differ in representational similarity from scenes with no alterations. High similarity after only one change can be simply explained by high perceptual similarity. However, four alterations did not significantly differ in neural similarity from no alterations either, even though perceptual similarity should be lowest. With all given caution in interpreting the lack of an effect, this would be consistent with larger PEs inducing increased original reactive reinstatement. Importantly, this effect is directly mirrored in memory accuracy, where performance is lowest after one change and then increases, while four changes produce comparable memory to no changes at all – consistent with pattern separation after large PEs.

While more research is needed to establish potential mechanisms behind this effect, one can speculate that it is supported by two processes. First, some automatic reactivation of the original target by the incoming mismatch target may always be expected based on pattern completion (Berntsen, 2009; Grande et al., 2019; Horner et al., 2015; Joensen et al., 2024; Moccia et al., 2025; Sun et al., 2025). However, this does not explain why larger PEs should make this process stronger. Second, however, the areas responsive to PE size may trigger this process: The IFG, which maintains the predictive model (Study 2; Fujitani et al., 2024), may do so more strongly or more protractedly if larger PEs are detected, and the hippocampus, as a mismatch detector, may directly trigger retrieval to facilitate comparisons.

Functionally, increased reinstatement of the original can serve to protect it from interference (Kuhl et al., 2010; Scully & Hupbach, 2020). This is especially relevant after large PEs, given their signalling of a new latent cause. Even though the new mismatch version is important in its context and needs to be encoded, the original is

also still relevant under different circumstances and thus needs to be protected against interference (Gershman et al., 2017). Thus, both memory traces should be stored separately from each other. Similar processes can be observed in event segmentation, where the strength of boundaries between events, signalled by hippocampal activity, is predictive of later memory (Barnett et al., 2024; Y. Lee & Chen, 2024).

It should be pointed out that we used naturalistic stimuli that extended for several seconds. We modelled the whole targets as epochs, capturing neural activity across the mismatching input, and not only immediately after the PE. This may allow the hippocampus to fulfil its functions in encoding and retrieval (B. E. Sherman & Turk-Browne, 2020) at different time points. It is conceivable that reactive reinstatement takes some time to develop and may therefore not be detectable with static and short stimuli. It would therefore be interesting to employ EEG, combined with multivariate techniques such as RSA. This allows tracking cortical reinstatement with high temporal precision and can therefore elucidate the time profile of original reactivation (Bramão et al., 2022; Liu et al., 2026), a point that will be further outlined in section 4.6.1.

Hypothesis 5 pertains to buffering factors that can remove, weaken, or enhance mnemonic effects of PE. Task demands can also take effect here, harmonizing the idea of separation after large PEs with studies that found integration after stronger original reactivation (Chanales et al., 2019). As these studies sometimes include an integration instruction at encoding, this will likely bias representational outcomes. Additionally, plausibility judgements (de Bruïne et al., 2025) and the correctness of the prediction (Pupillo et al., 2023) have been shown to reduce or eliminate effects, possibly due to a conscious rejection of the information. Due to the high importance of these mechanisms, buffering factors are discussed further as future research (section 4.6.3).

Lastly, Hypotheses 6-8 refer to how PE size relates to representational outcomes. The current set of experiments provided strong support for these hypotheses, adding to less explicit, but overall coherent results from previous studies (Bein, Duncan, et al., 2020; Brod et al., 2022; de Bruïne et al., 2025; Greve et al., 2017, 2019; Jainta et al., 2022; Siestrup et al., 2023; Siestrup & Schubotz, 2023; Vlasceanu et al., 2018, 2020). Nevertheless, testing the robustness of these outcomes across stimuli, retrieval intervals, instructions, PE size measurements, and memory tests will be highly valuable.

There are two open questions that are orthogonal to the Hypotheses in Table 1.

First, one issue that arises from the comparison of REPRISE with EMRC by Wahlheim and Zacks (2025) is the role of conscious change detection. While their claim that consciously registering that a change occurred is necessary to avoid interference does not align with predictive processing, it is certainly possible that noticing a mismatch can influence the process. However, it is unclear whether this induces a qualitative change, with different brain regions being involved and biasing representational consequences like a buffering factor, or whether explicitly detected changes simply amount to larger, but qualitatively similar PEs.

As a last caveat, recent research suggests that when faced with new input that overlaps with a previous episode, both integration and separation can happen at different time points of the episode (Liu et al., 2026). This study adds to proposals that different hippocampal subfields (Bein, Duncan, et al., 2020; Frank et al., 2020; Schapiro et al., 2017; Schlichting et al., 2014; Yassa & Stark, 2011) and different brain areas (Bein, Reggev, et al., 2020) may support integration and separation simultaneously. It should be stressed that Liu et al. (2026) created a situation where similar memories were not mutually incompatible, but overlapping. They presented an episode with two elements (first A, then B, then both together). Later, one of the elements was shown again, but with a new one (C and B). In this case, no prediction is triggered by element C, as it is novel, and element B occurring with it is not incompatible with element A still being relevant. In prediction tasks, however, the causal structure inherent to dialogues or stories triggers the prediction of the target when faced with the cue (A → B). If an alternate version of the target is played instead (A → B'), this will cause a PE. Whether this co-occurrence of integration and separation can also follow PEs is therefore yet to be established.

4.5.4 Strengths of REPRISE

While being mindful of the open questions within REPRISE that should be clarified by future research, I would argue that the model offers a comprehensive and parsimonious view on episodic PEs. Notably, it is only the second framework for specifically episodic PEs, after the suggestion by Wahlheim and Zacks (2025). Several other studies have relied on explanations such as reconsolidation or dopamine (see chapter 4.4), applied from conditioned or semantic PEs.

As a first strength, REPRISE is parsimonious. It does not require assumptions about molecular processes as reconsolidation theory does (section 4.4.1) and is

agnostic towards different basic theories of episodic memory encoding and retrieval. Instead, it is grounded in a substantial body of work investigating the basic mechanisms of memory integration or separation. These results are directly incorporated into the model and related to PE size.

Second, REPRISE is a novel contribution, but it incorporates well-established theoretical frameworks, bringing them together and applying them to episodic PEs. It includes the three components of PE size from PIMMS, PE size-dependent consequences for the original target from LCT (Gershman et al., 2017), and encoding of the mismatch target from SLIMM (van Kesteren et al., 2012).

Third, REPRISE offers a graded and dynamic view on mnemonic consequences of PEs. PE size is seen to crucially contribute to outcomes, and further dynamically moderated by post-PE buffering factors. This is a departure from the tacit assumption within many previous studies that all PEs are the same.

Fourth, as will be demonstrated in the next section, REPRISE can help understand why results have sometimes differed across studies. It offers an account for why some studies found no memory enhancement of PEs and can therefore also help to inform decisions about experimental designs.

4.5.5 Predictive Reinstatement as the Gateway to Prediction Errors

Not all previous studies have found a memory enhancement of PEs. In addition to a potential involvement of PE size and buffering factors, it is possible that studies have failed at the first hurdle, namely at producing predictions. REPRISE has two requirements at this crucial gateway: First, there has to be a prediction at all – if the prior is essentially flat, only a minimal, non-consequential PE will arise (Henson & Gagnepain, 2010; Quent et al., 2021). Second, for effects to be captured, this prediction needs contain what experimenters are testing for, so an episodic memory in the field of episodic PEs. There are at least three ways how prior studies may have failed to satisfy these requirements.

First, if participants do not learn the original episodes or rules, the prediction that the experimenter expects and attempts to measure during re-exposure to the cue will not be made. While many experiments tested whether original associations were encoded (or used piloting, as did the studies presented here), some studies did not, leaving it unclear whether participants acquired them. This is especially problematic in implicit learning tasks or when complex sequences are involved (e.g., Jano et al.,

2024; Ortiz-Tudela et al., 2023 for non-episodic PEs).

Second, some studies warn participants that they are about to experience the same stimuli as before, but with changes (Brod et al., 2015) or make change detection the task for participants (Long et al., 2016; Stawarczyk et al., 2020; Wahlheim & Zacks, 2019). In these cases, participants know that their predictions will often be false, preparing the brain for more uncertainty, which will make the prior considerably flatter, reducing PE size and its consequences.

Third, it is an intriguing, but to my knowledge hitherto undiscussed possibility that priors can be flattened by constraining task demands. In conditioned PEs (section 1.3.1), it has already been argued that explicit predicting can modulate the effects of signed compared to unsigned PEs (H. Chen et al., 2025; Ergo et al., 2020). In episodic and semantic PEs, it is noticeable that among those studies not finding memory-enhancing effects of PEs, many asked participants to make predictions – and constraining the possibilities of doing so naturally.

For instance, two such studies (Patel et al., 2022; Udeogu et al., 2022) presented participants with knowledge about personas. They were then asked to predict how these personas would behave in a novel scenario. Crucially, participants had to choose which of two given options was more likely. The presentation of these options alone may have caused an unintended PE if they did not align with the predictions generated automatically. Furthermore, potentially rich and detailed automatic predictions could not be maintained, being interrupted by the necessity to make a binary choice. This likely flattened the prior, as it is unlikely that the brain replaced the generated scenario in favour of an imposed option. Frustration with the task may additionally have led to an abandoning or weakening of predicting.

A similar example is the study by Hermann et al. (2021), where participants encoded naturalistic movies. In a later encoding run, these videos were paused and participants were asked to choose one of two continuations. One of the options was the same continuation that was shown in the video before (i.e., the original memory), and the other was a new continuation. This task may be frustrating, as a participant with perfect memory would have to choose between selecting the old continuation each time (knowing that this will be incorrect half of the time) and responding fully randomly. In both cases, the prior will likely be flattened by an abandoning of automatic prediction, frustration, and high uncertainty.

Similar problems occur when complex, abstract, or imperfect[10] associations are paired with a need for explicit prediction (Ortiz-Tudela et al., 2023; Turan et al., 2023). Here, another factor may be that participants develop a motivation to be right. When they find out that they are not, this can trigger a motivated buffering of generated PEs, constraining their effect on memory (H. Chen et al., 2025; Pupillo et al., 2023).

Explicit responding is not always problematic, however. In the studies by Brod and colleagues (Brod et al., 2018, 2022; de Bruïne et al., 2025), participants predicted numerical general-knowledge facts, but in a natural, non-constraining format (e.g., percentage of animal species that are insects[11]). Similarly, even an explicit recall of the originally encoded episodes does not weaken the effects of PEs, as long as it is phrased to participants not as a prediction that can be right or wrong (Stawarczyk et al., 2020). This study is strong support for explicit predicting being problematic: Even though the two studies use the same stimuli, design, and memory test, Stawarczyk et al. (2020) find a memory-enhancing effect of PEs, while Hermann et al. (2021) do not. The main difference is that participants are asked to recall the original memories in the former study but are asked to make a constraining prediction in the latter. This strongly underscores that this artificial task demand can interrupt naturalistic predicting.

Beyond these three points, it should be noted that a wealth of other influences can affect prior precision. Priors are adapted to the certainty of the environment (Rouhani et al., 2018) and reflect relative instead of absolute value (Bunzeck et al., 2010; but see Mason et al., 2016 for a failed replication). They may also scale with importance of the situation: Arguably, in fear conditioning paradigms, participants may have a pressing interest to anticipate whether they are about to receive an unpleasant electrical shock (Kalbe & Schwabe, 2020; Loock et al., 2025). Similarly, naturalistic and engaging stimuli (Brod et al., 2022; Sinclair & Barense, 2018; Stawarczyk et al., 2020; D. K. Varga et al., 2025) may elicit higher attention and stronger priors than arbitrary sequences of neutral images.

In summary, differences between studies can often be explained by a lack of predictive episodic reinstatement, which is the first and crucial gateway for any

[10] For instance, in Ortiz-Tudela et al. (2023), stimulus-outcome mappings are not 100% determined, but have contingency levels of 80% at most. The inherent randomness in such associations can be frustrating when predicting explicitly: If a participant knows that the beach will be followed by household items in only 80% of cases, they will probably still predict it each time, but knowing – and arguably (especially if monetary incentives are involved) frustratingly so – that the outcome will not always materialise even after essentially perfect behaviour.

[11] To allow reading on without the tension of not knowing the answer: It's around 70%.

mnemonic effects of PE to occur within the REPRISE framework.

4.6 Future Research

The study of episodic PEs is still relatively young and there are pressing questions yet to be addressed. I will point out some of these, starting with further investigation into the basic mechanisms within the REPRISE framework.

4.6.1 Testing REPRISE

Despite the theoretical grounding and support from empirical studies, the hypotheses of REPRISE listed in Table 1 (section 4.5.3) should be explicitly tested. Two broad approaches seem especially promising for this.

First, a series of studies could focus on the behavioural outcomes of PEs of different sizes. In a first step, it would be useful to validate the use of compound measures of PEs, especially skin conductance (Kalbe & Schwabe, 2020, 2021; Loock et al., 2025) and pupil dilation (Brod et al., 2018, 2022; Kafkas & Montaldi, 2015; Theobald & Brod, 2021). As these measures have hitherto been used mainly for conditioned and semantic PEs (Stemerding et al., 2022), it should be tested if they also reflect episodic PE size.

If these physiological measures are shown to reflect episodic PE size, future studies would not necessarily need to induce prior precision or obtain participant-ratings of prior accuracy anymore. Instead, they could capture natural trial-to-trial fluctuations of PE size. Different studies could then focus on different memory tests, including an inference test where participants are tested on associations between objects from the original and mismatch targets (given that they were never seen together, more accurate and faster responses are a sign of integration; Zeithamova et al., 2010), and a mnemonic similarity task where the true target is presented alongside perceptually similar lures to test for high-detail memory associated with pattern separation (Bein et al., 2021; Frank et al., 2020). Together, these tests would allow stronger conclusions about different representational outcomes after PEs of different size.

The second broad approach would be to leverage the high temporal resolution of EEG. Multivariate analyses on EEG data allow testing the unique contributions of predictive and reactive reinstatement, and assessing whether larger PEs lead to more reactive reinstatement. The use of image sequences rather than more complex stimuli

would allow high-precision measuring of representations before PEs as a baseline, predictive and reactive reinstatement during PEs, and representational outcomes after PE induction. This approach could be paired with any memory test, possibly informed by the findings of the previous line of work. Such studies would therefore amount to refining the REPRISE framework and better understanding how episodic PEs influence episodic memory.

4.6.2 The Predictive Mix

An intriguing and largely unaddressed question is how the "predictive mix" will be shaped in a given situation. While experimental findings, multivariate techniques such as RSA, and predictive looking (G. Kim et al., 2014; Long et al., 2016; D. K. Varga et al., 2025; Wahlheim et al., 2021) provide solid evidence that episodic content is used for prediction, this evidence was gathered in experiments where previously encoded episodes were the most effective path to prediction by design. In the face of theoretical accounts questioning a neat distinction between memory systems (Brigard et al., 2022; Renoult et al., 2019; Rubin, 2021) and due to the importance of semantic memory even in episodic recall (Cheng, 2024; Cheng et al., 2016), it may be worth explicitly investigating the interaction between several sources of prediction. Intuitively, this happens in everyday situations: When unwrapping a gift from a relative, one may predict what it will be, based on semantic knowledge about them, but also based on an episodic memory of last year's gift. Additionally, the hedonic value of the gift may be predicted alongside its content.

A full assessment of the contribution of episodic memory to naturalistic prediction would require the use of RSA with nothing less than all of a person's previous experiences. However, even without this admittedly ambitious approach, the question of the predictive mix can be partially addressed. For instance, if the recently demonstrated selectivity of the hippocampus for episodic mismatches is shown to be robust (D. K. Varga et al., 2025), hippocampal involvement can be seen as a correlate of episodic prediction. Another partial approximation within the current paradigm would be measuring not only how much original and mismatch targets differ in terms of meaning, but also in valence, akin to the hedonic outcome in conditioned PEs (section 1.3.1). Naturally, people behave not only differently than they used to but can also act better or worse than expected. Obtaining valence difference ratings between original and mismatch targets and regressing memory outcomes on them, in addition

to the previously used difference ratings targeting meaning, can help to elucidate unique contributions of episodic and valence-based PEs to memory outcomes. Similarly, sporting events (Antony et al., 2021, 2023) or elections (Boeltzig et al., 2026) are not just different, but also better or worse than predicted, making this another area for naturalistic future research.

A related question is which episodic memories will be selected for prediction in a given situation. It is plausible that memories with high overlap with the current situation will be automatically retrieved (Berntsen, 1996, 2009; Berntsen et al., 2013; Moccia et al., 2025), and thereby used for prediction. However, several memories can be relevant, and it is not currently clear whether inhibitory control (Anderson, 2003; Benoit et al., 2015) is employed to select only one at a time, or whether several memories can contribute to a prediction together. This could be tested in paradigms where more than one previously encoded episode is rendered relevant by the cue, while employing RSA to test which episode is reinstated, and univariate analyses to assess whether inhibitory control is deployed.

4.6.3 Post-Prediction Error Buffering Factors

The tacit assumption of most previous studies has been that a PE signal, once generated, will always trigger its mnemonic consequences. However, this may not be the case as it is conceivable that PE signals are modified after being generated. Therefore, REPRISE has a place for buffering factors that can weaken or eliminate PE signals, or influence which representational consequences they trigger.

One example is a study by de Bruïne et al. (2025) showing that even if a PE is evoked and measured in pupil dilation, it does not trigger memory effects if participants judge the information as implausible. This cannot be explained by prior precision or prior accuracy, as these would influence the PE signal itself, and must instead be due to post-PE processes. Similarly, PE size has been shown to have different effects depending on whether the prediction is correct. Specifically, if participants are wrong, a larger PE leads to less learning than a weaker PE (Pupillo et al., 2023). This could be explained by motivational factors, as participants may reject information that proved them wrong.

Many studies, including Studies 2 and 3, may have generated data on this phenomenon without analysing it. While these studies focussed on the time period of the incoming mismatch, it is possible that neural correlates after this period can

predict memory as well. Study 3 is a good example for this: The incoherent episodes may have elicited the large PE that they were designed to elicit, but processes after the input, such as a plausibility judgment (de Bruïne et al., 2025) may have buffered the behavioural effect of these PEs.

Similarly to plausibility, source credibility may play a role and make the process susceptible to ingroup bias. There are self-serving biases which distort memories in order to protect a positive view of oneself (Schacter et al., 2023). Importantly, as central identities are thought to merge with the self (Reese & Whitehouse, 2021), even public events such as elections are misremembered to protect a positive view of one's ingroup (Boeltzig et al., 2026). This would also imply that if someone from our ingroup behaves badly, we would be motivated not to remember that episode later. However, it is likely that a PE was evoked if the identity-based expectation was that ingroup members should behave well. Conversely, we could be inclined to protect our negative view of the outgroup, which would make us motivated not to learn from the large PE generated by an outgroup member behaving well. It is therefore an interesting question whether these social PEs are buffered after their generation or if they are encoded, but distorted over time or at retrieval. A study investigating this would directly apply the notion of PEs to the societal problem of polarization and help elucidate its underlying cognitive mechanisms.

A related open question is how inter-individual differences affect the process. Research on this is still scarce, but there have been suggestions that individuals higher on anxiety and neuroticism, and lower on risk taking and extraversion generally show less PE-induced learning (Kafkas et al., 2025). Similarly, people high on right-wing authoritarianism have shown less belief updating in one study (Sinclair et al., 2020), while the evidence was mixed in another (Vlasceanu et al., 2020). Additionally, higher depressive symptoms are associated with more learning from negative, but less learning from positive reward PEs (Rouhani & Niv, 2019). The related literature of event segmentation has also shown considerable inter-individual differences (Sava-Segal et al., 2023), including due to age (M. E. Smith et al., 2024), hinting at potential differences in the threshold of latent cause inference. Future research should therefore continue investigating which inter-individual factors and cognitive capabilities can moderate PE effects. Once critical factors have been robustly identified, neuroimaging studies can test whether they impact predictive reinstatement, PE detection, latent cause inference, or act as buffering processes.

4.6.4 Bridging the Gap to Future Thinking

A last broad area for future research is the quest to apply the findings from PE-based memory consequences to episodic future thinking. While predictive processing plays out on the time scale of milliseconds or seconds and without conscious awareness, future thinking is the process in which people imagine scenarios about events that are minutes or years into the future (Cole & Kvavilashvili, 2021; Schacter et al., 2017; K. K. Szpunar, 2010; P. M. Szpunar & Szpunar, 2016). While there is a growing body of research on this phenomenon, few studies follow up on these future simulations to assess what happens when these predictions are compared to reality. It is possible that something akin to a PE arises when simulations were incorrect, and that this could lead to memory outcomes that parallel those of the PEs studied in this thesis.

One study supporting the idea that principles of episodic PEs can be applied to future thinking asked basketball fans to list memorable seasons, games, and even individual plays. Using fine-grained analytical game data and betting odds, researchers computed how surprising these events were at the time and found that more surprising events were more likely to be recalled later (Antony et al., 2023). Following up on this study, a valuable next step would be to use participants' individual predictions and surprise instead of the "objective" unexpectedness. This can be realized in a longitudinal design where participants are asked to predict sports results, exam grades, elections, or their assessment of personal events. Then, the event itself would be observed (if it is a public event) or assessed by participants, allowing to compare prediction with reality. Based on the research presented here, event memory should be highest when aspects of the events were most unpredicted.

4.7 Implications

Preceding these potential avenues for future research, there are immediate implications of the findings presented in this thesis, both for experiments in the area of episodic PEs, and for practical applications with societal relevance.

4.7.1 Implications for Research

There are three clear recommendations emerging from this thesis for future studies. First, considering the strong influence of PE size on memory outcomes, future studies should not assume that all PEs they elicit are equal. Measuring PE size directly

via physiological measures (de Bruïne et al., 2025; Loock et al., 2025), with behavioural ratings of the material (Sinclair & Barense, 2018; Vlasceanu et al., 2018), or using measures such as large language model-based semantic similarity (Reimers & Gurevych, 2019) is therefore paramount. Naturally, this will not eliminate the problem that comparisons across studies are constrained by the absence of an absolute measure of PE size (Bunzeck et al., 2010; Rouhani et al., 2018). This problem can only be approached by careful deliberation, driven by theory, manipulation checks, and an evaluation of the paradigms, possibly informed by REPRISE.

Second, manipulation checks are needed to confirm that participants indeed use the induced associations for prediction. As incorporated in REPRISE, without a prediction, no PE-based memory effects can be observed. In episodic PEs, pilot testing can ensure that participants encode events, or participants can be tested on original memory before mismatches are induced. Alternatively, in video stimuli, eye tracking can be used to exploit predictive looking, the tendency of fixating on the area that is predicted to become important in the next instant (Eisenberg et al., 2018; Wahlheim et al., 2021). Response times in classification tasks (Greve et al., 2017) or reading durations (Haeuser & Kray, 2021, 2022) can also indicate that predictions are being formed. Lastly, multivariate techniques that directly track episodic reinstatement (Studies 2 and 3; Bein, Duncan, et al., 2020; G. Kim et al., 2014, 2017) are the gold standard in this area and can provide convincing evidence that participants are predicting based on episodic memories.

Third, the task that participants are given should be chosen carefully. As argued in section 4.5.5, conscious and restrictive predictions or explicit change detection can disrupt naturalistic prediction and should therefore be avoided if possible. In our experiments, we gave neither encoding instructions nor warnings before the changes. It can be argued that this is the best approach when attempting to mirror naturalistic prediction.

4.7.2 Wider Applications

The finding that moderate PEs lead to integration of new information into existing memories and that large PEs lead to separate encoding can be leveraged for several memory-based real-world problems. These differ in whether integration or separate encoding is desirable.

In classroom teaching, it may usually be ideal to encode distinct memories. For

instance, if a student thinks that the angles in a triangle add up to 120°, little is to gain from integrating the truth (180°) with the previous idea – the new information simply needs to be encoded as it is. This idea is supported by a series of studies showing increased fact learning after larger PEs (Brod, 2021; Brod et al., 2013, 2018, 2020, 2022; de Bruïne et al., 2025; Theobald et al., 2022; Theobald & Brod, 2021). These studies suggest that forming a stronger prior by requiring guesses from students is an effective way of inducing stronger PEs (Brod, 2021; Vlasceanu et al., 2020). These predictions can compensate for an otherwise weak prior (Brod et al., 2022), at least if naturalistic and unconstrained prediction is possible (see section 4.5.5). However, it can also happen that the prediction itself becomes a false memory and hinders updating (Bulatova & Fukuda, 2025), which could be more likely when it is important to be correct. Future research is therefore needed to establish under which circumstances active predicting can benefit or hinder learning.

As an alternative strategy, one could exploit the possibility that PEs enhance memory not just for the PE-evoking detail, but potentially also for stimuli before, after, or peripheral to the PE (Congleton & Berntsen, 2020b; Greve et al., 2019; Loock et al., 2025; Rouhani & Niv, 2021). The embedding of the fact in a surprising situation (based on visual aids, oral delivery, or a story) can therefore potentially enhance learning.

Another practical application that has been explored in the context of PEs is misinformation correction (Kemp et al., 2022, 2024). Here, it is possible that moderate PEs are more successful in inducing changes, so that the base of knowledge and experiences is slowly updated over time as new information is being integrated. Separate encoding after large PEs may mean that people know that there is a different view, but that this isolated episode does not feed back into attitudes or behaviour.

However, prior accuracy cannot be targeted to create moderate PEs, as it would defeat the purpose to make the correction message less true so that it aligns better with existing beliefs. Instead, prior precision can be leveraged, making priors less sharp or post-PE buffering factors can be removed by making the context less confrontational, using trustworthy sources, and framing the topic as a less contested. In fact, this tallies with previously identified successful strategies to counter polarization such as emphasizing shared identities with outgroups or showcasing positive intergroup contact (Voelkel et al., 2024).

Lastly, PEs may explain memory inaccuracies in eyewitness testimony. Conversations can be a major driver of memory change, and elements can be added or

deleted from a memory when talking to others who also witnessed the event (Andrews & Rapp, 2014; Coman & Hirst, 2015; Granhag et al., 2012; Hirst & Echterhoff, 2012). While the two processes of adding and deleting details have been attributed different processes, both can be explained by predictive processing: Items that are predicted based on a previous memory but are not mentioned by the conversation partner may be weakened (Vlasceanu et al., 2018), while unpredicted items that they mentioned are likely to be encoded. Moderate PEs, induced by plausible but false items (Loftus, 2005; Loftus & Pickrell, 1995), may therefore lead to an integration of new elements, and thus pose a larger risk for evoking false memories.

5 Conclusion

The study of episodic PEs is still in its relative infancy. In its development, the field has often been informed and inspired by the more established conditioned and semantic PEs. However, the evidence reviewed in this thesis underscores that the complexity of episodic prediction does not allow a simple translation from one type of PE to another. Instead, to fully appreciate mnemonic outcomes after episodic PEs, it is necessary to combine basic principles of predictive processing with empirically supported episodic memory operations.

In this vein, I have argued that PEs are not like happy families under the Anna Karenina principle: They are not all alike, and recognizing their differences is a crucial step to understanding how they impact episodic memory. Specifically, I have attempted to demonstrate that PE size is a decisive factor in determining memory outcomes and that a trade-off between integration and separation underlies these effects. Across three studies, using behavioural analyses as well as univariate and multivariate fMRI techniques, we have shown evidence for the idea that small PEs trigger no changes to memories, moderate PEs lead to the integration of a new event into similar memories, while large PEs lead to separate encoding. Based on these studies and the existing literature, I have proposed the REPRISE model, which offers a framework for understanding the effects of different PEs and how experimental manipulations can affect the process. Its suggestions should inform best practices for future research, including the quantification of the PE, testing both original and mismatch targets, and allowing naturalistic, unconstrained predicting.

The fact that PEs can modify memories and trigger new encoding is fascinating, given that predicting happens continuously without our conscious awareness. PEs are therefore a major driver of memory change, possibly more omnipotent than many other mechanisms, which involve conscious retrieval. The power of PEs is still underappreciated in the field of episodic memory and their relevance for classroom learning, eyewitness testimony, and misinformation warrants further research.

On a broader note, it may be comforting to know that while we are often frustrated by our mistakes, errors are a fundamental element of cognition. Far from being a regrettable sign of malfunction, they are a necessary source for learning. Thus, the information they encapsulate makes errors the crucial mediator between our minds and the ever-changing world around us.

Supplementary Materials

S1: Analysis of Difference Ratings Across Versions

When pulling together the in-lab participants in Studies 1 and 2, and the participants of the online difference ratings obtained for Study 2, each dialogue version was rated by 33-37 (M = 33.17, SD = 1.09) participants. These ratings were used to test whether the conditions to which we assigned each dialogue version actually produced a hierarchy of difference ratings along condition lines.

In an ANOVA with dialogue condition (surface low, surface high, gist low, gist high) as the sole factor, condition was significant, $F(3, 140) = 154.00$, $p < .001$. Tukey-corrected post-hoc tests found significant differences in all pairwise comparisons (ps < .001), following the intended hierarchy, with the exception of surface low and surface high conditions, which were not significantly different from each other (p = .230). Summary statistics are shown in Table S1, and Figure S1 shows condition differences in each dialogue that was written (note that this only includes the 36 dialogues used in Studies 1 and 2).

This analysis confirms that gist changes were rated as more impactful than surface ratings and that this behavioural measure was not sensitive for detecting surface changes of different magnitude. Note that the distinction between surface and gist was only used in Study 1 and that no study used the four pre-assigned dialogue conditions but employed the difference ratings instead.

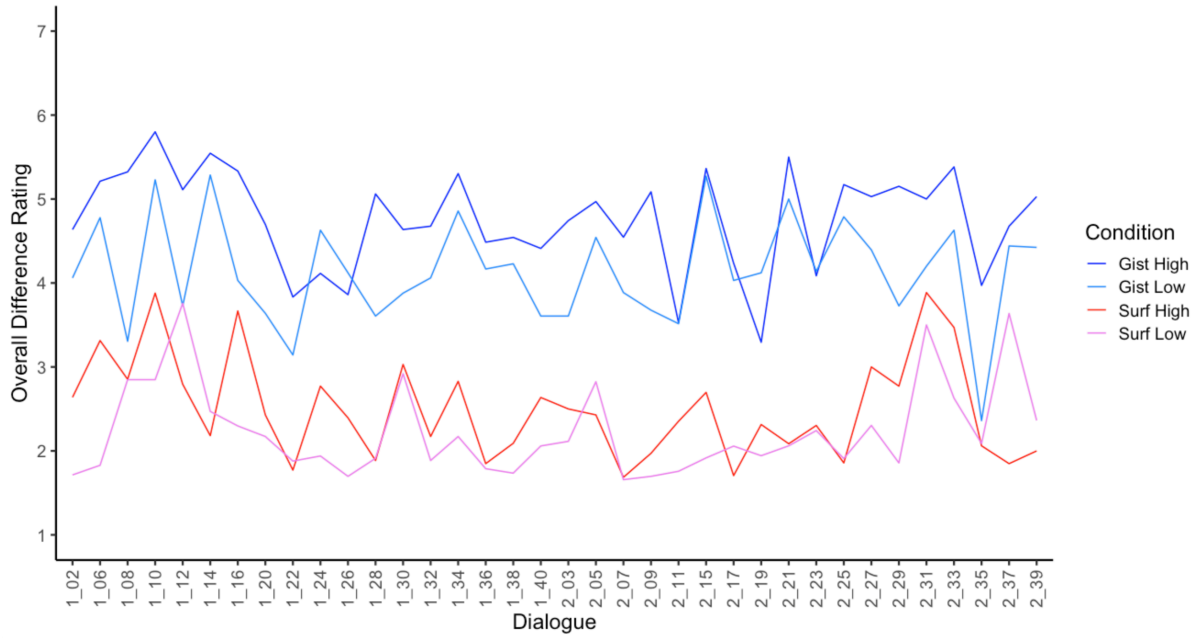
Table S1

Difference rating in each modification condition

Condition	*M*	*SD*
Surface low	2.24	0.55
Surface high	2.50	0.60
Gist low	4.14	0.63
Gist high	4.76	0.60

Figure S1

Mean difference ratings of each dialogue condition



Note. "Surf" in the legend is abbreviated for Surface. The dialogue numbers are largely arbitrary. The rating for each dialogue is based on 33-37 participants, 20 of those from the online rating study, the others from the lab-based Studies 1 and 2.

S2: Analysis of Cued Recall Versions

In Studies 1 and 2, after the recognition test, we also conducted a cued recall test, which was not analysed in the presented papers. Participants were presented with the beginning of the dialogue and were asked to verbally reproduce the target, so the part of the dialogue that they had heard in two versions. Regardless of recall accuracy, the data presented here pertains to how many recall attempts participants made. After a first recall, they could choose to also recall the other version. If they did not produce either version, these trials are coded with a zero.

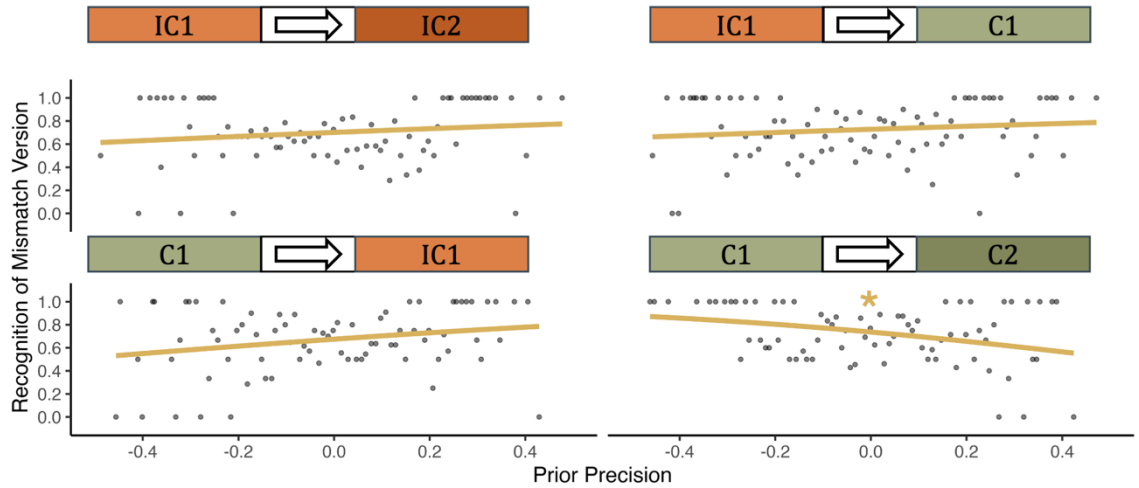
In both studies, higher difference ratings were clearly associated with more recall attempts (Study 1: $b = 0.07$, $SE = 0.01$, $t(802.85) = 7.14$, $p < .001$; Study 2: $b = 0.08$, $SE = 0.01$, $t(1046) = 11.77$, $p < .001$). This is in line with more distinct encoding after larger PEs: Participants were more aware of having encountered two versions and therefore attempted to recall both versions. This data is also presented in Pronoza et al. (2025).

S3: Behavioural Effects of Prior Precision in Study 3

In Study 2, prior precision enhanced original and mismatch recognition memory, while in the study by G. Kim et al. (2014), it decreased it. Unpublished data from Study 3 rather finds the latter pattern for the mismatch target. When letting prior precision in the IFG interact with the coherence status of both original and mismatch targets, there is a three-way interaction (b = -3.16, SD = 1.21, OR = 0.04, p = .009). Figure S2 depicts this interaction, showing the effect of prior precision depending on coherence status of the two targets. While there is a mild, non-significant positive association between prior precision and mismatch memory for three of the situations, only the negative effect of prior precision on memory if both targets are coherent reaches significance in Tukey-corrected post-hoc tests (p = .024). This situation is the same as in Study 2: A coherent original dialogue version is replaced by a different coherent dialogue version. As discussed in the text, the higher uncertainty and weaker encoding of dialogues in Study 3 may have led to this negative effect. Together, Studies 2 and 3 may therefore model the full U-shape of prior precision. Note that there were no effects of prior precision on original recognition in this study.

Figure S2

The effect of prior precision on mismatch target recognition in Study 3



Note. The plot depicts original mismatch recognition (0 = not recognised, 1 = recognised) as a function of prior precision, measured as original predictive reinstatement in the IFG. This is shown for each of the combinations of conditions between original and mismatch version (IC = inconsistent, C = consistent). Data is aggregated per participant and prior precision level.

List of Abbreviations

BA	Brodmann Area
EEG	Electroencephalography
EMRC	Event Memory Retrieval and Comparison Theory (Wahlheim & Zacks, 2025)
fMRI	Functional magnetic resonance imaging
IFG	Inferior Frontal Gyrus
LCT	Latent Cause Theory (Gershman et al., 2017)
mPFC	Medial Prefrontal Cortex
MTL	Medial Temporal Lobe
PE	Prediction Error
PIMMS	Predictive, Interactive Multiple Memory Systems (Henson & Gagnepain, 2010)
REPRISE	Reactivation and Prediction causing Integration and Separation in Episodic Memory
SLIMM	Schema-Linked Interaction Between Medial Prefrontal and Medial Temporal Regions (van Kesteren et al., 2012)
vmPFC	Ventromedial Prefrontal Cortex

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? *Psychonomic Bulletin & Review*, *30*(6), 2296–2304. https://doi.org/10.3758/s13423-023-02264-8

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

Agren, T. (2014). Human reconsolidation: A reactivation and update. *Brain Research Bulletin*, *105*, 70–82. https://doi.org/10.1016/j.brainresbull.2013.12.010

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. https://doi.org/10.1016/j.jml.2003.08.006

Anderson, M. C., & Hulbert, J. C. (2020). Active forgetting: Adaptation of memory by prefrontal control. *Annual Review of Psychology*, *72*(1), 1–36. https://doi.org/10.1146/annurev-psych-072720-094140

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

Antony, J. W., Dam, J. V., Massey, J. R., Barnett, A. J., & Bennion, K. A. (2023). Long-term, multi-event surprise correlates with enhanced autobiographical memory. *Nature Human Behaviour*, *7*(12), 2152–2168. https://doi.org/10.1038/s41562-023-01631-8

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougle, S. D., & Norman, K. A. (2021). Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *Neuron*, *109*(2), 377–390. https://doi.org/10.1016/j.neuron.2020.10.029

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

Baker, B., Lansdell, B., & Kording, K. P. (2022). Three aspects of representation in neuroscience. *Trends in Cognitive Sciences*, *26*(11), 942–958. https://doi.org/10.1016/j.tics.2022.08.014

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Barbeau, E. J., Chauvel, P., Moulin, C. J. A., Regis, J., & Liégeois-Chauvel, C. (2017). Hippocampus duality: Memory and novelty detection are subserved by distinct mechanisms. *Hippocampus*, *27*(4), 405–416. https://doi.org/10.1002/hipo.22699

Barnett, A. J., Nguyen, M., Spargo, J., Yadav, R., Cohn-Sheehy, B. I., & Ranganath, C. (2024). Hippocampal-cortical interactions during event boundaries support retention of complex narrative events. *Neuron*, *112*(2), 319–330. https://doi.org/10.1016/j.neuron.2023.10.010

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

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

Bauer, P. J. (2021). We know more than we ever learned: Processes involved in accumulation of world knowledge. *Child Development Perspectives*, *15*(4), 220–227. https://doi.org/10.1111/cdep.12430

Bauer, P. J., Esposito, A. G., & Daly, J. J. (2020). Self-derivation through memory integration: A model for accumulation of semantic knowledge. *Learning and Instruction*, *66*, Article 101271. https://doi.org/10.1016/j.learninstruc.2019.101271

Bäuml, K.-H. T. (2019). Context retrieval as a critical component in selective memory retrieval. *Current Directions in Psychological Science*, *28*(2), 177–182. https://doi.org/10.1177/0963721419827520

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

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., & Niv, Y. (2025). Schemas, reinforcement learning and the medial prefrontal cortex. *Nature Reviews Neuroscience*, *26*(3), 141–157. https://doi.org/10.1038/s41583-024-00893-z

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

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

Benoit, R. G., Hulbert, J. C., Huddleston, E., & Anderson, M. C. (2015). Adaptive top–down suppression of hippocampal activity and the purging of intrusive memories from consciousness. *Journal of Cognitive Neuroscience*, *27*(1), 96–111. https://doi.org/10.1162/jocn_a_00696

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

Berntsen, D. (1996). Involuntary autobiographical memories. *Applied Cognitive Psychology*, *10*(5), 435–454.

Berntsen, D. (2009). *Involuntary autobiographical memories: An introduction to the unbidden past*. Cambridge University Press.

Berntsen, D., Staugaard, S. R., & Sørensen, L. M. T. (2013). Why am I remembering this now? Predicting the occurrence of involuntary (spontaneous) episodic memories. *Journal of Experimental Psychology: General*, *142*(2), 426–444. https://doi.org/10.1037/a0029128

Bluck, S., Alea, N., Habermas, T., & Rubin, D. C. (2005). A TALE of three functions: The self–reported uses of autobiographical memory. *Social Cognition*, *23*(1), 91–117. https://doi.org/10.1521/soco.23.1.91.59198

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

Boeltzig, M., Schubotz, R., Cole, S., & Rathbone, C. (2026). Self-serving biases shape the relationship between future thinking and remembering of elections. *Communications Psychology*. https://doi.org/10.1038/s44271-026-00423-w

Bonetti, L., Carlomagno, F., Kliuchko, M., Gold, B. P., Palva, S., Haumann, N. T., Tervaniemi, M., Huotilainen, M., Vuust, P., & Brattico, E. (2022). Whole-brain computation of cognitive versus acoustic errors in music: A mismatch negativity study. *Neuroimage: Reports*, *2*(4), Article 100145. https://doi.org/10.1016/j.ynirp.2022.100145

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

Boyle, A. (2022). The mnemonic functions of episodic memory. *Philosophical Psychology*, *35*(3), 327–349. https://doi.org/10.1080/09515089.2021.1980520

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. *Cerebral Cortex*, *32*(22), 5020–5035. https://doi.org/10.1093/cercor/bhab529

Bramão, I., Karlsson, A., & Johansson, M. (2017). Mental reinstatement of encoding context improves episodic remembering. *Cortex*, *94*, 15–26. https://doi.org/10.1016/j.cortex.2017.06.007

Brigard, F. D. (2014). The nature of memory traces. *Philosophy Compass*, *9*(6), 402–414. https://doi.org/10.1111/phc3.12133

Brigard, F. D., Umanath, S., & Irish, M. (2022). Rethinking the distinction between episodic and semantic memory: Insights from the past, present, and future. *Memory & Cognition*, *50*(3), 459–463. https://doi.org/10.3758/s13421-022-01299-x

Brod, G. (2021). Predicting as a learning strategy. *Psychonomic Bulletin & Review*, *28*(6), 1839–1847. https://doi.org/10.3758/s13423-021-01904-1

Brod, G., Breitwieser, J., Hasselhorn, M., & Bunge, S. A. (2020). Being proven wrong elicits learning in children – but only in those with higher executive function skills. *Developmental Science*, *23*(3), Article e12916. https://doi.org/10.1111/desc.12916

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

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

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

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The influence of prior knowledge on memory: A developmental cognitive neuroscience perspective. *Frontiers in Behavioral Neuroscience*, *7*, Article 139. https://doi.org/10.3389/fnbeh.2013.00139

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

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

Brunec, I. K., Robin, J., Olsen, R. K., Moscovitch, M., & Barense, M. D. (2020). Integration and differentiation of hippocampal memory traces. *Neuroscience & Biobehavioral Reviews*, *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. *Frontiers in Human Neuroscience*, *4*, Article 25. https://doi.org/10.3389/fnhum.2010.00025

Bulatova, O., & Fukuda, K. (2025). Prediction-based false memory: Unconfirmed prediction can result in robust false memories. *Cognition*, *255*, Article 106013. https://doi.org/10.1016/j.cognition.2024.106013

Bunzeck, N., Dayan, P., Dolan, R. J., & Duzel, E. (2010). A common mechanism for adaptive scaling of reward and novelty. *Human Brain Mapping*, *31*(9), 1380–1394. https://doi.org/10.1002/hbm.20939

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

Carpenter, A. C., & Schacter, D. L. (2017). Flexible retrieval: When true inferences produce false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *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. *Journal of Experimental Psychology: General*, *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. *Psychology and 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. *Nature Communications*, *10*(1), Article 5363. https://doi.org/10.1038/s41467-019-13377-x

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among spatial memories triggers repulsion of hippocampal representations. *Current Biology*, *27*(15), 2307–2317. https://doi.org/10.1016/j.cub.2017.06.057

Chen, H., Claeys, W., Ergo, K., Calderon, C. B., & Verguts, T. (2025). Robust effects of signed prediction error on declarative memory. *PsyArXiv*.

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

Chen, Y., Zhao, C., Liu, R., & Li, Q. (2025). Unexpected feedback enhances episodic memory: Exploring signed and unsigned reward prediction errors with EEG. *NeuroImage*, *316*, Article 121303. https://doi.org/10.1016/j.neuroimage.2025.121303

Cheng, S. (2024). Distinct mechanisms and functions of episodic memory. *Philosophical Transactions B*, *379*(1913), Article 20230411. https://doi.org/10.1098/rstb.2023.0411

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

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/s0140525x12000477

Clark, A. (2014). Perceiving as predicting. In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and its modalities*. Oxford University Press.

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

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

Coman, A., & Hirst, W. (2015). Social identity and socially shared retrieval-induced forgetting: The effects of group membership. *Journal of Experimental Psychology: General*, *144*(4), 717–722. https://doi.org/10.1037/xge0000077

Congleton, A. R., & Berntsen, D. (2020a). 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

Congleton, A. R., & Berntsen, D. (2020b). The devil is in the details: investigating the influence of emotion on event memory using a simulated event. *Psychological Research*, *84*(8), 2339–2353. https://doi.org/10.1007/s00426-019-01215-0

Congleton, A. R., & Berntsen, D. (2022). How suspense and surprise enhance subsequent memory: the case of the 2016 United States Presidential Election. *Memory*, *30*(3), 317–329. https://doi.org/10.1080/09658211.2021.2013503

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261–288. https://doi.org/10.1037/0033-295x.107.2.261

Danek, A. H., Öllinger, M., Fraps, T., Grothe, B., & Flanagin, V. L. (2015). An fMRI investigation of expectation violation in magic tricks. *Frontiers in Psychology*, *6*, Article 84. https://doi.org/10.3389/fpsyg.2015.00084

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

de Bruïne, A., Tromp, M. V., Koornneef, A., Brod, G., & Jolles, D. (2025). The interactive effects of surprise and plausibility on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *51*(6), 954–967. https://doi.org/10.1037/xlm0001388

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, Article 548. https://doi.org/10.3389/fpsyg.2012.00548

Detre, G. J., Natarajan, A., Gershman, S. J., & Norman, K. A. (2013). Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia*, *51*(12), 2371–2388. https://doi.org/10.1016/j.neuropsychologia.2013.02.017

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

Drascher, M. L., & Kuhl, B. A. (2022). Long-term memory interference is resolved via repulsion and precision along diagnostic memory dimensions. *Psychonomic Bulletin & Review*, *29*(5), 1898–1912. https://doi.org/10.3758/s13423-022-02082-4

Ebbinghaus, H. (1885). *Über das Gedächtnis [Memory. A Contribution to Experimental Psychology]*. Wissenschaftliche Buchgesellschaft.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*(4), 319–345. https://doi.org/10.1016/s0005-7967(99)00123-0

Eisenberg, M. L., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications*, *3*(1), Article 53. https://doi.org/10.1186/s41235-018-0146-z

Elsey, J. W. B., Ast, V. A. V., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin*, *144*(8), 797–848. https://doi.org/10.1037/bul0000152

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

Ergo, K., Loof, E. D., & Verguts, T. (2020). Reward prediction error and declarative memory. *Trends in Cognitive Sciences*, *24*(5), 388–397. https://doi.org/10.1016/j.tics.2020.02.009

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories—The role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*(1), Article 11066. https://doi.org/10.1038/ncomms11066

Fawcett, J. M., & Hulbert, J. C. (2020). The many faces of forgetting: Toward a constructive view of forgetting in everyday life. *Journal of Applied Research in Memory and Cognition*, *9*(1), 1–18. https://doi.org/10.1016/j.jarmac.2019.11.002

Felleman, D. J., & Essen, D. C. V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of prediction error. *Neuroscience & Biobehavioral Reviews*, *68*, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004

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

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

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

Forcato, C., Rodríguez, M. L. C., Pedreira, M. E., & Maldonado, H. (2010). Reconsolidation in humans opens up declarative memory to the entrance of new information. *Neurobiology of Learning and Memory*, *93*(1), 77–84. https://doi.org/10.1016/j.nlm.2009.08.006

Frank, D., Kafkas, A., & Montaldi, D. (2022). Experiencing surprise: The temporal dynamics of its impact on memory. *The Journal of Neuroscience*, *42*(33), 6435–6444. https://doi.org/10.1523/jneurosci.1783-21.2022

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

Friston, K. J. (2019). Waves of prediction. *PLoS Biology*, *17*(10), Article e3000426. https://doi.org/10.1371/journal.pbio.3000426

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

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

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

Gisquet-Verrier, P., & Riccio, D. C. (2018). Memory integration: An alternative to the consolidation/reconsolidation hypothesis. *Progress in Neurobiology*, *171*, 15–31. https://doi.org/10.1016/j.pneurobio.2018.10.002

Gisquet-Verrier, P., & Riccio, D. C. (2019). Memory integration as a challenge to the consolidation/reconsolidation hypothesis: Similarities, differences and perspectives. *Frontiers in Systems Neuroscience*, *12*, Article 71. https://doi.org/10.3389/fnsys.2018.00071

Giuliano, A. E., Bonasia, K., Ghosh, V. E., Moscovitch, M., & Gilboa, A. (2021). Differential influence of ventromedial prefrontal cortex lesions on neural representations of schema and semantic category knowledge. *Journal of Cognitive Neuroscience*, *33*(9), 1928–1955. https://doi.org/10.1162/jocn_a_01746

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Grande, X., Berron, D., Horner, A. J., Bisby, J. A., Düzel, E., & Burgess, N. (2019). Holistic recollection via pattern completion involves hippocampal subfield CA3. *The Journal of Neuroscience*, *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. *Psychology, 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. *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

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge Is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, *148*(2), 325–341. https://doi.org/10.1037/xge0000498

Groves, I., Grella, S. L., Harley, C. W., Hardt, O., & Nadel, L. (2025). How prediction error drives memory updating: role of locus coeruleus–hippocampal interactions. *Trends in Neurosciences*, *48*(11), 865–876. https://doi.org/10.1016/j.tins.2025.09.003

Güler, B., Serin, F., & Günseli, E. (2025). Prediction error is out of context: The dominance of contextual stability in structuring episodic memories. *Psychonomic Bulletin & Review*, *32*, 2957–2968. https://doi.org/10.3758/s13423-025-02723-4

Haeuser, K. I., & Kray, J. (2021). Effects of prediction error on episodic memory retrieval: evidence from sentence reading and word recognition. *Language, Cognition and Neuroscience*, *38*(4), 558–574. https://doi.org/10.1080/23273798.2021.1924387

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

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. *Memory & Cognition*, *49*(7), 1387–1404. https://doi.org/10.3758/s13421-021-01173-2

Hirst, W., & Echterhoff, G. (2012). Remembering in conversations: The social sharing and reshaping of memories. *Annual Review of Psychology*, *63*(1), 55–79. https://doi.org/10.1146/annurev-psych-120710-100340

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, *35*(2), 209–223. https://doi.org/10.1111/mila.12281

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

Horner, A. J., & Doeller, C. F. (2017). Plasticity of hippocampal memories in humans. *Current Opinion in Neurobiology*, *43*, 102–109. https://doi.org/10.1016/j.conb.2017.02.004

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, *2*(1), 105–112. https://doi.org/10.3758/bf03214414

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, *14*(1–2), 47–53. https://doi.org/10.1101/lm.365707

Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: Updating or source confusion? *Memory*, *17*(5), 502–510. https://doi.org/10.1080/09658210902882399

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. *Quarterly Journal of Experimental Psychology*, *74*(3), 413–424. https://doi.org/10.1177/1747021820968483

Jacques, P. L. St., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences*, *110*(49), 19671–19678. https://doi.org/10.1073/pnas.1319630110

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

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

Jang, A. I., Nassar, M. R., Dillon, D. G., & Frank, M. J. (2019). Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*, *3*(7), 719–732. https://doi.org/10.1038/s41562-019-0597-3

Jano, S., Chatburn, A., Cross, Z. R., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2024). How predictability and individual alpha frequency shape memory: Insights from an event-related potential investigation. *Neurobiology of Learning and Memory*, *216*, Article 108006. https://doi.org/10.1016/j.nlm.2024.108006

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. *The Journal of Neuroscience*, *44*(18), Article e1740232024. https://doi.org/10.1523/jneurosci.1740-23.2024

Kafkas, A. (2021). Encoding-linked pupil response is modulated by expected and unexpected novelty: Implications for memory formation and neurotransmission. *Neurobiology of Learning and Memory*, *180*, Article 107412. https://doi.org/10.1016/j.nlm.2021.107412

Kafkas, A., & Montaldi, D. (2015). Striatal and midbrain connectivity with the hippocampus selectively boosts memory for contextual novelty. *Hippocampus*, *25*(11), 1262–1273. https://doi.org/10.1002/hipo.22434

Kafkas, A., & Montaldi, D. (2018). Expectation affects learning and modulates memory experience at retrieval. *Cognition*, *180*, 123–134. https://doi.org/10.1016/j.cognition.2018.07.010

Kafkas, A., Westerman, M., Cleto, K., Sabaityte, K., & Sergi, K. (2025). When personality meets surprise: individual differences in memory for unexpected events. *Frontiers in Psychology*, *16*, Article 1652428. https://doi.org/10.3389/fpsyg.2025.1652428

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

Kalbe, F., & Schwabe, L. (2021). Prediction errors for aversive events shape long-term memory formation through a distinct neural mechanism. *Cerebral Cortex*, *32*(14), 3081–3097. https://doi.org/10.1093/cercor/bhab402

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., 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. *Cognitive Research: Principles and Implications*, *7*(1), Article 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. *Cognitive Research: Principles and Implications*, *9*(1), Article 28. https://doi.org/10.1186/s41235-024-00546-w

Kerrén, C., Bramão, I., Hellerstedt, R., & Johansson, M. (2021). Strategic retrieval prevents memory interference: The temporal dynamics of retrieval orientation. *Neuropsychologia*, *154*, Article 107776. https://doi.org/10.1016/j.neuropsychologia.2021.107776

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

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

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

Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(2), 150–159. https://doi.org/10.1037/0278-7393.3.2.150

Kolibius, L. D., Josselyn, S. A., & Hanslmayr, S. (2025). On the origin of memory neurons in the human hippocampus. *Trends in Cognitive Sciences*, *29*(5), 421–433. https://doi.org/10.1016/j.tics.2025.01.013

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

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

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories. *Psychological Review*, *119*(3), 573–616. https://doi.org/10.1037/a0028681

Lee, K. M., Ferreira-Santos, F., & Satpute, A. B. (2021). Predictive processing models and affective neuroscience. *Neuroscience & Biobehavioral Reviews*, *131*, 211–228. https://doi.org/10.1016/j.neubiorev.2021.09.009

Lee, Y., & Chen, J. (2024). The relationship between event boundary strength and pattern shifts across the cortical hierarchy during naturalistic movie-viewing. *Journal of Cognitive Neuroscience*, *36*(11), 2317–2342. https://doi.org/10.1162/jocn_a_02213

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

Liedtke, N., Boeltzig, M., Siestrup, S., & Schubotz, R. I. (2026). Shaping memory from the start: Initial prediction errors during first encoding. *NeuroImage*, *325*, Article 121660. https://doi.org/10.1016/j.neuroimage.2025.121660

Liu, Z., Johansson, M., & Bramão, I. (2026). Episodic events are flexibly encoded in both integrated and separated neural representations. *Nature Communications*, *17*(1), Article 752. https://doi.org/10.1038/s41467-026-68473-6

Liu, Z., Johansson, M., Johansson, R., & Bramão, I. (2024). The effects of episodic context on memory integration. *Scientific Reports*, *14*(1), Article 30159. https://doi.org/10.1038/s41598-024-82004-7

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

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*(12), 720–725. https://doi.org/10.3928/0048-5713-19951201-07

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

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *The Journal of Neuroscience*, *36*(50), 12677–12687. https://doi.org/10.1523/jneurosci.1850-16.2016

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

Mack, M. L., Love, B. C., & Preston, A. R. (2018). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters*, *680*, 31–38. https://doi.org/10.1016/j.neulet.2017.07.061

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

Mason, A., Ludwig, C., & Farrell, S. (2016). Adaptive scaling of reward in episodic memory: a replication study. *The Quarterly Journal of Experimental Psychology*, *70*(11), 1–25. https://doi.org/10.1080/17470218.2016.1233439

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295x.102.3.419

Moccia, A., Plummer, M., Simpson, I., & Morcom, A. M. (2025). The effects of external cue overlap and internal goals on selective memory retrieval as revealed by electroencephalographic (EEG) neural pattern reinstatement. *European Journal of Neuroscience*, *62*(1), Article e70194. https://doi.org/10.1111/ejn.70194

Molitor, R. J., Sherrill, K. R., Morton, N. W., Miller, A. A., & Preston, A. R. (2020). Memory reactivation during learning simultaneously promotes dentate gyrus/CA 2,3 pattern differentiation and CA 1 memory integration. *The Journal of Neuroscience*, *41*(4), 726–738. https://doi.org/10.1523/jneurosci.0394-20.2020

Monfils, M.-H., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: Key to persistent attenuation of fear memories. *Science*, *324*(5929), 951–955. https://doi.org/10.1126/science.1167975

Mullally, S. L., & Maguire, E. A. (2014). Memory, imagination, and predicting the future. *The Neuroscientist*, *20*(3), 220–234. https://doi.org/10.1177/1073858413495091

Nadel, L., & Hardt, O. (2011). Update on memory systems and processes. *Neuropsychopharmacology*, *36*(1), 251–273. https://doi.org/10.1038/npp.2010.169

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

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227. https://doi.org/10.1016/s0959-4388(97)80010-4

Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, *9*(1), 1–22. https://doi.org/10.1016/0010-0277(81)90011-1

Newman, E. L., & Norman, K. A. (2010). Moderate excitation leads to weakening of perceptual representations. *Cerebral Cortex*, *20*(11), 2760–2770. https://doi.org/10.1093/cercor/bhq021

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

O'Brien, E. (2019). Enjoy it again: Repeat experiences are less repetitive than people think. *Journal of Personality and Social Psychology*, *116*(4), 519–540. https://doi.org/10.1037/pspa0000147

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. *Journal of Experimental Psychology: General*, *152*(8), 2160–2176. https://doi.org/10.1037/xge0001367

Osvath, M., & Johansson, M. (2024). A short natural history of mental time travels: a journey still travelled? *Philosophical Transactions B*, *379*(1913), Article 20230402. https://doi.org/10.1098/rstb.2023.0402

Parks, C. M., Mohawk, K. D., Werner, L. L. S., & Kiley, C. (2022). The time window of reconsolidation: A replication. *Psychonomic Bulletin & Review*, *29*(5), 2008–2013. https://doi.org/10.3758/s13423-022-02102-3

Patel, S. P., McCurdy, M. P., Frankenstein, A. N., Sklenar, A. M., Levy, P. U., Szpunar, K. K., & Leshikar, E. D. (2022). The reciprocal relationship between episodic memory and future thinking: How the outcome of predictions is subsequently remembered. *Brain and Behavior*, *12*(9), Article e2603. https://doi.org/10.1002/brb3.2603

Petersdotter, L., Miller, L., Johansson, M., & Hammar, Å. (2025). Trauma-analogue symptom variability predicted by inhibitory control and peritraumatic heart rate. *Scientific Reports*, *15*(1), Article 15215. https://doi.org/10.1038/s41598-025-99564-x

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

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

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

Pupillo, F., & Bruckner, R. (2023). Signed and unsigned effects of prediction error on memory: Is it a matter of choice? *Neuroscience & Biobehavioral Reviews*, *153*, Article 105371. https://doi.org/10.1016/j.neubiorev.2023.105371

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

Quent, J. A., Greve, A., & Henson, R. N. (2022). Shape of U: The nonmonotonic relationship between object–location memory and expectedness. *Psychological Science*, *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. *Neurobiology of Learning and Memory*, *179*, Article 107382. https://doi.org/10.1016/j.nlm.2021.107382

Rasmussen, A. S., Ramsgaard, S. B., & Berntsen, D. (2015). Frequency and functions of involuntary and voluntary autobiographical memories across the day. *Psychology of Consciousness: Theory, Research, and Practice*, *2*(2), 185–205. https://doi.org/10.1037/cns0000042

Reagh, Z. M., & Yassa, M. A. (2014). Repetition strengthens target recognition but impairs similar lure discrimination: evidence for trace competition. *Learning & Memory*, *21*(7), 342–346. https://doi.org/10.1101/lm.034546.114

Reese, E., & Whitehouse, H. (2021). The development of identity fusion. *Perspectives on Psychological Science*, *16*(6), 1398–1411. https://doi.org/10.1177/1745691620968761

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

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

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: The semantic–episodic distinction. *Trends in Cognitive Sciences*, *23*(12), 1041–1057. https://doi.org/10.1016/j.tics.2019.09.008

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

Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: Insights from functional brain imaging. *Annual Review of Psychology*, *63*(1), 101–128. https://doi.org/10.1146/annurev-psych-120710-100344

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences*, *17*, 114–123. https://doi.org/10.1016/j.cobeha.2017.07.016

Rouhani, N., & Niv, Y. (2019). Depressive symptoms bias the prediction-error enhancement of memory towards negative events in reinforcement learning. *Psychopharmacology*, *236*(8), 2425–2435. https://doi.org/10.1007/s00213-019-05322-z

Rouhani, N., & Niv, Y. (2021). Signed and unsigned reward prediction errors dynamically enhance learning and memory. *eLife*, *10*, Article e61077. https://doi.org/10.7554/elife.61077

Rouhani, N., Niv, Y., Frank, M. J., & Schwabe, L. (2023). Multiple routes to enhanced memory for emotionally relevant events. *Trends in Cognitive Sciences*, *27*(9), 867–882. https://doi.org/10.1016/j.tics.2023.06.006

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

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, Article 104269. https://doi.org/10.1016/j.cognition.2020.104269

Rowe, E. G., Tsuchiya, N., & Garrido, M. I. (2020). Detecting (un)seen change: The neural underpinnings of (un)conscious prediction errors. *Frontiers in Systems Neuroscience*, *14*, Article 541670. https://doi.org/10.3389/fnsys.2020.541670

Rubin, D. C. (2021). A conceptual space for episodic and semantic memory. *Memory & Cognition*, 1–14. https://doi.org/10.3758/s13421-021-01148-3

Rugg, M. D., & Renoult, L. (2025). The cognitive neuroscience of memory representations. *Neuroscience & Biobehavioral Reviews*, *179*, Article 106417. https://doi.org/10.1016/j.neubiorev.2025.106417

Sava-Segal, C., Richards, C., Leung, M., & Finn, E. S. (2023). Individual differences in neural event segmentation of continuous experiences. *Cerebral Cortex*, *33*(13), 8164–8178. https://doi.org/10.1093/cercor/bhad106

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. Houghton Mifflin.

Schacter, D. L. (2021). The seven sins of memory: An update. *Memory*, *30*(1), 1–6. https://doi.org/10.1080/09658211.2021.1873391

Schacter, D. L., & Addis, D. R. (2007a). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 773–786. https://doi.org/10.1098/rstb.2007.2087

Schacter, D. L., & Addis, D. R. (2007b). The ghosts of past and future. *Nature*, *445*(7123), 27–27. https://doi.org/10.1038/445027a

Schacter, D. L., & Addis, D. R. (2009). On the nature of medial temporal lobe contributions to the constructive simulation of future events. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1245–1253. https://doi.org/10.1098/rstb.2008.0308

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, *76*(4), 677–694. https://doi.org/10.1016/j.neuron.2012.11.001

Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic future thinking: mechanisms and functions. *Current Opinion in Behavioral Sciences*, *17*, 41–50. https://doi.org/10.1016/j.cobeha.2017.06.002

Schacter, D. L., Greene, C. M., & Murphy, G. (2023). Bias and constructive processes in a self-memory system. *Memory*, *32*(6), 656–665. https://doi.org/10.1080/09658211.2023.2232568

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), Article 20160049. https://doi.org/10.1098/rstb.2016.0049

Schiffer, A.-M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS ONE*, *7*(5), Article e36445. https://doi.org/10.1371/journal.pone.0036445

Schiller, D., & Phelps, E. A. (2011). Does reconsolidation occur in humans? *Frontiers in Behavioral Neuroscience*, *5*, Article 24. https://doi.org/10.3389/fnbeh.2011.00024

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. https://doi.org/10.1016/j.cobeha.2014.07.005

Schlichting, M. L., Zeithamova, D., & Preston, A. R. (2014). CA1 subfield contributions to memory integration and inference. *Hippocampus*, *24*(10), 1248–1260. https://doi.org/10.1002/hipo.22310

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

Schlüter, H., Hackländer, R. P., & Bermeitinger, C. (2019). Emotional oddball: A review on memory effects. *Psychonomic Bulletin & Review*, *26*(5), 1472–1502. https://doi.org/10.3758/s13423-019-01658-x

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

Scully, I. D., & Hupbach, A. (2020). Different reactivation procedures enable or prevent episodic memory updating. *Hippocampus*, *30*(8), 806–814. https://doi.org/10.1002/hipo.23159

Selvan, R. N., Cheng, M., Siestrup, S., Mecklenbrauck, F., Jainta, B., Pomp, J., Zahedi, A., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I. (2024). Updating predictions in a complex repertoire of actions and its neural representation. *NeuroImage*, *296*, Article 120687. https://doi.org/10.1016/j.neuroimage.2024.120687

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

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

Shing, Y. L., Brod, G., & Greve, A. (2023). Prediction error and memory across the lifespan. *Neuroscience & Biobehavioral Reviews*, *155*, Article 105462. https://doi.org/10.1016/j.neubiorev.2023.105462

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., Leeners, L. M., Pomp, J., Roehe, M. A., Zahedi, A., & Schubotz, R. I. (2025). Co-occurring positive and negative prediction error signals are independent of stimulus

repetition effects. *Cerebral Cortex*, *35*(4), Article bhaf078. https://doi.org/10.1093/cercor/bhaf078

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

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

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

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

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

Smith, M. E., Loschky, L. C., & Bailey, H. R. (2024). Eye movements and event segmentation: Eye movements reveal age-related differences in event model updating. *Psychology and Aging*, *39*(2), 180–187. https://doi.org/10.1037/pag0000773

Smith, T. A., Hasinski, A. E., & Sederberg, P. B. (2013). The context repetition effect: Predicted events are remembered better, even when they don't happen. *Journal of Experimental Psychology: General*, *142*(4), 1298–1308. https://doi.org/10.1037/a0034067

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

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

Stemerding, L. E., Ast, V. A. van, Gerlicher, A. M. V., & Kindt, M. (2022). Pupil dilation and skin conductance as measures of prediction error in aversive learning. *Behaviour Research and Therapy*, *157*, Article 104164. https://doi.org/10.1016/j.brat.2022.104164

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. *Human Brain Mapping*, *46*(2), Article e70150. https://doi.org/10.1002/hbm.70150

Sutherland, R. J., Lee, J. Q., McDonald, R. J., & Lehmann, H. (2020). Has multiple trace theory been refuted? *Hippocampus*, *30*(8), 842–850. https://doi.org/10.1002/hipo.23162

Szpunar, K. K. (2010). Episodic future thought. *Perspectives on Psychological Science*, *5*(2), 142–162. https://doi.org/10.1177/1745691610362350

Szpunar, P. M., & Szpunar, K. K. (2016). Collective future thought: Concept, function, and implications for collective memory studies. *Memory Studies*, *9*(4), 376–389. https://doi.org/10.1177/1750698015615660

Takasago, M., Kunii, N., Fujitani, S., Ishishita, Y., Tada, M., Kirihara, K., Komatsu, M., Uka, T., Shimada, S., Nagata, K., Kasai, K., & Saito, N. (2024). Auditory prediction errors in sound frequency and duration generated different cortical activation patterns in the human brain: an ECoG study. *Cerebral Cortex*, *34*(3), Article bhae072. https://doi.org/10.1093/cercor/bhae072

Theobald, M., & Brod, G. (2021). Tackling scientific misconceptions: The element of surprise. *Child Development*, *92*(5), 2128–2141. https://doi.org/10.1111/cdev.13582

Theobald, M., Galeano-Keiner, E., & Brod, G. (2022). Predicting vs. guessing: the role of confidence for pupillometric markers of curiosity and surprise. *Cognition and Emotion*, *36*(4), 731–740. https://doi.org/10.1080/02699931.2022.2029733

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1–12.

Tulving, E. (2002). Episodic Memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Turan, G., Ehrlich, I., Shing, Y. L., & Nolden, S. (2023). From generating to violating predictions: The effects of prediction error on episodic memory. *OSF*. https://doi.org/10.31234/osf.io/zm29a

Udeogu, O. J., Frankenstein, A. N., Sklenar, A. M., Levy, P. U., & Leshikar, E. D. (2022). Predicting and remembering the behaviors of social targets: how prediction accuracy affects episodic memory. *BMC Psychology*, *10*(1), 96. https://doi.org/10.1186/s40359-022-00801-z

Ufer, C., & Blank, H. (2024). Multivariate analysis of brain activity patterns as a tool to understand predictive processes in speech perception. *Language, Cognition and Neuroscience*, *39*(9), 1117–1133. https://doi.org/10.1080/23273798.2023.2166679

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

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

Varga, N. L., Cronin-Golomb, L., & Bauer, P. J. (2022). Self-derivation of new knowledge through memory integration varies as a function of prior knowledge. *Memory*, *30*(8), 971–987. https://doi.org/10.1080/09658211.2022.2068609

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. *Memory & Cognition*, *47*(3), 496–510. https://doi.org/10.3758/s13421-018-0887-4

Varga, N. L., Roome, H. E., Molitor, R. J., Martinez, L., Hipskind, E. M., Mack, M. L., Preston, A. R., & Schlichting, M. L. (2025). Differentiation of related events in hippocampus supports memory reinstatement in development. *Journal of Cognitive Neuroscience*, *37*(4), 853–894. https://doi.org/10.1162/jocn_a_02299

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

Vlasceanu, M., Morais, M. J., & Coman, A. (2020). The effect of prediction error on belief update across the political spectrum. *Psychological Science*, *32*(6), 916–933. https://doi.org/10.1177/0956797621995208

Voelkel, J. G., Stagnaro, M. N., Chu, J. Y., Pink, S. L., Mernyk, J. S., Redekopp, C., Ghezae, I., Cashman, M., Adjodah, D., Allen, L. G., Allis, L. V., Baleria, G., Ballantyne, N., Bavel, J. J. V., Blunden, H., Braley, A., Bryan, C. J., Celniker, J. B., Cikara, M., … Willer, R. (2024). Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity. *Science*, *386*(6719), Article eadh4764. https://doi.org/10.1126/science.adh4764

von Goethe, J. W. (1774). *The sorrows of young Werther*. Project Gutenberg.

von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [An analysis of the processes in the trace field]. *Psychologische Forschung*, *18*(1), 299–342. https://doi.org/10.1007/bf02409636

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

Wahlheim, C. N., 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. *Journal of Experimental Psychology: General*, *148*(1), 30–50. https://doi.org/10.1037/xge0000458

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

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2002). Life is pleasant—and memory helps to keep it that way! *Review of General Psychology*, *7*(2), 203–210. https://doi.org/10.1037/1089-2680.7.2.203

Wang, S.-H., & Morris, R. G. M. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annual Review of Psychology*, *61*(1), 49–79. https://doi.org/10.1146/annurev.psych.093008.100523

Werning, M. (2020). Predicting the past from minimal traces: Episodic memory and its distinction from imagination and preservation. *Review of Philosophy and Psychology*, *11*(2), 301–333. https://doi.org/10.1007/s13164-020-00471-z

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

Xue, G. (2022). From remembering to reconstruction: The transformative neural representation of episodic memory. *Progress in Neurobiology*, *219*, Article 102351. https://doi.org/10.1016/j.pneurobio.2022.102351

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

Yassa, M. A., & Reagh, Z. M. (2013). Competitive Trace Theory: A role for the hippocampus in contextual interference during retrieval. *Frontiers in Behavioral Neuroscience*, *7*, Article 107. https://doi.org/10.3389/fnbeh.2013.00107

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

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

Zeithamova, D., & Bowman, C. R. (2020). Generalization and the hippocampus: More than one story? *Neurobiology of Learning and Memory*, *175*, Article 107317. https://doi.org/10.1016/j.nlm.2020.107317

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

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

Curriculum Vitae

Education

Since 2022 **Doctoral student at Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience | UNIVERSITY OF MÜNSTER**
- Structured interdisciplinary postgraduate program
- Participation in workshops, masterclasses, lectures, and retreats

2019 – 2021 **Master of Science Psychology | LUND UNIVERSITY**
- Research-oriented program with an emphasis on cognitive neuroscience
- Master's Thesis: "The role of social identity in memory integration"

2014 – 2018 **Bachelor of Science Psychology | UNIVERSITY OF POTSDAM**
- Final grade: 1.1
- Focus on cognitive psychology
- 2016/17: Erasmus semester at the University of Wrocław (Poland)

2020 – 2026 **Selected Additional Coursework**
- Advanced Research Practices (Lund University, 2025-2026, 7.5 ECTS)
- Workshop on statistical parametric mapping for fMRI in SPM (University College of London, 2023, three days)
- Popular Science – History, Theory, Practice (Umeå University, Spring 2022, 7.5 ECTS)
- Language and Cognition (Umeå University, Autumn 2021, 7.5 ECTS)
- Elements of AI (2021) & Building AI (2022, Linköping University, 2 ECTS each)
- Prague Summer school on Crime, Law, and Psychology (2020, one week)

Research Positions

Since 2022 **PhD Student and Research Associate | UNIVERSITY OF MÜNSTER**
- Project on the effects of prediction errors on episodic memory in a *DFG*-funded project led by Ricarda Schubotz
- Taking over supervision and teaching

2021 – 2022 **Research Assistant | LUND UNIVERSITY**
- Developed and conducted four follow-up experiments based on my Master's thesis
- Programmed and implemented online data collection for studies on the effect of context on making novel inferences

2021 – 2022 **Academic Assistant | UNIVERSITY OF HAGEN**
- Supported all stages of research on perceived discrimination and confrontation behavior

- Assisted in teaching activities and preparation of materials

2021 – 2022 **Research Intern | UNIVERSITY OF GRANADA**
- Intern at the Memory & Language Research Group
- Supported data collection for a study on retrieval-induced forgetting and creativity, using EEG and transcranial stimulation

2017 – 2018 **Student Assistant | UNIVERSITY OF POTSDAM**
- Assisted in data collection with eye tracking in the Cognitive and Biological Psychology division

Research Grants

- Collaborative Grant from Society for Applied Research in Memory and Cognition (SARMAC) with Karine Malysheva for a study on future thinking in Ukraine (2,000 USD)
- Co-applicant for Internal Lund University Get Started Fund to kickstart new international collaborations, used to invite guests from Taras Shevchenko National University of Kyiv (15,000 SEK)

Prizes

2024 First Poster Prize at Tagung Experimentell Arbeitender Psycholog*innen (TeaP; 300€)

Personal Scholarships

2025
- Santander Mobility Fund to visit the International Conference on Cognitive Neuroscience (ICON) in Porto, Portugal (500€)
- Erasmus+ scholarship for training at Karolinska Institutet in Stockholm, Sweden (1500€)
- SAIL Mobility Fund (University of Münster) to visit the Society of Applied Research in Memory and Cognition (SARMAC) conference in Kildare, Ireland (1000€)
- DAAD stipend for invited presentation at Aarhus University, Denmark (311€)

2024
- Erasmus+ scholarship for training and teaching at Jagiellonian University of Kraków, Poland (1500€)

2023
- Erasmus+ scholarship for teaching at Aarhus University, Denmark (1500€)

Publications

Boeltzig, M., Siestrup, S., Bramão, I., & Schubotz, R. I. (2026). To update or to separate: Neural signatures and consequences of latent cause inference in episodic memory. *NeuroImage (under review)*.

Boeltzig, M., Schubotz, R., Cole, S., & Rathbone, C. (2026). Self-serving biases shape the relationship between future thinking and remembering of elections. *Communications Psychology*. https://doi.org/10.1038/s44271-026-00423-w

Liedtke, N., **Boeltzig, M.**, Siestrup, S., & Schubotz, R. I. (2026). Shaping memory from

the start: Initial prediction errors during first encoding. *NeuroImage, 325*, Article 121660. https://doi.org/10.1016/j.neuroimage.2025.121660

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

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

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

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

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

Conference Oral Presentations

Boeltzig, M. (2026). „We will be free“: The wartime positivity effect in collective future thinking in Ukraine. *Tagung Experimentell Arbeitender Psycholog*innen (TeaP). Tübingen, Germany*

Boeltzig, M. (2025). Partisan Predictions, Malleable Memory: Phenomenology and self-congruency bias in simulating and remembering elections. *Society of Applied Research in Memory and Cognition (SARMAC). Kildare, Ireland.*

Boeltzig, M. (2025). “I always knew it”: Self-serving biases moderate the relationship between future thinking and episodic remembering in the context of elections. *Generative Episodic Memory (GEM). Bochum, Germany.*

Boeltzig, M. (2025). The many meanings of memory updating - How prediction errors influence memory representations. *Tagung Experimentell Arbeitender Psycholog*innen (TeaP). Frankfurt, Germany.*

Conference Poster Presentations

Boeltzig, M., Liedtke, N., Schubotz, R.I. (2025). Neither Wrong Nor Right: Medium Prediction Errors Modify Episodic Memories. *Poster at International Conference on Cognitive Neuroscience (ICON). Porto, Portugal.*

Boeltzig, M., Liedtke, N., Schubotz, R.I. (2024). The effect of prediction errors on memory representations of naturalistic episodes. *Experimental Psychology Society (EPS). York, United Kingdom [Virtual Poster].*

Boeltzig, M., Liedtke, N., Schubotz, R.I. (2024). Representing old and new - The neural patterns of episodic memory updating. *Psychology and Brain (PUG). Hamburg, Germany.*

Boeltzig, M., Liedtke, N., Schubotz, R.I. (2024). That’s not what you said! Memory updating after prediction errors in naturalistic conversations. *Tagung experimentell arbeitender Psycholog:innen. Regensburg, Germany.*

Liedtke, N.*, **Boeltzig, M.***, Schubotz, R. I. (2023). Learning from quantified episodic prediction errors: Individual biases in gist revision. *Generative Episodic Memory: Interdisciplinary perspectives from neuroscience, psychology and philosophy. Bochum, Germany.*

* equal contributions

Invited Talks

2025 **It has begun: Flashbulb Memories of the Russian Invasion in Ukraine**
- McDonnel Foundation Collective Memory Group
- CON AMORE, Aarhus University

2024 **Prediction Errors in Naturalistic Conversations**
- Neuroscience & Cognition Group, Lund University
- Applied Memory Research Laboratory, Jagiellonian University of Kraków

2023 **Prediction Errors in Naturalistic Conversations**
- MRC Cognition and Brain Sciences Unit, University of Cambridge
- Montaldi Lab, University of Manchester
- York Episodic Memory Lab, University of York
- CON AMORE, Aarhus University
- Emotion and Neuroscience Group, University of Potsdam

2021 **The role of social identity in associative inference**
- Memory and Language Lab, University of Granada

Lab Visits

August 2025	Visiting Andreas Olsson at Karolinska Institutet Stockholm (Sweden) to discuss methods such as skin conductance and the use of virtual reality (2 weeks, funded by an Erasmus+ scholarship)
July 2024	Visiting Krystian Barzykowski at Jagiellonian University of Kraków (Poland) to initiate collaborations (1.5 weeks, funded by an Erasmus+ scholarship)
September 2023	Visiting Aidan Horner at University of York (United Kingdom) to initiate a collaboration for a future project and become familiar with Representational Similarity Analysis (3 weeks)
August 2023	Visiting Annette Bohn and Dorthe Berntsen at CON AMORE (University of Aarhus, Denmark) to teach a workshop on R and develop a project on Flashbulb Memories (2 weeks, funded by an Erasmus+ scholarship)

Organization & Outreach

January 2026	Initiated and co-organized the Kyiv Winter School of Memory, a one-week online event with high-profile lecturers delivering presentations on episodic, autobiographical, collective, and traumatic memory to 100 participants
Autumn 2024	Initiated and delivered one semester of weekly R seminars for Ukrainian students and faculty
Since 2021	Popular science presentations on memory at the German community of Malmö, the Perm State Library, and the Quantum Photonics live podcast

Teaching

- **Thesis Supervision** (University of Münster, ongoing): five completed Bachelor's theses, one completed Master's thesis, three ongoing Bachelor's and one ongoing Master's theses
- **Pain, stress, sexuality, emotion, motivation** (University of Münster, 2024-2026): Four seminars, 2 hours per week each
- **Practical Research Introduction** (University of Münster, winter semester 2023/24): Seminar, 3 hours per week
- **Introduction into R** (Taras Shevchenko University of Kyiv, fall 2023): Extra-curricular seminar for interested students and faculty, 10 sessions
- **Introduction into R** (Aarhus University, summer 2023): Workshop for faculty, 2 days
- **Tutorial in Statistics** (University of Potsdam, summer semesters 2016 & 2017): Weekly practical tutorials supplementing an introductory statistics lecture, 2 groups per week
- **Tutorial in Scientific Methods** (University of Potsdam, winter semester 2016/17): Practical tutorials on scientific skills for first-semester students, 2 hours per week

Skills

Languages: German (native), English (fluent), Spanish, Swedish, Russian (intermediate), Ukrainian, Polish (beginner)
Programming: R, Presentation (Neurobehavioral Systems), PsychoPy & Pavlovia, Python, MatLab

Declarations

Declarations by the candidate documenting **open-science activities** and **on the consideration of ethical aspects** as part of the doctoral process and disclosure of **personal contribution** for manuscripts completed by two or more authors (cumulative dissertations)

Doctoral candidate: Marius Boeltzig

Title of dissertation: Learning From The Unexpected: How Prediction Error Size Shapes Episodic Memory

1. Documentation of **open-science activities**

Manuscript 1

	yes	no	If yes, please specify the source
Pre-registration		x	
Publication of data	x		https://osf.io/8dnsg/overview?view_only=3fc9fd3393474fd8b82969623c6ce6f9
Publication of analysis scripts		x	
Publication of materials		x	
Open access publication	x		
Preprint		x	
Analysis scripts were not shared as no custom code was used. The materials are made available upon request but could not be deposited in an open database due to their involvement of professional speakers and ensuing licensing issues.			

Manuscript 2

	yes	no	If yes, please specify the source
Pre-registration		x	
Publication of data	x		https://osf.io/wnzs3/overview?view_only=8a7e9cf626aa42eeb9cb83c5e386ef41
Publication of analysis scripts		x	
Publication of materials		x	
Open access publication	x		
Preprint		x	
See Manuscript 1			

Manuscript 3

	yes	no	If yes, please specify the source
Pre-registration		x	
Publication of data	x		https://osf.io/mgu49/overview?view_only=1af278283f49433d8de09f566a2dff97
Publication of analysis scripts		x	
Publication of materials	(x)		
Open access publication	x		
Preprint		x	
The transcribed dialogues are published on OSF for this study, but not the full set of audio files (see Manuscript 1).			

2. Declaration on the consideration of ethical aspects

Study number	**Source (manuscript / chapter of dissertation):** e.g. study 1 in paper 2, study 1 described in chapter 4	**Was the study reviewed by an ethics commission?**	
		yes	**no**
1	Study 1	x	
2	Study 2	x	
3	Study 3	x	

3. Declaration of one's personal contribution to the submitted academic manuscripts by two or more authors

Manuscript 1

Title	Prediction errors lead to updating of memories for conversations	
Authors	Marius Boeltzig, Nina Liedtke, Ricarda I. Schubotz	
Publication status:	not yet submitted	
	submitted	
	in review	
	in revision	
	accepted	
	published	x
Journal	Memory	
Year of publication	2025	

Description of your own contribution in the case of joint authorship:

- Contributed to study design and research questions
- Contributed to data collection
- Solely responsible for data preparation, analysis, visualization
- Wrote first draft of manuscript and led revision process

Manuscript 2

Title	The benefit of being wrong: How prediction error size guides the reshaping of episodic memories	
Authors	Marius Boeltzig, Nina Liedtke, Sophie Siestrup, Falko Mecklenbrauck, Moritz F. Wurm, Inês Bramão, Ricarda I. Schubotz	
Publication status:	not yet submitted	
	submitted	
	in review	
	in revision	
	accepted	
	published	x
Journal	NeuroImage	
Year of publication	2025	

Description of your own contribution in the case of joint authorship:

- Contributed to study design and research questions
- Contributed to data collection
- Solely responsible for data preparation, analysis, visualization
- Wrote first draft of manuscript and led revision process

Manuscript 3

Title	To update or to separate: Neural signatures and consequences of latent cause inference in episodic memory	
Author(s)	Marius Boeltzig, Sophie Siestrup, Inês Bramão, Ricarda I. Schubotz	
Publication status:	not yet submitted	
	submitted	
	in review	x
	in revision	
	accepted	
	published	
Journal	NeuroImage	
Year of publication		

Description of your own contribution in the case of joint authorship:

- Contributed to study design and research questions
- Solely responsible for data collection
- Solely responsible for data preparation, analysis, visualization
- Wrote first draft of manuscript and led revision process

Malmö, 23.02.2026

Place, Date Signature of the doctoral candidate