Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Learning from the Unexpected: How Prediction Errors Shape Episodic Memory Formation

---

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

an der Universität Hamburg,

Fakultät für Psychologie und Bewegungswissenschaften,

Institut für Psychologie

vorgelegt von

Kaja Loock

Hamburg, 2025

Tag der mündlichen Prüfung: 11.07.2025

Promotionsprüfungsausschuss

Vorsitzender:	PD Dr. Patrick Bruns
1. Dissertationsgutachter:	Prof. Dr. Lars Schwabe
2. Dissertationsgutachter:	Prof. Dr. Sebastian Gluth
1. Disputationsgutachterin:	Prof. Dr. Anja Riesel
2. Disputationsgutachter:	Prof. Dr. Jan Wacker

Acknowledgements

Reflecting upon the last four years of this journey, I realize how much knowledge I have gained and how much I have matured personally. This would not have been possible without the extraordinary support of so many wonderful people. First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Lars Schwabe, for his invaluable guidance, for his unwavering support, and for his exceptional responsiveness to all my questions throughout every stage of my PhD. His mentorship has been fundamental not only my projects, but also to nurturing my scientific curiosity and growth as a researcher. It has been a true privilege to share this journey with amazing colleagues. I would like to thank Anna-Maria, Stefan, Felix, Valentina, Leah and Hendrik for introducing me into the world of academia and sharing your knowledge with me. Kai, Sterre, Elena, Sarah, Ivana, Marta, Li, Claudius, Jacqueline, Xiaoyu, Blazej, Jessica, and Coco, thank you for your support, thoughtful conversations, and for making even the most challenging days feel lighter and the good ones even better. Antonia, thank you for all the fun times we shared both inside and outside the lab. Your friendship and humor made this journey all the more enjoyable. Thank you, Fabian, for being a fantastic office mate during the countless hours in our office, for your steady encouragement and for creating such a supportive and collaborative environment. I would also like to acknowledge the many students who contributed to this work. Thank you, Tim, Mali, Konstantin, Anna B., Fatih, Chris, Hazem, Constantin, Philipp, Lilli, Klara, Johannes, Luigi, Farida, Mashid, Berivan, Sabrina, Jasmin, Flavia, Astrid, Georgia, Anna T., Fenja, Laura, Mattias, Stefanie, Jeannine, and Maik, for your assistance that did not go unnoticed and was greatly appreciated. I am deeply grateful to Maren, Hendrik, and Vanessa for being such thoughtful and encouraging companions - your honest feedback and our many chats were invaluable during the writing of this thesis. To my dear friends, thank you for your humor in all the right moments, and for always helping me see things more clearly. Your friendship, whether near or far, has kept me grounded, lifted my spirits, and reminded me of life beyond deadlines. I am endlessly grateful to my parents and my brother for their constant love, patience, and belief in me throughout this journey. Your support has been my foundation. A special thank you to my grandpa - your strength, curiosity, and pride in my work mean the world to me. Last, but certainly not least, I want to thank you, Vanessa, for standing beside me with so much love, patience, and faith. Through shared tears and joy, you have been my anchor, and I truly could not have done this without you.

Abstract

Updating memory in response to unexpected outcomes is critical for maintaining an adaptive internal model of the world. While prediction errors (PEs) have long been central to theories of reinforcement learning, emerging evidence indicates they also play a key role in shaping episodic memory, although their underlying mechanisms remain poorly understood. This thesis systematically investigates how PEs, particularly in aversive contexts, influence episodic memory formation by uncovering their cognitive and neural underpinnings. Across five studies, we employed modified versions of an incidental encoding-fear learning task to examine how expectancy violations modulate memory. Studies I and II replicated previous findings that unsigned PEs retrospectively enhance memory for predictive stimuli, and further demonstrated that these effects also occur prospectively and independently of physiological arousal. Studies III to V extended these findings by revealing that signed PEs affect memory formation in a direction-specific manner, with positive PEs enhancing and negative PEs attenuating memory. Study III investigated the temporal constraints of PE-induced memory enhancements by varying the delay between predictive cues and outcomes. Results showed that temporal proximity is not essential for PE-driven memory enhancements, suggesting that contingency, rather than contiguity, plays a critical role. Study IV examined the specificity of PE effects by testing whether memory enhancements also occur for uninformative, i.e., unpredictive, stimuli encountered near the PE event. The memory benefits were restricted to predictive cues, supporting a selective encoding mechanism susceptible to interference. Finally, Study V combined EEG with inhibitory continuous theta burst stimulation (cTBS) over the right superior parietal cortex (rSPC) to explore its causal role in PE-induced memory modulations and the neural dynamics surrounding PEs. Alpha and theta oscillations before the PE and stimulus reactivation after the PE predicted enhanced memory, depending on the direction of the PE. Interestingly, inhibiting the rSPC paradoxically boosted memory, suggesting that reduced top-down filtering may facilitate memory formation under surprise. Together, these findings suggest that PEs promote the selective encoding and consolidation of behaviorally relevant events into long-term memory by engaging attention, working memory, and salience networks. These insights provide translational potential for clinical interventions by identifying mechanisms through which maladaptive processes in fear-related disorders could be altered.

Contents

List of Figures

List of Abbreviations

ACC	Anterior cingulate cortex
ANOVA	Analyis of variance
CMS	Common mode sense
CR	Conditioned response
CS	Conditioned stimulus
cTBS	Continuous theta burst stimulation
DFT	Discrete Fourier-Transform filter
DMS	Delayed-matching-to-sample task
dlPFC	Dorsolateral prefrontal cortex
DRL	Driven right leg
EEG	Electroencephalography
EMG	Electromyography
ERP	Event-related potential
fMRI	Functional magnetic resonance imaging
FRN	Feedback Related Negativity component
GLMM	General linear mixed model
HC	Hippocampus
IBI	Interbeat interval
ICA	Independent component analysis
LC	Locus coeruleus
LMM	Linear mixed model
LTP	Long-Term Potentiation
MEP	Motor evoked potential
MH	Motor hotspot
MNI	Montreal Neurological Institute
mPFC	Medial prefrontal cortex
MPRAGE	Magnetization-prepared rapid acquisition gradient echo
MT	Motor threshold
MTL	Medial temporal lobe
MVPA	Multivariate pattern analysis
P3	Positivity 300 component

PE	Prediction error
PFC	Prefrontal cortex
PTSD	Post-traumatic stress disorder
rSPC	Right superior parietal cortex
(r)TMS	(Repetitive) transcranial magnetic stimulation
SCR	Skin conductance response
sPE	Signed prediction error
TAL	Talairach
TD	Temporal difference
UCS	Unconditioned stimulus
UI	Uninformative stimulus
uPE	Unsigned prediction error
VTA	Ventral tegmental area

1 General introduction

Human cognition is inherently predictive. Across perception, decision-making, and social interaction, the brain continuously generates expectations about future events drawing from prior experiences to guide future behavior (Clark, 2013; Friston, 2010). From a computational perspective, this process can be modeled as probabilistic inference under uncertainty: the brain constructs generative models of the world, continuously updating them to reduce prediction errors (PEs; Clark, 2013; Shohamy & Adcock, 2010). If an anticipated outcome does not occur or something unexpected happens, this expectancy violation produces a PE being fundamental in this framework. PEs signal mismatches between sensory input and internal models, thereby driving model updating (Den Ouden et al., 2009; Niv & Schoenbaum, 2008). PEs have long been recognized as drivers of reinforcement learning and decision making (Gläscher et al., 2010; Glimcher, 2011; Niv, 2009; Rescorla & Wagner, 1972; Schultz et al., 1997). In reinforcement learning, PEs are formally defined as the difference between expected and received reward and are used to update value estimates, optimizing future decisions (Niv, 2009; Rescorla & Wagner, 1972; Sutton, 1998). More recently, a growing body of evidence suggests that these learning signals, i.e., PEs, not only guide learning and decision-making but also serve a mnemonic function: Unexpected events, even if emotionally neutral, appear more likely to be encoded into long-term memory than predictable ones (Greve et al., 2017; Rouhani et al., 2018; Sinclair & Barense, 2019). This memory enhancing effect has been observed across various domains, such as reward-based learning (Ergo et al., 2020; Jang et al., 2019; Rouhani & Niv, 2021) and aversive learning (Den Ouden et al., 2009; Kalbe & Schwabe, 2020, 2022b), and may operate independently of emotional arousal, enhancing memory for neutral stimuli that merely occurred in the context of violated expectations (Ergo et al., 2020; Kalbe & Schwabe, 2020; Rouhani & Niv, 2021). This aligns with computational theories positing that memory systems should prioritize events with high learning value, i.e., experiences that offer maximal model updating (Courville et al., 2006; Gershman et al., 2014). PEs could serve precisely this function, marking moments of surprise that warrant increased representational precision or episodic encoding (Dunsmoor et al., 2022; Gershman, 2017). Dopaminergic and noradrenergic pathways likely play a key role in forwarding these salience signals to memory-relevant structures such as the hippocampus (HC; Düzel et al., 2009; McNamara & Dupret, 2017; Takeuchi et al., 2016).

Overall, these findings suggest that PEs act not only as learning signals but also as key drivers of episodic memory formation, prioritizing the encoding of events that carry informational value (Jang et al., 2019; Pupillo et al., 2023; Rouhani & Niv, 2021). This would

constitute a fundamental mechanism by which the brain ensures that surprising or behaviorally relevant experiences are retained for future use. Yet despite the growing interest in this phenomenon, the cognitive and neural mechanisms by which PEs enhance episodic memory remain largely unclear.

1.1 Mechanisms of adaptive memory

Human memory is not a passive archive of experiences but a highly selective, goal-oriented system (Shohamy & Adcock, 2010; Tulving, 1972). This is particularly evident in episodic memory, capturing personally experienced events in rich contextual detail (Tulving, 1972). Rather than storing all encountered information indiscriminately, the brain dynamically filters and encodes events based on their relevance to future behavior, a process referred to as adaptive memory (Gershman & Daw, 2017; Shohamy & Adcock, 2010). From an evolutionary perspective, such prioritization makes functional sense: Remembering every mundane detail of daily life would be inefficient, whereas selectively encoding behaviorally salient events, e.g., emotionally charged events, novel occurrences, or outcomes that defy expectation, enhances the organism's capacity to plan, predict, and adapt in future encounters (Christianson, 2014; Nairne & Pandeirada, 2008; Reisberg & Hertel, 2005; Shohamy & Adcock, 2010).

Adaptive memory formation is thought to be supported by the interaction of multiple cognitive and neural mechanisms. The HC has been established as a central structure in the formation of episodic memories which are rapidly acquired and rich in contextual detail (Burgess et al., 2002; Chadwick et al., 2010; Moscovitch et al., 2016). Indeed, research showed that the HC is functionally involved in memory retrieval, spatial learning and cognitive maps, prioritizing the encoding of valuable and contextual details of events (Chadwick et al., 2010; FeldmanHall et al., 2021; Schapiro et al., 2013). Moreover, hippocampal neurons are assumed to act as pointers and indices to distributed representations across the neocortex, linking the content, context and details of an encoded event (Moscovitch, 1995; Teyler & Rudy, 2007). Together with adjacent medial temporal lobe (MTL) cortices and the medial prefrontal cortex (mPFC), the HC supports the binding of multiple elements of an event into a narrative of complex and coherent memory representations (Allen & Fortin, 2013; Kroes & Fernández, 2012; Shohamy & Adcock, 2010). Specifically, the HC serves to detect regularities in the environment that are integrated with response options in the mPFC to form abstract knowledge (Garvert et al., 2017; Nieh et al., 2021). This abstract knowledge can guide future behavior in novel situations that only partially resemble existing episodic experiences (Kroes & Fernández, 2012). Recent research suggests that the MTL is involved in the imagination of future episodes

with important implications for decision-making (Schacter & Addis, 2007). Moreover, the relational structure of episodic memories also allows flexible retrieval of relevant information to guide goal-directed behavior in the future, highlighting the necessity of HC-, mPFC- and MTL-involvement in linking episodic memory and adaptive decision-making (FeldmanHall et al., 2021; Shohamy & Adcock, 2010).

1.1.1 Prioritization of salient events

Additionally, research of the past decades has demonstrated a critical impact of emotions on memory formation. Compared to neutral events, emotionally arousing events have been shown to be encoded faster and remembered more vividly and accurately (Cahill & McGaugh, 1996; Christianson, 2014; Reisberg & Hertel, 2005). The characteristics of emotional information, i.e., its elicited physiological arousal, may facilitate preferential processing of heightened attention to emotional content at encoding (Mather & Sutherland, 2011; Talmi, 2013). This emotional memory enhancement is largely attributed to a noradrenergic arousal-related activation of the amygdala that supports preferential encoding and enhances the distinctiveness and consolidation of relevant experiences (Clewett et al., 2014; Mather & Sutherland, 2011). Subsequently, the amygdala then modulates memory consolidation processes in other brain regions, i.e., the HC and neocortex (Buchanan, 2007; Cahill & McGaugh, 1995, 1998; Fastenrath et al., 2014; LaBar & Cabeza, 2006; McGaugh, 2018; McReynolds & McIntyre, 2012; Phelps, 2004).

Stress hormones such as glucocorticoids (e.g., cortisol) and catecholamines (e.g., noradrenaline), in response to an emotional event, signal the salience of an event, thereby influencing the strength of memory trace formation (Cahill & McGaugh, 1998; Krugers et al., 2012; Zerbes et al., 2019). Specifically, an emotional event triggers physiological arousal which in turn leads to a secretion of stress hormones in the adrenal glands, i.e., norepinephrine and glucocorticoids (Roozendaal et al., 2006). The interaction of norepinephrine and cortisol in the basolateral amygdala then modulates memory plasticity and consolidation processes in other brain areas, i.e., the HC, mPFC and MTL (McGaugh, 2000; Roozendaal et al., 2006). Similarly, arousal triggers locus coeruleus (LC) activity which leads to the release of noradrenaline that selectively strengthens prioritized memory representations by modulating local and functional network-level patterns of information processing (Clewett et al., 2018; Mather et al., 2016). The increase in noradrenaline enhances synaptic plasticity exciting local protein synthesis processes that enhance selective memory consolidation (Mather et al., 2016). Pupil dilation, a peripheral marker of arousal and LC-noradrenaline system activity, has been shown to reliably predict long-term memory formation, especially for emotionally arousing stimuli (Bergt et al.,

2018; Huang & Clewett, 2024). These findings suggest that arousal-related neuromodulatory activity may gate synaptic plasticity, supporting the selective consolidation of salient information.

1.1.2 Behavioral tagging account

While neuromodulatory arousal-based mechanisms such as noradrenergic activation support the selective consolidation of emotional memories, they also interact with synaptic plasticity processes that underlie long-term memory formation. Repeated stimulation of neurons in the HC can induce long-term potentiation (LTP) establishing long-term memory (Frey & Morris, 1997). As proposed by the *synaptic tagging and capture hypothesis* (Frey & Morris, 1997; Frey & Frey, 2008), an early-LTP triggers the formation of a transient, protein-synthesis-independent synaptic tag which decays in less than three hours at the stimulated neuron. This synaptic tag then captures the necessary plasticity-related proteins to support the development of late and persistent LTP (Frey & Morris, 1997). Intriguingly, this hypothesis implicates that a weak stimulus being insufficient to induce protein synthesis can lead to a persistent late-LTP, if the weak and a strong stimulus, with the first not being able and the latter being able to induce a synaptic tag, were applied in a relatively long-lasting associative time window on different synapses of the same neuron (Frey & Morris, 1997; Ballarini et al., 2009). A behavioral analogy, the *behavioral tagging hypothesis* proposes that the formation of LTM depends on two processes: The setting of a learning tag, and the synthesis of plasticity-related proteins which can enhance memory consolidation (Moncada et al., 2015). In particular, weak events are most likely to induce short-lived memories, which can lead to a persistent long-term memory if they occur in close temporal proximity to a salient, behaviorally significant event that induced the secretion of plasticity-related proteins (Ballarini et al., 2009; Dunsmoor et al., 2022; Moncada & Viola, 2007). Thus, behavioral tagging provides a compelling mechanism through which the brain selectively strengthens memories that occur in close temporal proximity to behaviorally significant and arousing events.

Importantly, adaptive memory does not solely depend on the emotional valence of an experience. It is also critically shaped by its informational value that describes the degree to which an event updates prior knowledge or signals a change in environmental contingencies. From this perspective, the impact of emotional events on memory may, in part, stem from their inherent unpredictability and the surprise they elicit (Trapp et al., 2018). In this sense, PEs, i.e., violations of expectation, may act as behaviorally significant signals that trigger mechanisms akin to those proposed by the behavioral tagging framework. Recent research suggests that PEs

may serve as powerful triggers for adaptive memory processes (Antony et al., 2021; Rouhani et al., 2018; Rouhani et al., 2023).

1.2 Prediction errors as modulators of learning and memory

PEs have long been recognized from the reinforcement learning domain as driving forces of memory updating (Gläscher et al., 2010; Glimcher, 2011; Niv, 2009; Rescorla & Wagner, 1972; Schultz et al., 1997). Acting as teaching signals to the brain, PEs trigger incremental learning processes that enable the brain to adaptively optimize future behavior (Bar, 2007; Bein et al., 2020; Clark, 2013; Ergo et al., 2020; Trapp et al., 2018). PEs in the framework of reinforcement learning, i.e., reward PEs signalling the value of an unexpected outcome, are conveyed by midbrain dopaminergic neurons and drive incremental learning of future values of stimuli by supporting decision making based on basal-ganglia structures (Barto, 1995; Montague et al., 1996; Schultz et al., 1997; Niv & Schoenbaum, 2008; Niv, 2009). Specifically, the relevant information from a new experience is used to update either situation-action values (in model-free approaches) or the parameters of an internal model, which can then dynamically compute future values (in model-based approaches; Gershman & Daw, 2017; Gläscher et al., 2010). Specifically, PEs enable cumulative learning in Pavlovian conditioning as suggested by the Rescorla Wagner model (Miller et al., 1995; Rescorla & Wagner, 1972; Yau & McNally, 2023). In Pavlovian conditioning, an inherently neutral stimulus (e.g., ring of a bell) is paired with an unconditioned stimulus (UCS; e.g., an electric shock). Throughout the conditioning process, i.e., multiple associations, the neutral stimulus turns into a conditioned stimulus (CS) that elicits a conditioned response (CR; e.g., fear) by itself. Fitting into the Rescorla Wagner model, this learning process is based on adjusting values by discrepancies between prediction and outcome (Den Ouden et al., 2012; Miller et al., 1995; Niv, 2009; Niv & Schoenbaum, 2008):

$$V_{\{new\}} = V_{\{old\}} + \eta \cdot \left(R - V_{\{old\}}\right)$$

$V_{\{new\}}$ represents the updated value of an event incorporating the new information provided by the actual outcome. Here, $V_{\{old\}}$ describes the current, i.e., old, value of the prediction before the update. η is the learning rate ($0 \leq \eta \geq 1$) determining the magnitude of the update induced by the new experience. High learning rates emphasize recent experiences, allowing the model to rapidly update its predictions based on new information. However, this also results in faster discounting of past experiences. In contrast, low learning rates lead to more gradual learning,

requiring the accumulation of multiple past and new experiences before significantly altering predictions. *R* is a scalar quantity of the goodness of the observed outcome. In sum, the term $R\text{-}V_{\{old\}}$ represents the PE indicating the discrepancy between the predicted and the actual outcome (Niv & Schoenbaum, 2008).

An extended model of the basic approach mentioned above incorporates temporal difference (TD) learning, i.e., a TD PE (Maes et al., 2020; Niv & Schoenbaum, 2008; O'Doherty et al., 2003). TD learning acknowledges that real-world experiences unfold as a continuous stream of information rather than as discrete, isolated trials. Within this ongoing stream of information, predictive cues and rewarding outcomes often occur at different timepoints. The aim at each time point is to make accurate predictions about future outcomes based on the current state and the history of preceding stimuli. This means that, if predictions are accurate, the value predicted at time *t* should equal the sum of two components which are (i) the expected immediate reward at time *t*+1 (which could be zero) and (ii) the predicted value of future rewards from time *t*+1 onward (Dayan, 1993; Maes et al., 2020; Niv & Schoenbaum, 2008). Thus, the TD error (i.e., δ at time $t + 1$) is defined as the difference between predicted and actual outcomes (Sutton, 1988):

$$\delta(t+1) = \text{outcome}(t+1) + \text{prediction}(t+1) - \text{prediction}(t)$$

Specifically, the TD error is computed as the sum of the actual outcome received at time *t*+1 and the updated prediction of future rewards at that time, minus the prediction made at the previous time step *t* (Dayan, 1993; Seymour et al., 2004). An error that is unequal to zero signals a mismatch between expected and experienced outcome. This error can be incorporated into learning the value $V(t)_{\{new\}}$ of a stimulus as follows:

$$V(t)_{\{new\}} = V(t)_{\{old\}} + \eta \cdot [\big(\text{outcome}(t+1) + \text{prediction}(t+1) - \text{prediction}(t)\big)]$$

Here, *V(t)* is the prediction at time point *t*, η is the learning rate, outcome($t + 1$) is the outcome at the succeeding time point and prediction($t + 1$) is the prediction at time point *t* (Niv & Schoenbaum, 2008). Thus, TD learning assumes that PEs modulate reinforcement learning by being a driving force of adjusting models to guiding future behavior (Maes et al., 2020; Schultz, 2016; Seymour et al., 2004).

1.2.1 Unsigned and signed PEs

PEs can be considered as (i) unsigned or (ii) signed PEs reflecting different computational and functional characteristics in learning and memory processes (Gurunandan et al., 2025; Pupillo & Bruckner, 2023; Rouhani et al., 2023). While unsigned PEs (uPEs) take the absolute magnitude of deviation between prediction and outcome into account (ranging between 0 and 1), signed PEs (sPEs) account for the direction of deviation containing information about the value of the outcome (positive vs. negative, ranging from -1 to 1).

In the reward domain, the uPE is assumed to play a critical role in learning and memory, irrespective of the PE's direction, i.e., good or bad unexpected outcomes (Rouhani & Niv, 2021; Rouhani et al., 2018). In contrast to signed PEs, which carry information about outcome valence (Schultz, 2017), uPEs primarily signal surprise (Hayden et al., 2011). This form of PE has been shown to enhance learning by increasing attention to outcomes that deviate from expectation, thereby improving model updating and memory encoding (Hayden et al., 2011; Rouhani & Niv, 2021; Rouhani et al., 2018). uPEs are underpinned by distinct neural signals that drive learning about outcomes in various domains, e.g., perception, motor function and reward (Den Ouden et al., 2012; Fiorillo, 2013; Fouragnan et al., 2018). Accumulating evidence suggests that uPEs enhance learning rate adaptation and episodic memory strength in proportion to the degree of unexpectedness. Indeed, greater uPEs, regardless of whether the outcome was better or worse than expected, has been shown to correlate with improved subsequent recall, supporting the idea that surprise signalled by a PE can boost memory encoding (Ergo et al., 2020; Greve et al., 2017; Jang et al., 2019; Metcalfe, 2017; Stanek et al., 2019). This effect seems to be due to the surprise elicited by the outcome (Steinberg et al., 2013). Specifically, unexpected events are assumed to trigger phasic activation of the LC resulting in a concommitant norepinephrine and dopamine release in the LC, ultimately leading to increased plasticity in the HC (Clewett & Murty, 2019; Jordan & Keller, 2023; Takeuchi et al., 2016). These neurotransmitters are thought to increase synaptic plasticity and boost network excitability (Mather & Sutherland, 2011; Lisman et al., 2011). Importantly, uPEs may serve an adaptive function by updating internal models about environmental volatility and uncertainty (Behrens et al., 2007). When large uPEs are detected, the brain may shift into a high learning mode, allocating greater cognitive and mnemonic resources to the unexpected event leading to rapid learning (Behrens et al., 2007).

While uPEs reflect the magnitude of surprise regardless of outcome valence, sPEs capture additional information about the direction of the mismatch, i.e., whether the outcome was better or worse than expected (Ergo et al., 2020). These valence-specific signals

are particularly important in the reward domain, where outcomes can vary between reward and punishment (Schultz, 2017). Specifically, reward sPEs influence dopaminergic firing from the midbrain, particularly from the ventral tegmental area (VTA), which projects to key memory- and motivation-related regions such as the HC, nucleus accumbens, striatum and prefrontal cortex (PFC; Bayer & Glimcher, 2005; Sharpe et al., 2017; Watabe-Uchida et al., 2017). When outcomes are better than expected, i.e., a positive PE, dopamine neurons increase firing, whereas outcomes worse than expected, i.e., negative PEs, lead to suppression of dopaminergic activity (Keiflin & Janak, 2015; Rouhani et al., 2023; Schultz, 2016). These dopaminergic signals are essential for adjusting internal models of reward contingencies and for guiding future decisions by updating expectations to improve the reflection of actual outcomes (Eshel et al., 2016; Rouhani et al., 2023).

Over time and learning, dopaminergic sPEs can shift from the outcome itself to the cue that predicts it, reflecting temporal difference learning (Doll et al., 2015; Montague et al., 1996). This dynamic transfer supports the framework of synaptic and behavioral tagging, where initially weak and neutral cues acquire motivational salience and become preferentially encoded into memory after being tagged by a salient event (Ballarini et al., 2009; Moncada & Viola, 2007). Furthermore, the PE-induced dopaminergic response may also be associated with increased neural plasticity and ultimately enhanced memory consolidation, e.g., in the HC (Cahill & McGaugh, 1998; Lisman et al., 2011; McGaugh, 2000; Rouhani et al., 2023, Trapp et al., 2018). However, the impact of reward PEs on memory may also be modulated by individual differences in dopaminergic tone, cognitive traits such as working memory capacity, or even motivational states (Krebs et al., 2009; Murty & Adcock, 2014).

Recently, accumulating evidence has further expanded upon the memory enhancing effects of reward sPEs (Ergo et al., 2020; Greve et al., 2017; Jang et al., 2019; Rouhani & Niv, 2021; Rouhani et al., 2018). Memory encoding seems to be improved for the predictive event in the case of positive reward sPE, i.e., better than expected, and to decline if outcomes are worse than expected, i.e., negative reward sPEs (De Loof et al., 2018; Jang et al., 2019). Neurally, this memory modulating effect of reward sPEs is linked to activity in the ventral striatum (Calderon et al., 2021; Pine et al., 2018; Ripollés et al., 2018) and increased alpha and beta oscillations (Ergo et al., 2019).

1.2.2 PEs related to aversive events and their role in memory formation

While most research has focused on the role of PEs in reward-based learning, similar computational mechanisms have been proposed for aversive, perceptual, and even social domains (Corlett et al., 2022; Den Ouden et al., 2012; Tzovara et al., 2018), suggesting that PEs

constitute a general translational learning signal. Emotional, in particular aversive, events are frequently characterized by unpredictability (Herry et al., 2007; Seligman et al., 1971; Trapp et al., 2018), which often give rise to PEs. Accordingly, a growing body of research suggests that PEs related to aversive events, i.e., when an aversive outcome occurs unexpectedly or when an expected aversive outcome is unexpectedly omitted, play a role in modulating the strength of memory (Kalbe & Schwabe, 2020, 2022b). Importantly, these effects are distinct from the well-established memory boost associated with novel or salient events (Schlüter et al., 2019; Sinclair & Barense, 2018). Aversive PEs, particularly uPEs, enhance memory for preceding stimuli that are inherently neutral, suggesting a different underlying mechanism (Kalbe & Schwabe, 2020, 2022b; Rouhani & Niv, 2021; Rouhani et al., 2018).

Specifically, it has been demonstrated that uPEs related to aversive events enhance episodic memory. An initial study employed a combined Pavlovian fear conditioning and incidental memory paradigm in which participants were asked to predict the occurrence of electric shocks (i.e., shock vs. no-shock) for two different stimulus categories (Kalbe & Schwabe, 2020). The results showed that large uPEs were linked to enhanced memory for stimuli preceding the PE. In a follow-up study, using an adapted version of this paradigm that also assessed sPE, the authors found that negative PEs related to aversive events (corresponding with 'unexpected shock omissions') were associated with an improvement of memory for preceding stimuli whereas positive aversive PEs (corresponding with 'unexpected shock occurrences' impaired memory (Kalbe & Schwabe, 2022b). Positive PEs were assumed to lead to heightened processing of the unexpected electric shock which in turn diverted attention away from the encoded stimulus and disrupted mnemonic processing (Iglesias et al., 2013; Kalbe & Schwabe, 2022b; Pearce & Hall, 1980), whereas negative PEs may serve as relief signals due to the omission of an expected outcome, triggering neuromodulatory responses that prioritize memory consolidation via the LC-VTA-hippocampal loop (Takeuchi et al., 2016). This finding suggests that effects of sPEs related to aversive events on memory depend on their direction (positive vs. negative PEs) which are also linked to different underlying neural network activations (Kalbe & Schwabe, 2022b). It is important to note that the definition of sPEs differed between the reward and aversive domains. In reward-based paradigms, positive PEs mostly refer to outcomes that are better than expected (e.g., receiving an unexpected reward), whereas negative PEs reflect outcomes that are worse than expected (e.g., reward omission or punishment). In contrast, within the aversive domain, as also applied in the present studies, positive PEs correspond to unexpected occurrences of an aversive outcome (i.e., shock), while negative PEs reflect unexpected omissions of such outcomes.

The PE-driven modulation of memory may arise from the role of PEs in marking event boundaries, segmenting a continuous experience into discrete episodes (Laing & Dunsmoor, 2025; Rouhani et al., 2020; Zacks et al., 2007; Zacks & Swallow, 2007). This aligns with predictive coding frameworks, where the brain continuously compares incoming sensory data to internal models and updates these models upon detecting prediction violations (Barrett & Simmons, 2015; Friston, 2005). These mismatches engage brain regions such as the mPFC, angular gyrus, and precuneus, which are involved in schema updating and event segmentation (van Kesteren et al., 2012; Vogel et al., 2018). Consequently, events characterized by large PEs are more likely to be stored as independent memory episodes instead of being incorporated into pre-existing schema networks suggesting that PEs lead to a shift in mnemonic processing (Bein et al., 2020).

Given that aversive events, e.g., shocks, also contain an arousal component due to their unpredictability and emotional valence, there is some overlap between PE-related and arousal-related effects on memory formation (Braem et al., 2015; Ferreira-Santos, 2016; Ganesh et al., 2024; Rouhani et al., 2023). PE-related arousal has been shown to drive noradrenergic and dopaminergic neuromodulation, particularly via phasic activation of the LC and VTA affecting plasticity in memory-relevant structures like the HC (Clewett & Murty, 2019; Takeuchi et al., 2016). Moreover, aversive PEs may engage additional structures such as the amygdala, insula and periaqueductal gray which are critical for processing interoceptive threat signals and modulating attention and salience (Kolada et al., 2023; McHugh et al., 2014; Roy et al., 2014; McCutcheon et al., 2019). Hence, recent evidence was further able to disentangle the link between PEs, physiological arousal and episodic memory (Kalbe & Schwabe, 2020, 2022b). Indeed, arousal related to the outcome (shock vs. no-shock) was associated with enhanced memory for preceding stimuli. This finding dovetails with evidence pointing out that noradrenaline released during emotional arousal modulates activity in the basolateral amygdala which has been shown to be critically involved in emotional memory formation (Hermans et al., 2014; LaBar, 2003; Phelps, 2004; Roozendaal & Hermans, 2017). Subsequently, memory formation processes are strengthened in areas such as HC and prefrontal cortex (PFC; LaBar & Cabeza, 2006; McGaugh & Roozendaal, 2002; Pape & Pare, 2010). Intriguingly, there is striking evidence that PEs facilitate memory formation beyond the mere effects of physiological arousal suggesting that there is an arousal-based and a prediction-based route to memory (Kalbe & Schwabe, 2020, 2022b; Rouhani et al., 2023). However, the prediction-related and arousal-related routes to memory may not be entirely independent. Emerging evidence suggests that prediction errors can manifest in outcome-related arousal

(Spoormaker et al., 2012) and indicates that physiological arousal may be modulated by environmental uncertainty, which may in part arise from PEs (de Berker et al., 2016). These findings indicate that arousal may partly be driven by PE signals. Nevertheless, the PE-related and arousal-related routes to episodic memory seem to be at least partly independent of each other (Kalbe & Schwabe, 2020, 2022b). Although the PE effects on episodic memory formation are fundamental to our understanding of adaptive memory, the brain mechanisms underlying the impact of PEs related to aversive events on memory for preceding events are less clear.

1.3 Mechanisms of prediction error-related memory modulations

In general, PEs involve a complex assortment of processes and neural structures, including the striatum encoding dopaminergic projections induced by PEs, the mPFC and anterior cingulate cortex (ACC) that are associated with error monitoring, and the dorsolateral prefrontal cortex (dlPFC) linked to the active maintenance of information in short-term memory (Alexander & Brown, 2019; Calderon et al., 2021; Delgado et al., 2008; Gläscher et al., 2010; Haque et al., 2020; Li et al., 2011; Maier et al., 2019; Matsumoto et al., 2007; Pine et al., 2018). Although PEs have received attention regarding their general neural correlates, it remains unclear which neural mechanisms drive their modulatory effects on episodic memory. However, initial evidence from a functional magnetic resonance imaging (fMRI) study (Kalbe & Schwabe, 2022b) suggests that the PE effects on memory are associated with a reduced activation of the MTL, which is implicated in memory formation for expectancy- and schema-congruent information and associated with the schema network (Bein et al., 2020; Davachi & Wagner, 2002; Eichenbaum, 2004; van Kesteren et al., 2012; Vogel et al., 2018), as well as an enhanced crosstalk of the salience and frontoparietal network (Kalbe & Schwabe, 2022b). These findings are in line with the notion that PEs may create event boundaries interrupting the sequential integration of events preferentially by downregulating the schema network and upregulating the salience network. While fMRI provides high spatial resolution to identify relevant brain areas and network activations, its temporal resolution is rather limited, making it hard to capture neural processes occurring around the PE event.

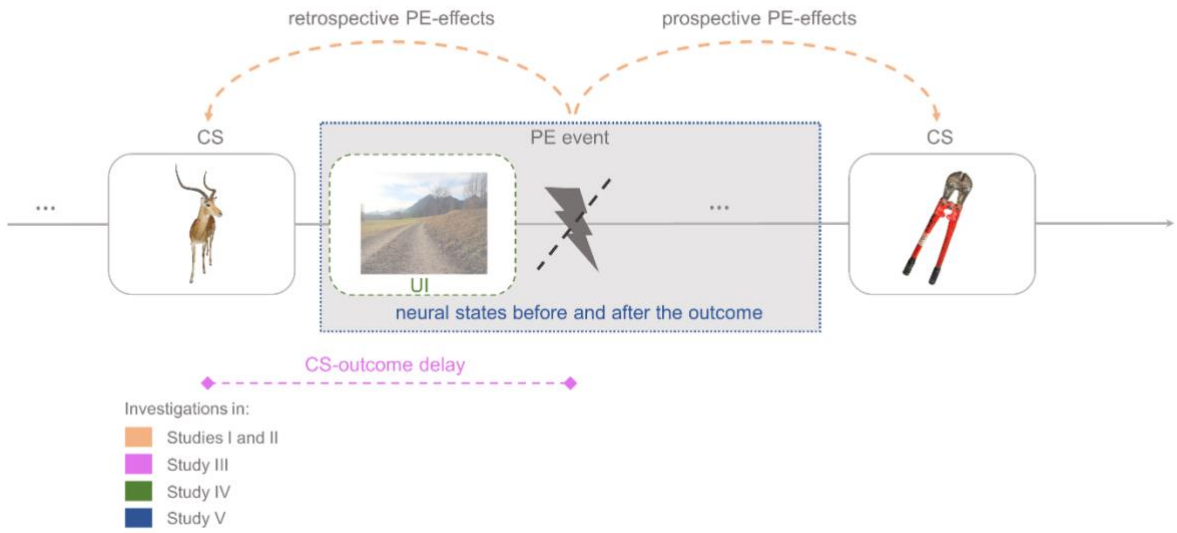
1.4 Research scope and rationale of the current thesis

Overall, PEs signal a mismatch between expected and actual outcomes and are thus a fundamental computational signal in reinforcement learning models (Gläscher et al., 2010; Glimcher, 2011; Schultz et al., 2007). Accumulating evidence suggests that PEs emerge as a core mechanism that guides not only learning but also memory formation (Antony et al., 2021; Corlett et al., 2022; Ergo et al., 2020; Rouhani et al., 2023). Recent theoretical and empirical

advances demonstrate that PEs, particularly those arising from deviations in threat, can modulate episodic memory encoding, at least partly independent of physiological arousal (Greve et al., 2017; Haque et al., 2020; Jang et al., 2019; Kalbe & Schwabe, 2020, 2022b; Rouhani et al., 2018; Rouhani & Niv, 2021). From a computational perspective, PEs dynamically enable the updating of internal models to guide future behavior (Clark, 2013; Niv, 2009). However, the cognitive and neural mechanisms through which aversive PEs influence memory remain poorly understood. The overarching aim of this thesis is to investigate the cognitive and neural mechanisms by which PEs shape episodic memory formation, particularly in the context of aversive learning (see Figure 1).

Figure 1

Schematic representation of experimental modulations in the studies



Note. Participants underwent a combined Pavlovian fear-conditioning and incidental memory paradigm which was slightly modified over the course of the Studies I to V. Trialwise, participants had to indicate if (Studies I and II) and how likely (Studies III to V) they expected a shock to follow a conditioned stimulus (CS) inducing a prediction error (PE). Studies I and II aimed at investigating retrospective and prospective PE-effects on memory formation. In Study III, the delay between CS and outcome (shock vs. no-shock), i.e., the PE event, was critically varied between 0 and 10 s to examine the time-dependency of PE-effects on memory. In Study IV, an uninformative stimulus (UI) was added between the CS and the outcome to investigate the selectivity of the PE-induced memory modulations. Study V also used a variable CS-outcome delay (0-10 s) but no UI and critically investigated the interaction of neural states surrounding the PE event and memory modulations. Three dots indicate the start of a new trial. Pictures taken from "The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research" by Brodeur et al. (2010) and from "Introducing the Open Affective Standardized Image Set (Oasis)" by Kurdi et al. (2016). CC BY 4.0.

First, we wanted to assess the robustness of PE-related effects on memory within the aversive domain. In Studies I and II, we aimed at replicating the previously reported PE-effects

on episodic memory (Kalbe & Schwabe, 2020, 2022b; see Figure 1). To this end, participants completed a combined incidental encoding-fear learning task in which they predicted whether a stimulus would be followed by an electric shock using one excitatory conditioned stimulus (CS^+) and one inhibitory conditioned stimulus (CS^-). Simultaneously, we measured skin conductance responses (SCRs) as an index of physiological arousal allowing us to probe potential arousal-related effects on memory and to disentangle the respective contributions of PEs and arousal to episodic memory formation. 24 hours after encoding, memory was tested in a surprise recognition test. Specifically, we sought to replicate the previously reported effect of PEs related to aversive events, that were unsigned and binary, i.e., attaining 0 for correct predictions and 1 for incorrect predictions, on the subsequent memory for predictive stimuli presented before the PE using comprehensive behavioral analysis. Second, we investigated whether the memory-enhancing effects of PEs are limited to stimuli that immediately preceded the PE or whether PEs may boost memory also for stimuli encountered shortly after a PE. In particular, we asked whether PEs may exert not only the previously reported retrospective effects (Kalbe & Schwabe, 2020), but also prospective effects on memory formation (see Figure 1).

In Study III, we aimed to further elucidate the temporal dynamics of PE effects on memory. Previous studies typically employed short and fixed stimulus-outcome intervals (see Kalbe & Schwabe, 2020, 2022b), leaving it unclear whether the retrospective PE effect on memory relies on the close temporal proximity between the predictive cue and the unexpected outcome, or whether the effect is robust to temporal delays. To address this, we systematically manipulated the delay between the predictive stimulus and the outcome that triggered a PE, i.e., the CS-outcome delay. This allowed us to test whether maintaining the representation of the predictive cue over a longer interval is necessary for occurrence of the memory-enhancing effects of PEs. From a cognitive perspective, increasing this CS-outcome delay would tax working memory systems which are required to keep the stimulus representation accessible for integration with the outcome event (Cohen et al., 2014). If predictive stimuli must be actively maintained until the PE occurs to benefit from enhanced encoding, then increasing the interval should attenuate or abolish this effect. This would implicate working memory processes, such as active maintenance, as crucial for PE-related memory modulation and echoes broader debates in computational neuroscience regarding the interplay between working memory and reinforcement learning processes. Notably, Collins and Frank (2012) proposed that behaviors often attributed to reinforcement learning may reflect contributions from working memory, especially in tasks involving short trial sequences or rapid updating. Indeed, it has been

demonstrated that working memory and reinforcement learning are not strictly separable systems but dynamically interact, with working memory often supporting slower reinforcement-based updating depending on task demands and cognitive load (Collins & Frank, 2012). In line with this view, working memory may serve as a cognitive workspace that temporarily holds the predictive stimulus in a state that allows it to be updated in response to a (surprising) outcome. Here, we examined (i) the extent to which PE effects on memory depend on the time interval between the stimulus and the PE, i.e., CS-outcome delay (see Figure 1), and (ii) whether these effects depend on the interval between encoding and testing, i.e., whether the PE effects emerge during memory encoding or during consolidation. To these ends, Study III manipulated the delay between CS and outcome across a range of 0 to 10 s and tested recognition memory either 24 hours after encoding, as in Studies I and II, or immediately after encoding. In addition to these timing-related changes, we introduced two key modifications to our experimental paradigm from Studies I and II: First, participants were presented with three CSs, with one CS^- and two CS^+ with different shock contingencies. Second, participants were asked to rate their shock expectancy on a continuous scale from 0 to 100 enabling us to also measure sPEs. Both adjustments served to achieve an adequate distribution of positive and negative PEs, which have been proposed to exert distinct effects on memory formation (Kalbe & Schwabe, 2022b; Rouhani et al., 2023; Rouhani & Niv, 2021).

However, a critical question concerns whether this PE-induced memory enhancement is exclusive to the preceding predictive stimulus itself. To the best of our knowledge, no study yet has explored the link between the unexpected outcome and the preceding information by adding an uninformative, i.e., unpredictive, component. There are two contradictory, possible frameworks that may explain how the memory of information associated with a PE is strengthened. When an individual experiences an unexpected outcome, memory might generally be enhanced within a certain time window to collect more evidence for future predictions which could be retroactively generated by the mismatch between a prediction and the observed outcome. This would imply superior memory not only of an informative target, but also uninformative stimuli linked to the PE event. In contrast, a conceptual alternative would be a causal, specific, and exclusive link between the PE and its predictive information. The causal link would imply superior memory only of the informative target and no memory advantage of uninformative stimuli. In Study IV, we addressed this question by testing (i) whether the PE-induced memory enhancement is restricted to the predictive stimuli carrying relevant information about the outcome or (ii) whether it reflects a non-selective encoding boost (see Figure 1). To test these competing hypotheses, we used a modified version of the paradigm

from Study III: In some of the encoding blocks entirely uninformative stimuli (UI) were presented between the predictive stimulus and the outcome (i.e., potential PE) enabling us to assess whether the memory enhancement extends to these non-predictive stimuli. Importantly, we ensured that the UI were introduced only after the association between the predictive cues and the outcomes was established, i.e., making use of the blocking effect (Fanselow, 1998). We hypothesized that if PEs induce a transient window of enhanced encoding, memory should also be increased for uninformative stimuli presented shortly after the predictive stimulus. Alternatively, if the PE effects are specific to the encoding of the predictive stimulus, memory should not be increased for the uninformative stimulus presented between the predictive stimulus and the PE.

While Studies I to IV focused on behavioral investigations of the cognitive mechanisms underlying PE-induced memory formation, the neural mechanisms remain elusive. Specifically, it is unclear how the brain integrates unexpected outcomes to prioritize preceding, neutral events into long-term memory. Understanding this process at a neural level is crucial for building computational models of adaptive memory and for identifying potential intervention targets in memory disorders. Previous fMRI work (Kalbe & Schwabe, 2022b) provided first hints, such as an increased crosstalk between frontoparietal networks and the salience network, but its temporal resolution is insufficient to resolve the rapid dynamics likely involved in PE-driven memory formation. In Study V, we aimed to further explore the neural states that may support PE-driven memory enhancement. Specifically, two plausible mechanisms may underly these effects (see Figure 1): First, PEs might trigger a post-encoding reactivation of the predictive stimulus thereby enhancing its encoding. This mechanism is supported by evidence linking post-encoding reactivation to later recall (Staresina et al., 2013; Tambini et al., 2020). Alternatively, the neural state shortly before the PE could drive the memory enhancing effects by sustaining a neural representation of the predictive stimulus when the PE occurs. This mechanism is in line with behavioral tagging models proposing that pre-activated representations can be strengthened by salient events (Kalbe & Schwabe, 2022a; Moncada et al., 2015). Neural markers of this maintenance mechanism may include alpha oscillations that are linked to attention (Payne & Sekuler, 2014), theta oscillations that are associated with binding of associative memory and reinstatement memory as well as a reactivation of the stimulus representations (Kota et al., 2020; Nyhus & Curran, 2010; Staudigl & Hanslmayr, 2013). To these ends, we employed the same paradigm as in Study III with a variable CS-outcome delay and combined this task with multivariate electroencephalography (EEG) analysis and comprehensive behavioral analysis. However, correlational EEG findings cannot

establish whether these maintenance-related processes are necessary for the PE effects on memory. To investigate whether neural maintenance mechanisms play a causal role in PE-driven memory enhancement, we applied inhibitory continuous theta burst stimulation (cTBS) over the superior parietal cortex, given its significant role in attentional control, working memory maintenance, and goal-directed updating (Corbetta et al., 1995; D'Esposito & Postle, 2015; Ester et al., 2015; Koenigs et al., 2009; Wager & Smith, 2003). Critically, the stimulation was administered before participants completed the fear-conditioning and incidental memory task. If sustained stimulus maintenance is necessary for the PE effect on memory to emerge, then disrupting this region before the incidental encoding-fear learning task should reduce or even abolish the memory enhancement for stimuli preceding a PE.

2 Retrospective and prospective effects of PEs related to aversive events on subsequent memory

This chapter was published in modified form in: 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

2.1 Background

Every day, we are constantly exposed to a continuous stream of information, yet only selected experiences are retained in long-term memory. One well-established path to memory formation is emotional arousal: Emotionally charged events are more vividly remembered, likely due to amygdala-driven modulation of hippocampal memory processes (Cahill & McGaugh, 1998; McGaugh, 2018). Beyond arousal, recent work suggests that PEs also serve as drivers of memory formation (Antony et al., 2021; Kalbe & Schwabe, 2020; Rouhani et al., 2018). While PE- and arousal-related memory effects sometimes overlap, PE-driven memory enhancements seem to involve distinct neural processes (Kalbe & Schwabe, 2020, 2022b; Rouhani et al., 2023), bearing practical implications. Whereas arousal-related effects on memory might be altered pharmacologically, PE-related effects could be targeted by modulating expectations.

Critically, PE-related memory effects cannot be fully explained by attention or salience alone. Unlike "oddball" effects, where surprising stimuli themselves are better remembered, PE effects selectively enhance memory for predictive stimuli that precede the surprising outcome (Kalbe & Schwabe, 2020, 2022b; Rouhani et al., 2018; Rouhani & Niv, 2021), suggesting a mechanism that retrospectively modulates memory formation. A major question related to the PE effects on memory concerns, however, whether the PE-related memory boost is selective to the predictive stimulus preceding the PE event. If so, then there should be no memory enhancement for either stimuli following the PE. Alternatively, it could be hypothesized that PEs open a transient window of enhanced mnemonic processing that may also enable better memory for nonpredictive stimuli, suggesting prospective PE effects. Whereas the selective memory enhancement would involve mnemonic efficiency, the latter would reflect a "better-safe-than-sorry" mechanism making sure that all stimuli that occurred in the surrounding of an unexpected emotional event are preferentially stored in memory.

To address this, Studies I and II employed an incidental encoding-fear conditioning paradigm, they saw a stream of initially neutral stimuli from different categories that were associated with a differential probability of electric shocks. We asked participants to predict the

occurrence of an electric shock which then allowed us to calculate PEs based on their expectation and its deviation from the actual outcome (shock vs. no-shock). Memory for the presented pictures was tested 24 hours later. Study I aimed to replicate the previously reported retrospective PE effects on memory and to test whether the memory-enhancing PE effect extends to stimuli presented after the PE, i.e., prospective PE effects. Study II served to replicate the findings of Study I. Additionally, we measured autonomic arousal in both Studies to test whether the observed PE effects go beyond the well-known effects of arousal on memory formation.

2.2 Study I

The objectives of Study I were two-fold: first, this study aimed to replicate the enhancing effects of PEs related to aversive events on the memory for (predictive) stimuli that preceded the PE. Second, we aimed to test whether the memory-enhancing effects of PEs are limited to stimuli that immediately preceded the PE or whether PEs may enhance memory also for stimuli encountered shortly after a PE. In other words, we asked whether PEs may, in addition to the previously reported retrospective effects (Kalbe & Schwabe, 2020), exert prospective effects on memory formation. To this end, participants completed a combined incidental encoding-fear learning task in which they predicted whether a stimulus would be followed by an electric shock. During this task, we measured skin conductance responses (SCRs) as an indicator of physiological arousal, enabling us to probe potential arousal effects on memory formation. Twenty-four hours after encoding, memory was tested in a surprise recognition test.

2.2.1 Methods

2.2.1.1 Participants

Eighty-four healthy volunteers with normal or corrected to normal vision participated in this study (age: M = 25.11 years, SD = 3.57 years, range = 18–33 years). Participants were fluent German speakers, had no current illnesses, no life-time history of any mental or neurological disorders and did not take any prescriptive medication as assessed in a standardized telephone interview. Furthermore, women being pregnant were excluded from participation. Six participants were excluded from the analyses because they did not return for the second experimental day or due to technical failure during the experiment, resulting in a final sample of n = 78. This sample was part of a larger study on emotional learning processes (Kalbe & Schwabe, 2022a). The sample size was based on an a priori power calculation using G*Power (3.1.9.6; Faul et al., 2009). Based on previous research by Kalbe and Schwabe (Kalbe & Schwabe, 2022a), we assumed d_z = .45 as a point estimate for the expected PE effects. The

power calculation showed that a sample of at least 67 participants is required to detect an effect of the expected size in a two-tailed paired *t*-test with a statistical power of 0.95. In line with these assumptions, a post hoc power simulation using the R-package simR (Green & MacLeod, 2016) for the observed effects of subsequent PEs and our final sample size of 78 participants yielded a power of 0.99 based on 1000 simulations. All participants provided written informed consent before participation and received a monetary reimbursement of 20€ at the end of the study. The study was approved by the ethics committee of the Faculty of Psychology and Human Movement Science at the Universität Hamburg and carried out in line with the Declaration of Helsinki.

2.2.1.2 Materials

For this study, we used the same stimulus set as in Kalbe and Schwabe (2022a). It consisted of 180 color photographs of animals and 180 color photographs of tools isolated on white backgrounds. These photographs were taken from already existing databases (Bank of Standardized Stimuli; Brodeur et al., 2010, 2014), McGill Calibrated Color Image Database (Olmos & Kingdom, 2004), SUN database (Xiao et al., 2010), Konklab (Konkle et al., 2010), which were developed for nonemotion research on cognition, vision, and psycholinguistics. All stimuli were assumed to be of neutral valence, and each object or animal represented a unique exemplar of its category. Importantly, each photograph was only presented once and there were not two different photographs of, for example, cats or screwdrivers. From this pool, 30 photographs of animals and 30 photographs of tools were randomly drawn and used on the first experimental day and 120 photographs for encoding tasks unrelated to the purpose of the present study. These unrelated tasks took place before and after the relevant learning paradigm and importantly, did not contain any predictions or aversive events, thus making it highly unlikely that these tasks interfered with the memory paradigm of interest here. The remaining 180 photographs served as lures for the surprise recognition test on the second experimental day. The order in which individual items were presented was randomized across participants.

2.2.1.3 Procedure

The study consisted of 2 days, with an encoding session on the first experimental day and a recognition test on the second experimental day, 24 hr later (see Figure 2).

Upon arrival on the first experimental day, participants provided written informed consent and received written instructions indicating that they were going to see a series of photographs of animals and tools and that some of them might be followed by a brief electric shock. Participants were instructed to try to predict whether a shock would follow the current

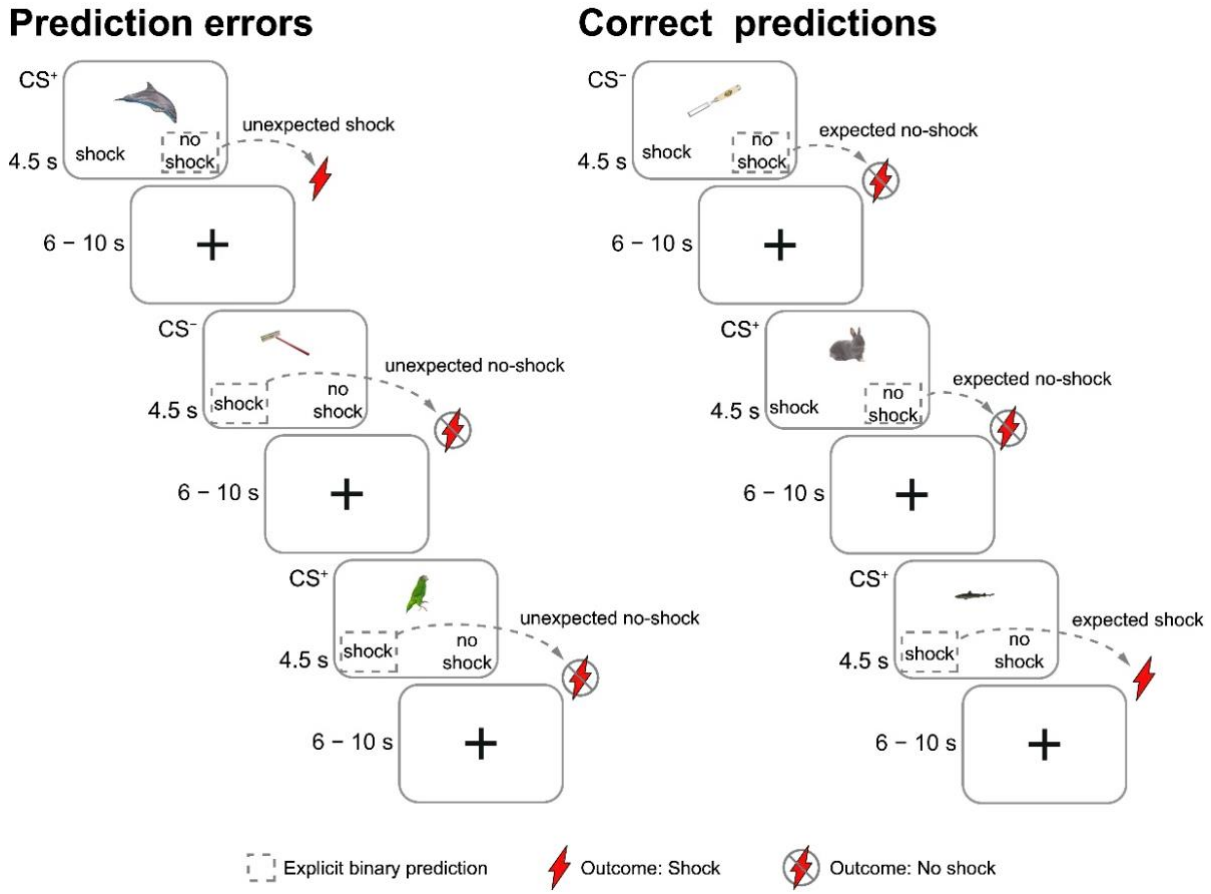
photograph (“shock” vs. “no shock” response). Importantly, participants were neither informed about the shock contingencies, nor that a subsequent memory test would follow on the second day.

To record SCRs as indicator of physiological arousal and conditioned fear, electrodes were placed on the distal phalanx of the second and third finger of the left hand. Skin conductance was measured using the MP-160 data acquisition and analysis BIOPAC system (BIOPAC systems, Goleta, California, United States). For electrical stimulation, we used the STM-200 stimulator module connected to the MP-160. A stimulation electrode was placed on the back of the right hand near the wrist. Stimulation intensity was adjusted individually to be unpleasant but not painful using a standardized procedure. More specifically, a total of twelve 200-ms single pulse shocks were administered, with an initial intensity of 10 V. After each trial, participants rated whether the received shock had been painful in a forced choice fashion using the “left” (“not painful”) and “right” (“painful”) keys. Whenever a shock was rated as not painful, its intensity for the next trial was increased slightly. Analogous, when participants rated the shock as painful, it was decreased slightly. The aim was to choose an intensity that was unpleasant but not painful to the participants.

During the encoding session, 30 photographs of animals and 30 photographs of tools were presented in a pseudorandomized order, so that no more than three pictures of the same category appeared in a row. In each trial, a photograph was shown in the center of a computer screen for 4.5 s, during which participants were asked to make their binary prediction about the occurrence of an electric shock using the “1” and “2” buttons on the keyboard, corresponding to *no shock* and *shock*, respectively (see Figure 2). Critically, shock contingencies were linked to the item category, such that one image category served as excitatory conditioned stimulus (CS^+) and the other one served as inhibitory conditioned stimulus (CS^-). The assignment of tools or animals as CS^+ or CS^- was counterbalanced across participants. In CS^+ trials the shock contingency was two-thirds, resulting in 20 out of 30 trials that included a shock. In CS^- trials, no shocks were administered. Each trial was followed by a black fixation cross centered on white background for 8 ± 2 s, which enabled measuring the relatively slow SCRs elicited by the photographs and the shocks. After the experimental task which lasted approximately 12 min, electrodes were removed and participants were asked to rate the intensity of the shocks on a scale from 1 (*not unpleasant at all*) to 10 (*extremely unpleasant*).

Figure 2

Experimental Procedure of Study I and II



Note. Participants completed a combined Pavlovian fear-conditioning and incidental memory paradigm. They saw initially neutral pictures from two different categories, one of which was associated with receiving an electric shock with a shock-contingency of 67% (CS^+). In each trial, they were asked to make their binary prediction about the occurrence of an electric shock. Critically, in Study I, the 200ms-shock occurred after stimulus offset while the shock coterminated with the predictive stimulus in Study II (4.3 sec after stimulus onset). On a second experimental day, memory was tested for these in items in a surprise recognition test. Pictures taken from "The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research" by Brodeur et al. (2010) and from "Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos" by Brodeur et al. (2014). CC BY 4.0.

On Experimental Day 2, 22–26 hr after the encoding session, participants returned for a surprise recognition test. First, they completed a short questionnaire to assess whether they anticipated a memory test and then rated how surprised they were about the recognition test on a scale from 1 (*not surprised at all*) to 5 (*very surprised*). Next, they received written instructions explaining details of the recognition test. During the recognition test, participants were presented all pictures they had seen on Experimental Day 1 (90 pictures of animals and 90 pictures of tools) as well as 180 "new" pictures (90 pictures of animals and 90 pictures of tools) that had not been presented on the previous day. Each trial started with a central black fixation cross on a white background for 1.5 ± 0.5 s, followed by an "old" or "new" picture presented centrally on the computer screen. For each item, participants had to indicate whether

the currently presented picture was *definitely old*, *maybe old*, *maybe new*, or *definitely new* by pressing the "1," "2," "3," or "4" button on the keyboard, respectively. There were no time restrictions for participants' responses.

2.2.1.4 Data analysis

For each trial, we derived binary PEs which were calculated as the absolute value of the difference between participants' explicit binary shock expectancy ratings (coded 0 when no shock was expected and coded 1 when a shock was expected) and the actual outcome of the trial (coded 0 when no shock occurred and 1 when a shock occurred in the current trial). Therefore, the resulting unsigned PE is also binary, attaining 0 for any correct prediction (i.e., either an expected shock or an expected shock omission) and 1 for any incorrect prediction (i.e., either an unexpected shock or an unexpected shock omission). SCRs were analyzed using Continuous Decomposition Analysis in Ledalab Version 3.4.9 (Benedek & Kaernbach, 2010). We derived the average phasic driver within a specified response window. First, the skin conductance signal was downsampled to a resolution of 50 Hz and optimized using four sets of initial values to increase the goodness of the model. For the anticipatory SCR, the response window was set from 0.5 to 4.5 s after stimulus onset. For the outcome-related SCR, the response window was set from 4.5 to 7.9 s after stimulus onset. More importantly, aversive electric stimulation always occurred exactly 4.5 s after stimulus onset; thus, leaving the anticipatory SCR unbiased by the shock itself. The minimum amplitude threshold was set to 0.01 µS for both the anticipatory and the outcome-related SCR. Resulting estimates of average phasic driver within each response window were returned in μS. Notably, these estimates are sensitive to interindividual baseline skin conductance differences because of physiological factors such as the thickness of the corneum (Figner & Murphy, 2011). To account for these interindividual baseline differences, we therefore standardized both the anticipatory and the outcome-related SCR by dividing the average phasic driver estimated in each trial by the maximum average phasic driver for each participant observed in each of the 60 trials. Due to an experimenter error, SCR data for one additional participant was missing.

To investigate how PEs impacted the ability to recognize pictures presented during incidental encoding on the next day, we fitted generalized linear mixed models (GLMMs) with a logit link function using the lme4 R package (Bates et al., 2015). Compared with a "classic" analysis of proportions of binary recognition per condition and per participant, GLMMs have several advantages, such as increased statistical power and being less prone to spurious results (Dixon, 2008; Jaeger, 2008). Following guidelines to maximize the generalizability of these models, we included the maximal random effects structure, treating subjects as random effects

for both the intercept and all slopes of the fixed effects included in the model (Barr et al., 2013). The recognition of an individual item was treated as the binary dependent variable, coded "0" for misses and "1" for hits. In line with previous research on episodic memory (Bartlett et al., 1980; Kalbe & Schwabe, 2022b), our analysis focused on high-confidence responses, that is, only trials in which participants indicated that they were very sure were considered. Such high-confidence recognitions have been linked to a hippocampus-based recollection rather than only familiarity with an item, which is assumed to depend on the perirhinal cortex (Eichenbaum et al., 2007). Accordingly, we computed hit rates (i.e., recognizing an item as "surely old") and category-based false alarm rates referring to conditioning on category level. (CS+ vs. CS−; please note that new items of the category that had been a CS+ during encoding have never been paired with the shock.) We fitted models using different sets of independent variables, including subsequent PEs, previous PEs, anticipatory and outcome-related arousal, and the explicit shock prediction. To further elucidate the effects of PEs on episodic memory, we computed *subsequent PEs* and *previous PEs* to investigate retro- and prospective effects of PEs on subsequent memory. In the analysis of retrospective PE effects, subsequent PEs were linked to the memory of the preceding stimulus in the same trial, that is, PE of Trial 3 referred to the predictive item of Trial 3. In the analysis of potential prospective PE effects, previous PEs were linked to the memory of the stimulus in the following trial, that is, the PE in Trial 3 was used as predictor for subsequent memory of the item presented in Trial 4.

2.2.1.5 Transparency and openness

The materials, data, and R analysis scripts are publicly available on the Research Data Management System of University of Hamburg and can be accessed at https://www.fdr.uni-hamburg.de/record/14147 (Loock et al., 2024). This study was not preregistered.

2.2.2 Results

2.2.2.1 Successful fear conditioning

An analysis of SCR data confirmed that fear was successfully induced for CS^+ items (see Figure 3A). On average, participants showed significantly higher anticipatory SCRs to CS^+ items ($M = 0.21$, $SD = 0.01$) compared to CS^- items ($M = 0.17$, $SD = 0.01$); $t(76) = 3.87$, $p < .001$, $d = 0.42$. Furthermore, outcome-related SCRs were significantly higher for shocked items ($M = 0.42$, $SD = 0.01$) compared to unshocked items ($M = 0.10$, $SD = 0.01$), $t(76) = 21.29$, $p < .001$, $d = 3.15$. Explicit shock ratings showed that participants learned the shock contingencies very well. On average, incorrect predictions were made in 29% ($SD = 0.11$) of all trials with substantially more PEs for CS^+ ($M = 0.47$, $SD = 0.10$) compared to CS^- items ($M = 0.11$, $SD =$

0.16), $t(77) = 19.05$, $p < .001$, $d = 2.71$ (see Figure 3B). As expected, PEs decreased as the task progressed, $r(58) = -0.53$, $p < .001$.

2.2.2.2 General memory performance

As expected, participants on average were moderately surprised by the recognition test (see Supplemental Material). Overall, participants performed very well in the recognition task, as indicated by significantly higher hit rates ($M = 0.42$, $SD = 0.17$) than category-based false alarm rates ($M = 0.23$, $SD = 0.10$), $t(77) = -7.54$, $p < .001$, $d = 1.37$ (see Figure 3C). Importantly, the category-based false alarm rate for the CS+ items ($M = 0.23$, $SD = 0.12$) was comparable to the false alarm rate for items from the CS− category ($M = 0.23$, $SD = 0.13$), $t(77) = 0.51$, $p = .611$, $d = 0.07$. Note that new items have never been paired with a shock and false alarms relate to the whole category. As expected, the average hit rate for items from the CS+ category ($M = 0.48$, $SD = 0.20$) was significantly higher than for items from the CS− category ($M = 0.35$, $SD = 0.20$), $t(77) = 5.59$, $p < .001$, $d = 0.65$. These finding dovetails with the assumption that memory advantages for CS+ items are attributed to increased physiological arousal. When we used the signal detection theory-based parameter *d'*, *d'* was increased for CS^+ items ($M = 1.50$, $SD = 0.55$) compared to CS^- items ($M = 1.06$, $SD = 0.53$; $t(75) = 6.24$, $p < .001$, $d = 0.81$).

In order to test whether participants expected the recognition test on the second experimental day, they rated their level of surprise related to the recognition test on a scale ranging from 1 (not surprised at all) to 5 (very surprised). On average, participants were moderately surprised ($M = 3.05$, $SD = 0.97$). Five participants chose the 'not surprised at all' option. Because excluding them did not affect the results, they were included in all analyses.

2.2.2.3 Modelling recognition performance at item level

So far, we showed that CS^+ items were better remembered after 24 h than CS^- items. To test whether PEs drive the emotional memory enhancement, we computed GLMMs at item level treating the binary recognition of an item as the dependent variable.

In a first minimal model, we tested whether uPEs that followed a CS contribute to subsequent recognition of this CS item. Therefore, we treated the binary subsequent PE as the sole independent variable to predict the binary recognition of an item. This revealed that episodic memory was indeed enhanced for trials in which an incorrect shock prediction for the predictive target has been made ($z = 4.90$, $p < .001$, $\beta = 0.90$; see Figure 3D). To rule out that those PE effects were due to the explicit shock prediction, we also computed a model that included the explicit shock prediction and the binary subsequent PE as independent variables to predict the binary item recognition. Critically, even after controlling for the shock prediction

we found a memory enhancing effect of subsequent PEs in this model ($z = 4.72$, $p < .001$, $\beta = 0.76$). To investigate the possibility that the effects of physiological arousal and the effects of PEs on memory might be confounded, we added both measures of arousal (i.e., anticipatory and outcome related SCRs) to the minimal model that featured only the binary subsequent PE as the sole independent variable. This model revealed no significant effect of anticipatory SCRs on item recognition ($z = -1.47$, $p = .142$, $\beta = -0.54$). Larger outcome-related SCRs, on the other hand, were associated with better item recognition ($z = 4.20$, $p < .001$, $\beta = 1.45$). For subsequent PEs, our results showed that recognition was improved significantly ($z = 4.36$, $p < .001$, $\beta = 0.76$), suggesting that subsequent PEs enhanced memory even after controlling for arousal.

In a next step, we tested whether uPEs that preceded an item and were actually related to the previous item may exert also prospective effects, contributing to the recognition of the item following the PE. To this end, we treated the binary previous PE as the sole independent variable to predict the binary recognition of the following item. This revealed that episodic memory was indeed enhanced for items following a PE, $z = 3.48$, $p < .001$, $\beta = 0.57$. In addition, we also computed a model that included the explicit shock prediction and the previous PE as independent variables to predict the binary item recognition. Critically, even after controlling for the shock prediction the memory enhancing effect of previous PEs remained in this model ($z = 2.68$, $p = .007$, $\beta = 0.37$). Same as in the analysis of subsequent PE effects, we added both measures of arousal (i.e., anticipatory and outcome related SCRs) to the minimal model that featured only the binary previous PE as the sole independent variable to rule out confounds with physiological arousal. This revealed no significant effect of anticipatory SCRs on item recognition ($z = -1.48$, $p = .140$, $\beta = -0.53$). Larger outcome-related SCRs, on the other hand, were associated with better item recognition ($z = 4.19$, $p < .001$, $\beta = 1.50$). For previous PEs, there was a strong trend in the direction of memory enhancement, which did, however, not reach significance anymore ($z = 1.95$, $p = .051$, $\beta = 0.29$).

Additionally, we sought to determine whether the subsequent PE and previous PE reflect distinct mechanisms. Therefore, we added both the subsequent PE and the previous PE as independent variables. Estimates obtained revealed that subsequent PEs, $z = 4.85$, $p < .001$, $\beta = 0.74$, and previous PEs, $z = 3.11$, $p = .002$, $\beta = 0.44$, showed a positive relationship with item recognition, suggesting that both contribute distinctly to memory recognition. In a follow-up model, we added anticipatory arousal and outcome-related arousal as predictors. Again, anticipatory SCR did not influence item recognition significantly, $z = -1.58$, $p = .114$, $\beta = -0.62$, while larger outcome-related SCRs were associated with better item recognition, $z = 3.95$, $p < .001$, $\beta = 1.42$. While our analyses showed that item recognition was still significantly enhanced

by subsequent PEs, $z = 4.12$, $p < .001$, $\beta = 0.72$, previous PEs did not predict item recognition significantly anymore, $z = 1.20$, $p = .230$, $\beta = 0.18$, suggesting that PEs directly associated with a (preceding) item exert an effect that is distinct from the effect of arousal, while the effects of PEs on the memory of a subsequently presented item appear to be at least partly driven by outcome-related arousal.

2.2.3 Conclusion

The findings of this study replicate the previously reported enhancing effects of PEs associated with aversive events on subsequent memory for the stimulus encountered before the PE. In line with these previous reports (Kalbe & Schwabe, 2020, 2022b), these PE effects could not be explained by mere increases of arousal. Interestingly, beyond these retrospective memory enhancements of PEs, we obtained also first evidence that PE enhance memory not only for items preceding the PE but also for items that followed a PE. In contrast to the retrospective effects of PEs, however, these prospective PE effects appeared to be at least partly driven by (outcome-related) arousal.

2.3 Study II

Study I provided initial evidence for a prospective effect of PEs on subsequent memory for stimuli encountered shortly after the PE. Study II served to replicate this prospective PE effect on memory, as well as the retrospective PE effect that was shown to go beyond the well-established effects of arousal on memory.

2.3.1 Methods

2.3.1.1 Participants

Eighty-four healthy volunteers participated in this study (age: $M = 25.17$ years, $SD = 4.26$ years, range = 18–34 years). Exclusion criteria were the same as those in Study I. Three participants were excluded from the analyses due to technical failure during the experiment, resulting in a final sample of $n = 81$. None of the participants had participated in Study I. This sample is part of a larger study on emotional learning processes (Kalbe & Schwabe, 2022a). The target sample size was based on an a priori power calculation with identical parameters as in Study I, showing that a sample of 67 participants is sufficient to detect a medium-sized effect ($d_z = .45$) of subsequent PEs with a power of .95. In addition, we performed a post hoc power simulation using the R-package simR (Green & MacLeod, 2016). For the observed subsequent PE effects and our final sample size of 81 participants, it yielded a power of 0.99 based on 1000 simulations. All participants provided written informed consent before participation and received a monetary reimbursement of 30€ at the end of the study. The study was approved by

the ethics committee of the Faculty of Psychology and Human Movement Science at the University of Hamburg and carried out in line with the Declaration of Helsinki.

2.3.1.2 Materials

We used the same stimulus set as in Study I. The stimulus set consisted of 60 photographs (i.e., 30 animals and 30 tools) on the first experimental day and 180 photographs that were used as lures in the recognition test on the second experimental day. Same as in Study I, the order of stimulus presentation was randomized across participants. The assignment of photograph category (i.e., tools or animals) as CS^+ or CS^- was counterbalanced across participants.

2.3.1.3 Procedure

The procedure of Study II was largely identical to the procedure of Study I, except that we changed the timing of the electric shock during the incidental encoding-fear learning session to make our procedure more comparable with previous learning paradigms (Dunsmoor et al., 2015; see Kalbe & Schwabe, 2022a). Specifically, in Study II, a 200-ms-electric shock occurred 4.3 s after stimulus onset and thus coterminated with the predictive stimulus during the learning task.

2.3.1.4 Data analysis

The statistical analysis was identical to Study I.

2.3.1.5 Transparency and openness.

The materials, data, and R analysis scripts are publicly available on the Research Data Management System of University of Hamburg and can be accessed at https://www.fdr.uni-hamburg.de/record/14147 (Loock et al., 2024). This study was not preregistered.

2.3.2 Results

2.3.2.1 Successful fear conditioning

Descriptively, participants showed higher anticipatory SCRs to CS^+ items (M = 0.16, SD = 0.01) compared to CS^- items (M = 0.15, SD = 0.01). However, this descriptive difference was not statistically significant, $t(80) = 1.03$, $p = .308$, $d = 0.09$ (see Figure 3E). We also used a Through-to-Peak Analysis of the anticipatory SCR data, a more traditional approach of SCR analysis (Boucsein, 1992; Kalbe & Schwabe, 2022a), instead of a continuous decomposition analysis. This analysis showed that SCRs were significantly higher in response to CS+ compared to CS- items, ($t(80) = 2.75$, $p = .007$, $d = 0.27$). Outcome-related SCRs were significantly higher for shocked items (M = 0.34, SD = 0.01) compared to unshocked items (M = 0.10, SD = 0.01), $t(80) = 16.46$, $p < .001$, $d = 2.21$). Explicit shock ratings showed that

participants learned the shock contingencies very well. On average, incorrect predictions were made in 29% (SD = 0.12) of all trials with substantially more PEs for CS^+ items (M = 0.45, SD = 0.09) compared to CS^- items (M = 0.12, SD = 0.18), $t(80) = 17.40$, $p < .001$, $d = 2.41$ (see Figure 3F). Similarly to Study I, PEs decreased as the task progressed, $r(58) = -0.48$, $p < .001$.

2.3.2.2 General memory performance

As expected, Participants on average were moderately surprised by the recognition test (see Supplemental Material). Again, participants performed very well in the recognition task, as indicated by significantly higher hit rates (M = 0.43, SD = 0.16) than category-based false alarm rates (M = 0.26, SD = 0.10), $t(80) = -7.77$, $p < .001$, $d = 1.25$ (see Figure 3G). Importantly, the category-based false alarm rate for the CS+ items (M = 0.27, SD = 0.13) was comparable to the false alarm rate for items from the CS^- category (M = 0.26, SD = 0.14), $t(80) = 0.69$, p = .494, d = 0.10. As expected, the average hit rate for items from the CS^+ category (M = 0.51, SD = 0.22) was significantly higher than for items from the CS^- category (M = 0.35, SD = 0.17), $t(80) = 6.84$, $p < .001$, d = 0.81, in line with the findings of Study I. The sensitivity measure d' was also significantly higher for CS^+ items (M = 1.38, SD = 0.67) than for CS^- items (M = 1.04, SD = 0.51; $t(80) = 5.24$, $p < .001$, d = 0.58).

Again, participants rated their level of surprise related to the recognition test on a scale ranging from 1 (not surprised at all) to 5 (very surprised). On average, participants were moderately surprised (M = 3.17, SD = 1.14). Eight participants chose the 'not surprised at all' option. Because excluding them did not affect the results, they were included in all analyses.

2.3.2.3 Modelling recognition performance at item level

To further elucidate whether PEs enhance memory for preceding and subsequent stimuli, we computed the same GLMMs as in Study I to predict the binary recognition of an item.

We started with a minimal model, in which we tested whether uPEs that followed a CS contribute to subsequent recognition of this CS item. Therefore, we treated the binary subsequent PE as the sole independent variable to predict the binary recognition of an item. Again, this analysis revealed that memory was enhanced for trials in which a PE occurred (z = 4.58, $p < .001$, β = 1.02; see Figure 3H). To rule out that those PE effects were only due to the explicit shock prediction, we also computed a model that included the explicit shock prediction and the binary subsequent PE as independent variables to predict the binary item recognition. Critically, even after controlling for the shock prediction we were able to replicate the memory enhancing effect of subsequent PEs in this model (z = 4.49, $p < .001$, β = 0.78). To investigate

the possibility that the effects of physiological arousal and the effects of PEs on memory are confounded, we added both measures of arousal (i.e., anticipatory and outcome related SCRs) to the minimal model that featured only the binary subsequent PE as the sole independent variable. This analysis showed no significant effect of anticipatory SCRs on item recognition ($z = 1.26$, $p = .209$, $\beta = 0.42$) nor of outcome-related SCRs ($z = 1.53$, $p = .127$, $\beta = 0.49$). Most importantly and in line with Study I, our results showed for subsequent PEs that recognition was significantly enhanced ($z = 4.95$, $p < .001$, $\beta = 0.87$), suggesting that subsequent PEs enhanced memory even after controlling for arousal.

Study I provided initial evidence for prospective effects of PEs, that is, memory enhancing effects of PEs for stimuli encoded after the PE. To test whether we can replicate this effect, we treated the binary previous PE as the sole independent variable to predict the binary recognition of the following item. This analysis revealed that episodic memory was indeed enhanced for items following an incorrect prediction (i.e., a PE; $z = 5.03$, $p < .001$, $\beta = 0.74$). In addition, we also computed a model that included the explicit shock prediction and the previous PE as independent variables to predict the binary item recognition. Critically, even after controlling for the shock prediction, we obtained a memory enhancing effect of previous PEs in this model ($z = 3.46$, $p < .001$, $\beta = 0.50$). Same as in Study I, we added both measures of arousal (i.e., anticipatory and outcome-related SCRs) to the minimal model that featured only the binary previous PE as the sole independent variable to rule out confounds with physiological arousal. Interestingly, this model revealed neither a significant effect of anticipatory SCR ($z = 1.57$, $p = .116$, $\beta = 0.61$) nor of outcome-related SCR on item recognition ($z = 1.19$, $p = .234$, $\beta = 0.39$). Importantly, however, we obtained a significant effect of previous PEs on item recognition in this model ($z = 4.73$, $p < .001$, $\beta = 0.69$).

Additionally, we sought to investigate whether the subsequent PE and previous PE reflect distinct mechanisms. Therefore, we added both the subsequent PE and the previous PE as independent variables to an additional model. Estimates obtained revealed that subsequent PEs, $z = 4.77$, $p < .001$, $\beta = 0.87$, and previous PEs, $z = 4.56$, $p < .001$, $\beta = 0.67$, showed a positive relationship with item recognition, dovetailing with the results of Study 1 suggesting that both contribute distinctly to memory recognition. In a follow-up model, we added anticipatory arousal and outcome-related arousal as predictors. Interestingly, neither anticipatory SCRs, $z = 1.77$, $p = .078$, $\beta = 0.67$, nor outcome-related SCRs influenced item recognition significantly, $z = 0.48$, $p = .629$, $\beta = 0.16$. Most strikingly, our analyses showed that item recognition was significantly enhanced by both, the subsequent PEs, $z = 4.65$, $p < .001$, β

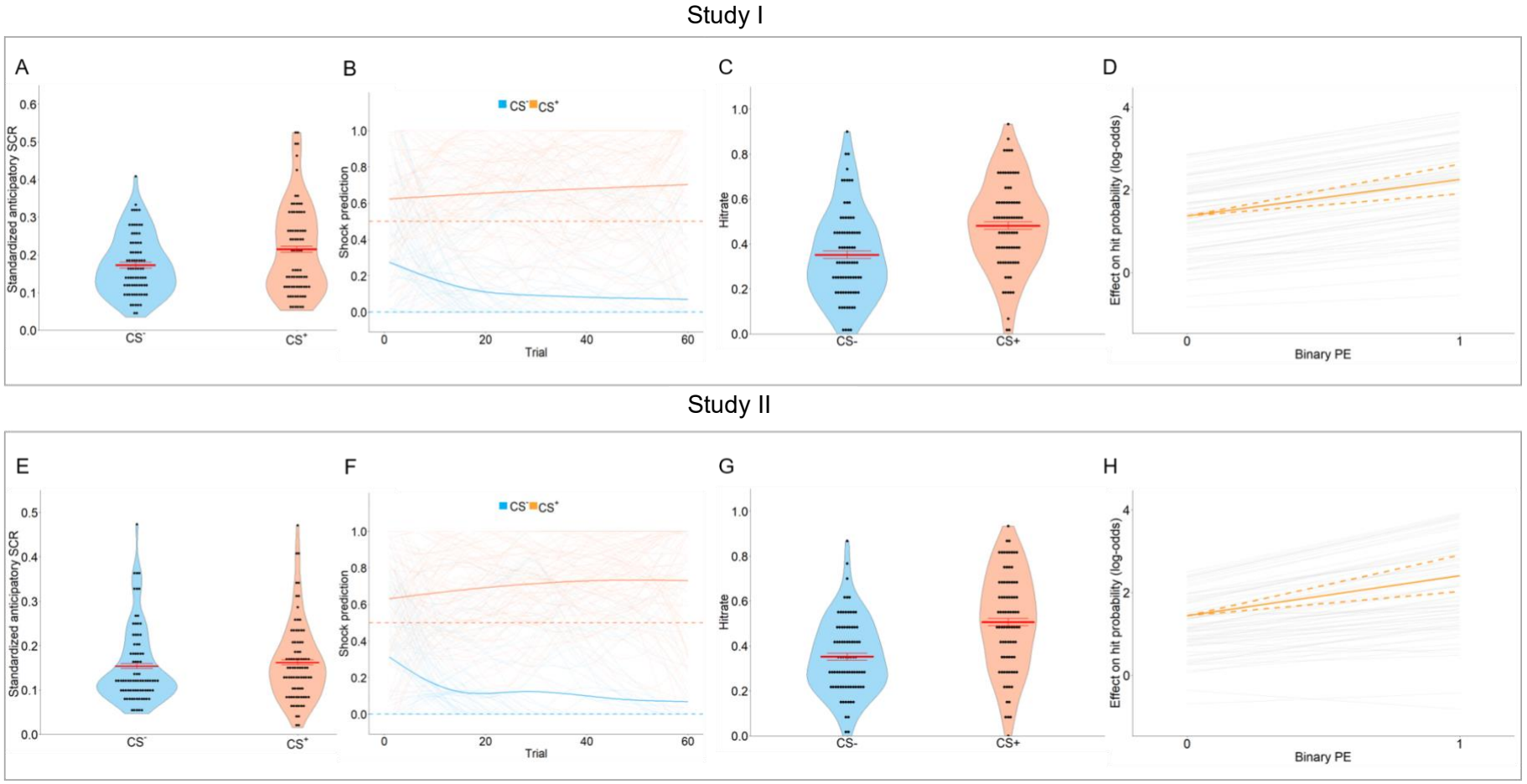
= 0.86, and previous PEs, $z = 4.38$, $p < .001$, $\beta = 0.65$, suggesting that both types of PEs seem to be distinct from the effect of arousal.

2.3.3 Conclusion

The findings of Study II replicate the previously reported enhancing effects of (unsigned) PEs associated with aversive events on subsequent memory for the stimulus encountered before the PE. Moreover, our findings further replicate the prospective memory enhancement induced by PEs. In contrast to Study I, which suggested that the prospective PE effects on memory may be due to arousal, the findings of Study II show that both the retrospective PE effects and the prospective PE effects on memory were independent of physiological arousal, as measured by SCR. The failure to replicate the effects of arousal on memory that we observed in Study I suggests that arousal effects on memory might be less robust than those of PEs on memory.

Figure 3

Results of Study I and Study II



Note. Anticipatory skin conductance responses and hitrates were always significantly higher for items from the CS^+ category compared to items from the CS^- category in both Study I (A,C) and Study II (E,G), confirming that the fear conditioning procedure was successful. Additionally, shock contingencies were learned well in Study I (B) and II (*F*; thick lines) approaching the underlying shock contingencies (dotted lines). GLMMs revealed that subsequent PEs significantly enhanced recognition memory for predictive stimuli in both studies (D,G). Black dots show individual data. Thick red bar represents group mean, while thin red bars show ±1 standard error of the mean.

3 Is the PE-induced memory enhancement time dependent?

This chapter was published in modified form in: 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

3.1 Background

In line with previous studies (Kalbe & Schwabe, 2020; Kalbe & Schwabe, 2022b; for a review, see Rouhani et al., 2023), the results of Studies I and II show consistently that PEs associated with aversive events enhance subsequent memory, beyond arousal effects on memory. Notably, the results of Studies I and II indicated further that PE effects may extend to stimuli following the PE, suggesting a brief time window of enhanced processing induced by the PE that lasts for at least a few seconds.

Expanding on these observations, Study III focused on the temporal dynamics underlying PE-related memory effects. Specifically, we aimed to elucidate whether the retrospective PE effect on memory depends on the temporal proximity between the predictive stimulus and the unexpected outcome, i.e. whether maintaining the representation of the predictive cue over time is necessary for the PE effect to occur. From a theoretical perspective, learning can be based on either temporal contiguity or contingency (Schultz, 2006). If PE-related memory effects rely on temporal contiguity, this would imply that memory for the predictive cue (either preceding or following the PE) is enhanced primarily due to its mere temporal proximity to the surprising outcome. To test this, we (i) systematically manipulated the CS-outcome delay, varying it between 0 and 10 seconds. Additionally, we (ii) examined whether PE effects on memory depend on the time interval between encoding and testing in order to assess whether the PE effects emerge during memory encoding or consolidation. Thus, participants' recognition memory was tested either immediately after encoding or 24 hours later, as in Studies I and II (and all previous studies on PEs related to aversive events and memory (see Kalbe & Schwabe, 2020, 2022b)). Additionally, we sought to examine the impact of positive and negative PEs on memory, given that they exert distinct effects on memory formation (Kalbe & Schwabe, 2022b; Rouhani et al., 2023; Rouhani & Niv, 2021). To this end, we modified our experimental paradigm in two ways to achieve an adequate distribution of both positive and negative PEs: First, we presented three CSs, with one CS^- and two CS^+ with different contingencies. Second, participants were asked to rate their shock expectancy on a continuous scale from 0 to 100 (see Pine et al., 2018).

3.2 Methods

3.2.1 Participants

One hundred twenty-three healthy volunteers participated in this study (age: $M = 25.95$ years, $SD = 4.33$ years, range = 18–35 years). Exclusion criteria were identical to those in Studies I and II. Five participants were excluded from the analyses due to technical failure during the experiment or because they did not return for the second experimental day, resulting in a final sample of $n = 118$. Importantly, none of the participants had participated in Study I or II. The target sample size was based on previous findings of signed aversive PE effects on episodic memory formation (Kalbe & Schwabe, 2022b). Because we modified the experimental design, in particular by adding a variable time interval between CS and outcome and by adding the between-factor retention interval (immediate vs. 24 hr delay), we doubled the reported sample size of Kalbe and Schwabe (2022b). Thus, we expected a sample of 120 participants to be sufficient to detect a power of at least 0.90. In line with these assumptions, a post hoc power simulation using the R-package simR (Green & MacLeod, 2016) for the observed effects of the subsequent sPE and our final sample size of 118 participants yielded a power of 0.92 based on 1000 simulations. All participants provided written informed consent before participation and received a moderate monetary reimbursement (up to 50 €) at the end of the study. The study was approved by the ethics committee of the Faculty of Psychology and Human Movement Science at the Universität Hamburg and carried out in line with the Declaration of Helsinki.

3.2.2 Materials

In Study III, we used stimuli from Kalbe and Schwabe (2022b) but added more stimuli due to the increased trial number. Stimuli were taken from existing image databases, that is, Bank of Standardized Stimuli (Brodeur et al., 2010, 2014), McGill Calibrated Color Image Database (Olmos & Kingdom, 2004), SUN database (Xiao et al., 2010), Konklab (Konkle et al., 2010), and open-online sources. In total, the stimulus set consisted of 810 pictures of vehicles, tools, and clothes, isolated on white background. All stimuli were assumed to be emotionally neutral and represented a unique exemplar of its category. From this pool, 120 pictures of vehicles, 120 pictures of tools, and 120 pictures of clothes were randomly drawn and used during encoding on the first experimental day. From the remaining 450 pictures, 180 randomly chosen pictures (60 pictures per category) served as lures for the surprise recognition test on the second experimental day. The allocation of images as encoding items or lures was randomized across participants and thus unique per participant. In addition, the order in which individual items were presented was randomized across participants.

3.2.3 Procedure

The study consisted of two parts, with an encoding session on the first experimental day and a recognition test on the second day (see Figure 4A). Depending on the experimental group, the recognition test took place either immediately after the encoding session or 22–26 hr later. Participants were pseudorandomly assigned to the two groups (immediate group: 20 men, 36 women, M_{age}: 25.93 (SD = 4.13); 24-hr-delay group: 23 men, 37 women, M_{age}: 25.97 (SD = 4.54)). Upon arrival on the first experimental day, participants provided written informed consent and received written instructions that they were going to see a series of photographs of vehicles, tools, and clothes and that some of them might be followed by a brief electric shock. They were then instructed to predict how likely a shock would be to follow the current picture (see Figure 4A). Therefore, they were requested to adjust the slider on the screen to a value corresponding to their prediction of the shock probability (ranging from 0% to 100%). Importantly, participants were neither told about the shock contingencies, nor that their memory would be tested later on. They were informed that their predictions would not affect the probability that a shock would occur, but that they should aim at improving their predictions over the task. Unbeknownst to the participants, the probabilities of a shock were linked to the image categories. One category served as CS^{a+} (67% shock probability), one as CS^{b+} (33% shock probability), and one as CS^{-} (0% shock probability). The assignment of image categories (i.e., vehicles, tools, and clothes) to the CS categories (i.e., CS^{a+}, CS^{b+} and CS^{-}) was counterbalanced across participants and groups.

To measure SCRs as indicator of physiological arousal through the incidental encoding-fear learning task, we placed disposable, pregelled snap-electrodes on the thenar and hypothenar eminence of the left hand (see Kalbe & Schwabe, 2022b). Skin conductance was measured using the MP-160 BIOPAC system (BIOPAC systems, Goleta, California, United States). For electrical stimulation, we used the STM-100C module connected to the MP-160. A stimulation electrode was placed on the back of participants' right lower leg, approximately 20 cm above the ankle. Before the learning task, stimulation intensity was adjusted individually to be unpleasant but not painful as described in Study I.

The encoding session on the first experimental day consisted of four blocks with 90 trials each. In each block, 30 pictures of vehicles, 30 pictures of tools, and 30 pictures of clothes were presented in a pseudorandomized order, so that no more than three pictures of the same category appeared in a row. On each trial, a picture was shown in the center of a computer screen for 4.5 s, during which participants were asked to make their prediction about the probability of an electric shock (Figure 4A). Therefore, a slider was presented beneath each

image which could be adjusted to any integer value between 0% and 100% using the computer mouse. After stimulus offset, a black dot was presented centrally on the screen which coterminated with the 200 ms-outcome (shock vs. no-shock). Critically, the duration of the dot's presentation ranged randomly between 0 and 10 s per trial to vary the critical CS-outcome delay. Each trial was followed by a black fixation cross centered on gray background for 6.5 ± 1.5 s, which enabled us to measure the relatively slow SCRs elicited by the pictures and the shocks. Between blocks there were short breaks (1–2 min) during which participants had the chance to recalibrate the shock intensity, if required. Each encoding block lasted approximately 25 min, resulting in a total duration of 100 min for the entire incidental encoding-fear learning session.

The surprise recognition test took place either immediately after the encoding session or on the next day, 22–26 hr later, depending on the experimental group. Same as in Study I and II, participants completed a short questionnaire to assess whether they anticipated a memory test and then rated how surprised they were about the recognition test on a scale from 1 (not surprised at all) to 5 (very surprised). In the recognition test, participants saw all pictures they had seen during the encoding session (120 pictures of vehicles, 120 pictures of tools and 120 pictures of clothes) as well as 180 "new" pictures (60 pictures of vehicles, 60 pictures of tools, and 60 pictures of clothes) that had not been presented before in a randomized order. Each trial started with a central white fixation cross on a white background for 1.5 ± 0.5 s, followed by an "old" or "new" picture presented centrally on the computer screen for 6 s. For each item, participants were instructed to indicate whether the currently presented picture was definitely old, maybe old, maybe new, or definitely new by pressing the "1," "2," "3," or "4" button on the keyboard, respectively.

3.2.4 Data analysis

For each trial, we derived uPEs which were calculated as the absolute value of the difference between participants' continuous explicit shock expectancy ratings (ranging from 0, corresponding to full confidence that no shock would occur, to 1, corresponding to full confidence that a shock would occur) and the actual binary outcome of the trial (coded 0 when no shock occurred and 1 when a shock occurred in the current trial). The resulting uPE is, therefore, ranging between 0 and 1. Because the modified paradigm of Study III allowed us to measure continuous PEs we derived also an sPE. We focused our analyses on sPE effects because these represent a more accurate measure of the PE and allow a distinction between positive and negative PEs. Analyses using the uPE as predictor are presented in the Supplemental Material.

The sPE was calculated as the relative value of the difference between the binary outcome of a trial (1 for shock and 0 for nonshock) and the explicit shock prediction and could take any integer value between −1 and 1. Importantly, the sign of the sPE contained information about the outcome's value: A negative sPE (sPE < 0) could only occur in unshocked trials corresponding to unexpected shock omissions, whereas positive sPEs (sPE > 0) could only occur in shocked trials corresponding to unexpected shock occurrence.

Again, SCRs were analyzed using continuous decomposition analysis in Ledalab Version 3.4.9 (Benedek & Kaernbach, 2010) and calculated in the same way as in Studies I and II. Deviations were due to altered response windows as a consequence of the variable CS-outcome delay: For the anticipatory SCR, the response window was set from 0.5 after stimulus onset until the onset of the outcome (shock vs. no-shock) and could thus vary depending on the CS-outcome delay. Outcome-related SCR was analysed between 0.5 s and 4.5 s after outcome onset, in line with Studies I and II. Resulting estimates of the average phasic driver within each response window were returned in μS. Post-hoc comparisons of ANOVAs were always Bonferroni-corrected.

Again, we fitted GLMMs with a logit link function using the lme4 R package (Bates et al., 2015) and treated subjects as random effects for both the intercept and all slopes of the fixed effects included in the model (Barr et al., 2013). The recognition of an individual item was treated as the binary dependent variable, coded "0" for misses and "1" for confident hits. We fitted models using different sets of independent variables, including subsequent PEs, previous PEs, anticipatory and outcome-related arousal, the explicit shock prediction, the retention interval, and the CS-outcome delay. In line with Studies I and II, we also distinguished between subsequent and previous PEs with the former referring to PE effects on the recognition of the preceding stimulus and the latter referring to PE effects on the recognition of the following stimulus. If not indicated otherwise, all analyses were collapsed across both retention intervals.

3.2.5 Transparency and openness

The materials, data, and R analysis scripts are publicly available on the Research Data Management System of University of Hamburg and can be accessed at https://www.fdr.uni-hamburg.de/record/14147 (Loock et al., 2024). This study was not preregistered.

3.3 Results

3.3.1 Successful fear conditioning

SCR data confirmed the expected fear learning process. Specifically, anticipatory SCR differed significantly between conditioning categories, $F(2, 234) = 4.18$, $p = .016$, partial $\eta^2 =$

0.002 (Figure 4D). Post hoc paired *t*-tests showed that participants showed higher anticipatory SCRs to CS^{a+} items ($M = 0.09$, $SD = 0.01$) compared to CS^{-} items ($M = 0.08$, $SD = 0.02$), $t(117) = 2.53$, $p = .012$, $d = 0.11$. Anticipatory SCRs did not differ significantly between CS^{b+} items ($M = 0.08$, $SD = 0.01$) and CS^{-} items, $t(115) = 0.79$, $p = .434$, $d = 0.04$. Outcome-related SCRs were significantly higher for shocked trials ($M = 0.19$, $SD = 0.10$) compared to unshocked trials ($M = 0.05$, $SD = 0.10$), $t(115) = 10.83$, $p < .001$, $d = 1.40$.

Explicit shock ratings further showed that participants learned the shock contingencies over the task very well (see Figure 4C). Participants had a significantly higher shock expectancy for CS^{a+} ($M = 0.71$, $SD = 0.17$) compared to CS^{b+} ($M = 0.47$, $SD = 0.12$), $t(117) = 13.27$, $p < .001$, $d = 1.52$, and for CS^{b+} compared to CS^{-} ($M = 0.13$, $SD = 0.17$); $t(117) = 18.16$, $p < .001$, $d = 2.11$. In addition, PEs were equally distributed around zero (Figure 4B) suggesting a sufficient number of positive and negative PEs that could be analyzed.

3.3.2 General memory performance

Again, participants were moderately surprised by the recognition test (see Supplemental Material). Overall, participants performed very well in the recognition task, as indicated by significantly higher hit rates ($M = 0.46$, $SD = 0.21$) than category-based false alarm rates ($M = 0.21$, $SD = 0.16$), $t(353) = 17.02$, $p < .001$, $d = 1.29$.

Importantly, while false alarm rates were comparable between conditioning categories, $F(2, 234) = 2.34$, $p = .098$, partial $\eta^2 = 0.004$, hit rates differed significantly between conditioning categories, $F(2, 234) = 9.61$, $p < .001$, partial $\eta^2 = 0.016$ (see Figure 4E) suggesting that memory but not the response bias differed between categories. Post hoc paired *t*-tests revealed that the average hit rate for CS^{a+} items ($M = 0.49$, $SD = 0.11$) was significantly higher than for CS^{b+} items ($M = 0.44$, $SD = 0.10$), $t(117) = 3.97$, $p < .001$, $d = 0.24$, and CS^{-} items ($M = 0.43$, $SD = 0.12$; $t(117) = 3.75$, $p < .001$, $d = 0.29$). The average hit rate for CS^{b+} items did not differ from the hit rate for CS^{-} items, $t(117) = 0.52$, $p = .605$, $d = 0.04$. When using the signal detection theory-based parameter *d'* instead of the hit rate, recognition memory was higher for both CS^{a+} items ($M = 1.47$, $SD = 0.66$) and CS^{b+} items ($M = 1.43$, $SD = 0.68$) compared to CS^{-} items ($M = 1.27$, $SD = 0.63$; vs. CS^{a+}: $t(112) = 3.34$, $p = .001$, $d = 0.30$; vs. CS^{b+}: for $t(112) = 2.63$, $p = .010$, $d = 0.24$; main effect CS category: $F(2,224) = 6.43$, $p = .002$, partial $\eta^2 = 0.016$).

Moreover, when taking the retention interval into account, recognition memory performance differed significantly between the immediate and 24 hr delayed groups, as expected. Participants who underwent the recognition test immediately after the encoding session (hit rate: $M = 0.53$, $SD = 0.13$; *d'*: $M = 1.53$, $SD = 0.71$) had a significantly better

recognition memory than participants who performed the recognition test about 24 hr after encoding (hit rate: $M = 0.38$, $SD = 0.14$, $t(348.6) = 7.05$, $p < .001$, $d = 0.74$; d': $M = 1.24$, $SD = 0.57$, $t(329.11) = 4.24$, $p < .001$, $d = 0.43$) reflecting the well-known decline in memory over time. There was no interaction between CS type and retention interval, hit rate: $F(2, 232) = 0.19$, $p = .830$, partial $\eta^2 = 0.000$; d': $F(2, 222) = 1.43$, $p = .241$, partial $\eta^2 = 0.004$, suggesting that the differential memory performance for stimuli from the three CS categories did not differ between the immediate and 24-hr-delayed groups.

Again, participants rated their level of surprise related to the recognition test on a scale ranging from 1 (not surprised at all) to 5 (very surprised). On average, participants were moderately surprised ($M = 2.88$, $SD = 1.10$). Fifteen participants chose the 'not surprised at all' option. Because excluding them did not affect the results, they were included in all analyses.

3.3.3 Modelling recognition performance at item level

To elucidate the mechanisms of episodic memory formation, we again fitted GLMMs with recognition of an item as the binary dependent variable and added certain independent predictors in a step-wise manner, similarly to Studies I and II.

We started with a first minimal model, in which we tested whether trial-wise subsequent sPEs contribute to later recognition. Therefore, we treated the sPE (ranging from −1 to 1) following on a CS as the sole independent variable to predict the binary recognition of this CS item. Estimates obtained revealed that sPEs ($z = 2.53$, $p = .012$, $\beta = 0.08$) showed a positive relationship with item recognition. To rule out that the PE effects were confounded with the shock prediction, we also computed a model where we added the explicit shock prediction as a predictor to the previous model. When controlling for the explicit shock prediction, the memory enhancing effect of the subsequent sPEs remained significant ($z = 4.20$, $p < .001$, $\beta = 0.15$). In a follow-up model, we added anticipatory arousal and outcome-related arousal as predictors. While anticipatory SCRs were associated with decreased memory ($z = -3.95$, $p < .001$, $\beta = -0.42$), outcome-related SCRs did not influence item recognition significantly ($z = 1.05$, $p = .295$, $\beta = 0.12$). After controlling for arousal effects on memory, the sPE effect did not reach statistical significance anymore ($z = 1.44$, $p = .150$, $\beta = 0.05$).

To examine whether sPE effects on memory are dependent on the retention interval and the CS-outcome delay, we included the retention interval and the CS-outcome delay as well as their interaction as predictors in an additional set of models.

First, we set up a model that treated the subsequent sPE and the retention interval and their interaction as independent variables to predict the binary recognition of an item. When controlling for the retention interval, we obtained a significant effect of subsequent sPEs on

memory ($z = 2.04$, $p = .041$, $\beta = 0.09$) and a nonsignificant Retention Interval × Subsequent sPE interaction ($z = -0.32$, $p = .753$, $\beta = -0.02$), suggesting that these "retrospective" sPE effects are not dependent on the interval between encoding and test.

In a next step, we set up a model that treated the subsequent sPE, the CS-outcome delay and their interaction as independent variables to predict the binary recognition of an item. This revealed that memory was neither influenced by the CS-outcome delay ($z = -1.57$, $p = .117$, $\beta = -0.01$) nor by the CS-Outcome Delay × Subsequent sPE interaction ($z = 1.56$, $p = .120$, $\beta = 0.01$). These findings suggest that the CS-outcome delay does not influence memory and does not modulate the subsequent sPE effects on memory (see Figure 4F).

In an additional model, we treated the CS-outcome delay, the retention interval, the sPE and their interaction as independent variables. Estimates obtained showed no significant CS-Outcome Delay × Retention Interval × sPE interaction ($z = 0.70$, $p = .481$, $\beta = 0.01$).

Next, we performed additional models in which we treated the previous sPE as the sole independent variable to predict the binary recognition of the following item. This revealed no significant effect of previous sPEs on memory for items following the PE ($z = 0.21$, $p = .835$, $\beta = 0.00$). Again, we added anticipatory and outcome-related SCRs to the former model to investigate confounds with physiological arousal. This revealed a significant impairing effect of anticipatory SCRs on item recognition ($z = -4.54$, $p < .001$, $\beta = -0.47$), whereas we obtained no effect of outcome-related SCRs on memory formation ($z = 1.67$, $p = .090$, $\beta = 0.16$). The previous sPE effect remained nonsignificant ($z = 0.57$, $p = .570$, $\beta = 0.01$). To rule out that the previous sPE effects were confounded with the shock prediction, we also computed a model where we added the explicit shock prediction and the previous sPE as predictors. When controlling for the explicit shock prediction, the effect of the previous sPE remained nonsignificant ($z = 0.58$, $p = .563$, $\beta = 0.01$).

In a next step, we tested whether the effect of sPEs on the memory for items following the sPE are dependent on the retention interval. Estimates obtained revealed that neither the (previous) sPE influence recognition of the following item significantly ($z = 0.26$, $p = .795$, $\beta = 0.01$) nor the Retention Interval × Previous sPE interaction ($z = -0.34$, $p = .737$, $\beta = -0.01$). Notably, there was a significant effect of the retention interval on CS recognition ($z = 4.35$, $p < .001$, $\beta = 0.33$).

We added both previous and subsequent sPEs as independent variables to an additional model to investigate whether subsequent PEs and previous PEs reflect distinct mechanisms. Estimates showed that only subsequent sPEs, $z = 2.49$, $p = .013$, $\beta = 0.08$, showed a positive relationship with item recognition while the association of previous sPEs and item recognition

remained non-significant, z = -0.09, p = .931, β = -0.00. In a follow-up model, we added anticipatory arousal and outcome-related arousal as predictors. In this model, neither outcome-related arousal, z = 1.04, p = .301 , β = 0.11, nor subsequent sPEs, z = 1.45, p = .149 , β = 0.05, nor previous sPEs influenced item recognition significantly, z = 0.43, p = .668 , β = 0.01, while item recognition was significantly impaired by anticipatory SCRs, z = -3.90, p < .001, β = -0.41.

3.3.4 Effects of uPEs

We started with a first minimal model, in which we tested whether uPEs contribute to later recognition. Therefore, we treated the uPE following on a CS as the sole independent variable to predict the binary recognition of this CS item, irrespective of the CS-outcome interval. This analysis showed no effect of uPE on episodic memory, z = 1.19, p = .232, β = 0.09. To investigate the possibility that the effects of physiological arousal and the effects of PEs on memory might be confounded, we added both measures of arousal (i.e., anticipatory and outcome related SCRs) to the minimal model that featured only the subsequent uPE as the sole independent variable. Interestingly, this revealed a significant impairing effect of anticipatory SCRs on item recognition, z = -4.33, p < .001, β = -0.44, while we did not find an effect of outcome-related SCRs on memory formation, z = 0.97, p = .335, β = 0.09. The uPE effect on item recognition remained non-significant, z = 1.26, p = .208, β = 0.09. To rule out that the subsequent uPE-effects were confounded with the shock prediction, we also computed a model where we added the explicit shock prediction and the subsequent uPE as predictors. When controlling for the explicit shock prediction, the effect of the subsequent uPE remained non-significant, z = 0.96, p = .335, β = 0.06.

To examine whether uPE effects on memory are dependent on the retention interval and the CS-outcome interval, we included the retention interval and the CS-outcome delay as predictors in an additional set of models.

First, we set up a model that treated the subsequent uPE and the retention interval as independent variables to predict the binary recognition of an item. When controlling for the retention interval, we obtained no significant effect of subsequent uPEs on memory, z = 1.16, p = .245 , β = 0.09, a non-significant *retention interval* × *subsequent uPE* interaction, z = 0.43, p = .664, β = 0.03, suggesting that these 'retrospective' PE effects are not dependent on the interval between encoding and test, but a significant effect of the retention interval on memory, z = 4.36, p < .001, β = 0.32.

In a next step, we set up a model that treated the subsequent uPE and the CS-outcome delay including their interaction as independent variables to predict the binary recognition of

an item. This revealed that episodic memory was impaired by the CS-outcome delay, z = -2.16, p = .031, β = -0.01, suggesting that increasing intervals led to lower recognition of the preceding item. The subsequent uPE-effects on memory remained non-significant when controlling for the CS-outcome-delay, z = 0.27, p = .788, β = 0.02, and there was no significant *CS-outcome delay×subsequent uPE* interaction, z = 1.27, p = .206, β = 0.01. In an additional model we treated the CS-outcome interval, the retention interval, the uPE and their interaction as independent variables. Estimates obtained showed a non-significant *CS-outcome delay×retention interval×uPE* interaction, z = -.13, p = .900, β = -0.00.

As Studies I and II suggested also prospective effects of PEs, we performed additional models in which we treated the previous uPE as the sole independent variable to predict the binary recognition of the following item. This revealed that episodic memory was indeed enhanced for items following a PE, z = 2.68, p = .007, β = 0.10. To rule out that the previous uPE-effects were confounded with the shock prediction, we also computed a model where we added the explicit shock prediction and the previous uPE as predictors. When controlling for the explicit shock prediction, the effect of the previous uPE remained significant, z = 2.72, p = .007, β = 0.09. Again, we added anticipatory and outcome-related SCRs to the former model featuring the previous uPE to investigate confounds with physiological arousal. Again, this revealed a significant impairing effect of anticipatory SCRs on item recognition, z = -4.41, p < .001, β = -0.46, while we did not find an effect of outcome-related SCRs on memory formation, z = 1.63, p = .103, β = 0.15. Importantly, however, even after controlling for arousal, our results showed a memory enhancing effect associated with uPE on the recognition of the following item, z = 2.77, p = .006, β = 0.10.

In a next step, we tested whether the previously observed effect of PEs on the memory for items following the PE would be dependent on the retention interval. Estimates obtained revealed that the (previous) uPE was associated with a memory enhancement for the following item, z = 2.67, p = .008, β = 0.09, even when controlling for the retention interval, suggesting that these 'prospective' PE effects are are already seen shortly after encoding and remain after 24 hours.

To investigate whether the subsequent uPE and previous uPE reflect distinct mechanisms, we added both uPEs as independent variables to an additional model. Estimates obtained revealed that only previous uPEs, z = 2.26, p = .024, β = 0.07, showed a positive relationship with item recognition while the association of subsequent uPEs and item recognition remained non-significant, z = 1.17, p = .241, β = 0.09. In a follow-up model, we added anticipatory arousal and outcome-related arousal as predictors. In this model, neither outcome-related arousal, z =

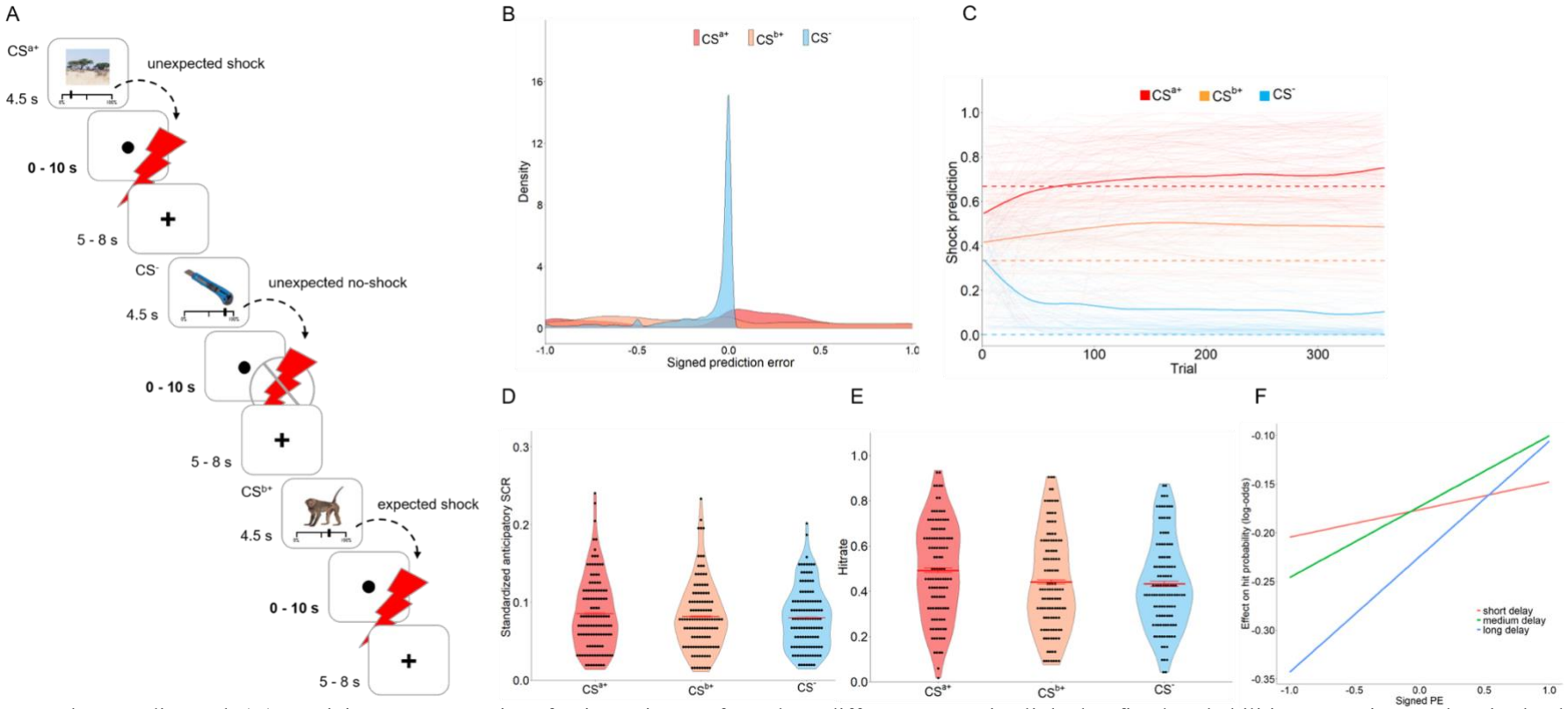
0.96, p = .340, β = 0.09, nor subsequent uPEs influenced item recognition significantly, z = 1.23, p = .219, β = 0.09. while item recognition was significantly impaired by anticipatory SCRs, z = -4.25, p < .001, β = -0.43, and significantly enhanced by previous uPEs, z = 2.24, p = .025, β = 0.07.

3.4 Conclusion

The findings of Study III replicate previously reported enhancing effects of (signed) PEs associated with aversive events on memory for the stimulus encountered before the PE. Our findings, however, show no prospective (signed) PE effects on memory, whereas such prospective effects are observed for uPE, in line with Studies I and II. Interestingly, we found that the retrospective sPE effects on episodic memory seem to be (a) independent of the time interval between the stimulus and the PE and (b) emerge already when memory is tested shortly after encoding, suggesting that they are not consolidation dependent. Thus, we obtain first evidence for retrospective PE effects to be time resistant and that they rather emerge during encoding than during consolidation processes. Notably, the results of Study III showed that the sPE effects were at least partly related to arousal because these effects disappeared when we controlled for arousal. The latter might be due to the extended CS-outcome interval.

Figure 4

Experimental procedure and results of Study III



Note. In the encoding task (A), participants saw a series of unique pictures from three different categories linked to fixed probabilities to receive an electric shock (CS^{a+}=67%, CS^{b+}=33%, and CS^{-}=0%). On each trial, participants indicated their shock expectation on a continuous scale from 0 to 100 %. The delay with which the outcome (shock vs. no-shock) occurred after stimulus-offset was varied between 0 and 10 sec. (B) PEs were equally distributed around zero. (C) Participants' mean shock expectancy ratings (thick lines) approached the true shock probabilities (dotted lines) relatively fast. (D, E) Mean standardized anticipatory SCR and hitrates confirmed successful fear conditioning, as reflected in significantly elevated SCR and increased hitrates of CS^{a+} compared with CS^{-} items. Black dots show data from individual participants. Thick red bar represents group mean, thin red bars show ±1 standard error of the mean. (*F*) The CS-outcome delay did not interact significantly with the subsequent sPE-effect on item memory suggesting that PEs seem to be independent of the time between stimulus and outcome. Pictures taken from "Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos" by Brodeur et al. (2014) and from "SUN database: Large-scale scene recognition from abbey to zoo" by Xiao et al. (2010). CC BY 4.0.

4 Is the PE-related memory enhancement specific to the predictive stimulus?

This chapter was published in modified form in: 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

4.1 Background

Studies I-III consistently showed that PEs following inherently neutral events enhance the subsequent memory of these events. However, a central open question is whether this PE-related memory enhancement is specific to the (predictive) stimulus that evoked the PE or whether the PE trigger a transient window of enhanced memory encoding for all events, including stimuli that are uninformative, that occur in this time window.

The findings of Study III revealed that the PE-related memory enhancement is still observed even with an CS-outcome delay that extends up to 10 s. There might be two possible explanations for this observation: (i) PEs might retroactively and exclusively enhance the memory of a predictive stimulus presented up to 10 seconds earlier or (2) PEs might induce a transient window of enhanced encoding that extends to all stimuli occurring within that time window including those that are not predictive or informative. Disentangling these accounts is critical, especially in light of theories of learning that distinguish between temporal contiguity and contingency (Schultz, 2006). While temporal contiguity refers to the temporal proximity of two events, contingency captures the predictive relationship between them. If PE effects on memory are primarily driven by contingency, only predictive stimuli, and not merely temporally adjacent ones, i.e., uninformative stimuli, should benefit from enhanced encoding. While Studies I and II found prospective PE effects, i.e., enhanced encoding for stimuli presented after the PE indicating a window of enhanced processing, these effects were not observed in Study III. This inconsistency raises the question whether such effects are truly driven by a broad encoding window or by a selective mechanism tied to stimulus relevance.

In Study IV, we aimed to determine whether PEs are exclusively linked to the predictive stimulus or whether PEs trigger a window in which memory formation is (retrospectively) strengthened for all stimuli, including entirely uninformative ones. To this end, we used a modified paradigm, in which we presented in some of the encoding blocks entirely uninformative stimuli between the predictive stimulus and the outcome (i.e., potential PE). If the PE induces a transient window of enhanced encoding, then memory should also be enhanced for uninformative stimuli presented shortly after the predictive stimulus. In contrast, if the PE-

related memory enhancement is contingency-dependent, then memory should not be enhanced for the uninformative stimulus presented between the predictive stimulus and the PE. Moreover, because previous research (Kalbe & Schwabe, 2020, 2022b; and our Studies I and II) suggested that the PE effects might be at least partly independent of physiological arousal but included only SCR as the only arousal measure, we included here, in addition to SCR, respiratory responses and heart rate as further measures of arousal.

4.2 Methods

4.2.1 Participants

Eighty-two healthy volunteers participated in this study (age: M = 25.45 years, SD = 3.98 years, range = 18–35 years). Exclusion criteria were identical to those in Studies I-III. Four participants were excluded from the analyses due to technical failure during the experiment or because they did not return for the second experimental day, resulting in a final sample of n = 78. Importantly, none of the participants had participated in Studies I-III. The target sample size was based on an a priori power calculation using G*Power (3.1.9.6; Faul et al., 2009). Based on previous research by Kalbe and Schwabe (2020), we assumed d_z = .39 as a point estimate for the expected effects. A two-tailed paired t-test with α = .05 required at least 72 participants to achieve a statistical power of 0.90. In line with this calculation, a post hoc power simulation using the R-package simR (Green & MacLeod, 2016) for the observed effects of subsequent PEs and our final sample size of 78 participants yielded a power of 0.84 based on 1000 simulations. All participants provided written informed consent before participation and received a moderate monetary reimbursement (up to 50 €) at the end of the study. The study was approved by the ethics committee of the Faculty of Psychology and Human Movement Science at the University of Hamburg and carried out in line with the Declaration of Helsinki.

4.2.2 Materials

We used the same stimulus set as in Study III, that is, 810 pictures of vehicles, tools, and clothes, isolated on white background, but added 270 pictures of outdoor scenes, resulting in a total set of 1,080 stimuli. Outdoor scenes were taken from image databases, that is, Bank of Standardized Stimuli (Brodeur et al., 2010, 2014), McGill Calibrated Color Image Database (Olmos & Kingdom, 2004), SUN database (Xiao et al., 2010), Konklab (Konkle et al., 2010), and open-online sources. Again, 120 randomly drawn pictures from each of the categories vehicles, tools, and clothes were used during encoding on the first experimental day. In addition, 180 randomly drawn pictures of scenes were shown during these encoding sessions. From the remaining 540 pictures, 270 randomly chosen pictures, that is, 60 pictures of tools, 60 pictures

of vehicles, 60 pictures of clothes, and 90 pictures of scenes, served as lures for the surprise recognition test on the second experimental day. All images were assumed to be emotionally neutral. The allocation of images to CS categories or lures as well as the order in which individual items were presented was randomized across participants and unique per participant.

4.2.3 Procedure

The procedure was largely comparable to Study III but contained some important differences, in particular the inclusion of encoding blocks in which an uninformative stimulus (UI) was presented between CS and outcome (see Figure 5A). Furthermore, the CS-outcome interval was kept constant in this study. Same as in Studies I and II, this experiment consisted of two sessions, with an encoding session on the first experimental day and a recognition test on the following day, about 22–26 hr after encoding.

Upon arrival on the first experimental day, participants provided written informed consent and received written instructions that they were going to see a series of photographs of vehicles, tools, clothes, and scenes and that some of them might be followed by a brief electric shock. Again, they were then instructed to predict how likely a shock would be to follow a picture by adjusting a slider on the screen to a value corresponding to their prediction of the shock probability (ranging from 0% to 100%). Importantly, participants were neither told about the shock contingencies, nor that their memory would be tested later on. They were informed that their predictions would not affect the probability that a shock would occur, but that they should aim at improving their predictions over the task. In line with Study III and unbeknownst to the participants, the probabilities of a shock were linked to the image categories. One category served as CS^{a+} (67% shock probability), one as CS^{b+} (33% shock probability), and one as CS^{-} (0% shock probability). The assignment of image categories (i.e., vehicles, tools, and clothes) to the CS categories (i.e., CS^{a+} ,CS^{b+} and CS^{-}) was counterbalanced across participants and groups. We also added a new, stimulus category (UI), which was uninformative with respect to the occurrence of an electric shock. Scene images were always used as UI to make sure that these sufficiently distinct from the CS categories (i.e., CS^{a+} ,CS^{b+} and CS^{-}).

To measure SCRs and to apply the electric shocks, we used the same equipment as in Study III and followed an identical procedure. Before the learning task, stimulation intensity was adjusted individually to be unpleasant but not painful as described in Study I.

The encoding session on the first experimental day consisted of four blocks with 90 trials each, with the critical difference to Study III that we used two different types of blocks: UI-blocks versus no-UI blocks (see Figure 5A). In blocks 1 and 3, referred to as no-UI blocks, participants saw pictures from the three CS categories only, whereas blocks 2 and 4 additionally

contained UI stimuli between the CS and outcome (UI-blocks). We presented UI stimuli only in blocks 2 and 4 to rule out conditioning to the UI stimuli. More specifically, participants were assumed to learn the specific associations to the CS in the first encoding block (in which no UI stimuli were presented). The addition of the UI after the CS in the second block should not lead to conditioning to the UI stimulus, according to the classic blocking effect (Fanselow, 1998; Kamin, 1968, 1969). The third block, in which the CS was again presented without the UI stimulus, was further supposed to refresh the specific CS-outcome association, underlining that CS contingencies were independent of the UI. The inclusion of two blocks including UI stimuli between CS and outcome and two blocks not containing these UI stimuli, which were apart from the UI stimulus identical, allowed us further to directly assess the effect of the UI stimulus within Study IV and to link the findings of Study IV to those of Studies I-III.

In Blocks 1 and 3, 30 pictures of vehicles, 30 pictures of tools, and 30 pictures of clothes were presented in a pseudorandomized order. On each trial, a picture was shown in the center of the computer screen for 4.5 s, during which participants were asked to make their prediction about the probability of an electric shock. Therefore, a slider which could be adjusted to any integer value between 0% and 100% using the computer mouse was presented beneath each image. After stimulus offset, a black fixation cross was presented centrally for 4.5 s on the screen and which was immediately followed by the 200 ms outcome (shock vs. no-shock). Between trials, the fixation cross was presented on the screen for 6.5 ± 1.5 s, which again enabled us to measure the relatively slow (anticipatory) SCRs. Blocks 2 and 4 differed from blocks 1 and 3 in the inclusion of UI stimuli. During the delay of 4.5 s between CS and outcome participants did not see a fixation cross, but a picture from the UI category. Importantly, the UI was presented centrally on the screen without a slider and had no influence on the CS-shock contingencies, leaving the UI completely uninformative for the shock predictions.

Again, there were short breaks (1–2 min) between blocks during which participants had the chance to recalibrate the shock intensity, if required. Each encoding block lasted approximately 25 min, resulting in a total duration of 100 min for the entire incidental encoding-fear learning session.

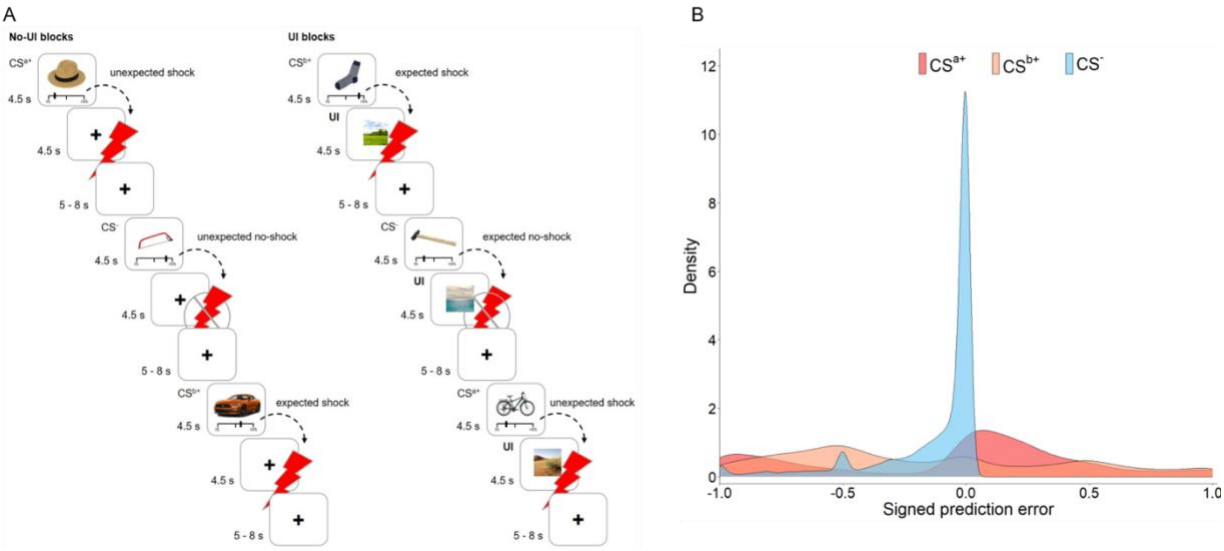
Data of respiratory frequency were collected continuously during the encoding session using a BioNomadix Respiratory Transducer (BIOPAC Systems, Goleta, California, United States) that was wrapped around the participants' upper torso approximately 5 cm below the arm pit at the point of maximum respiratory expansion and connected to the BioNomadix Respiratory Transmitter (BIOPAC Systems, Goleta, California, United States). For the measurement of heart rate, we used a NIBP100D noninvasive blood pressure monitoring system

(BIOPAC Systems, Goleta, California, United States) connected to the MP160 module. A blood-pressure cuff was placed on the participants' left arm and a double finger cuff sensor was placed on the index and middle finger of the left hand to measure heart rate continuously.

The surprise recognition test took place on the next day, 22–26 hr after encoding. Same as in Studies I-III, participants completed a short questionnaire to assess whether they anticipated a memory test and then rated how surprised they were about the recognition test on a scale from 1 (not surprised at all) to 5 (very surprised). In the recognition test, participants saw all 540 pictures they had seen during the encoding session (120 pictures of vehicles, 120 pictures of tools, 120 pictures of clothes, and 180 pictures of scenes) as well as 270 "new" pictures (60 pictures of vehicles, 60 pictures of tools, 60 pictures of clothes, and 90 pictures of scenes) that had not been presented before in a randomized order. Each trial started with a central white fixation cross on a white background for 1.5 ± 0.5 s, followed by an "old" or "new" picture presented centrally on the computer screen for 6 s. Again for each item, participants were asked to indicate whether the currently presented picture was definitely old, maybe old, maybe new, or definitely new by pressing the "1," "2," "3," or "4" button on the keyboard, respectively.

Figure 5

Experimental procedure of Study IV



Note. In the encoding task (A), participants saw a series of unique pictures from three different categories (clothes, vehicles, tools) linked to fixed probabilities to receive an electric shock (CS^{a+}= 67%, CS^{b+}=33%, and CS^{-}=0%). On each trial, participants indicated their shock expectation on a continuous scale from 0 to 100 %. Critically, in UI-blocks an UI-stimulus appeared between CS and outcome while in no-UI blocks a black fixation cross was presented on the screen. (B) PEs were equally distributed around zero. Pictures taken from "Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos" by Brodeur et al. (2014) and from "SUN database: Large-scale scene recognition from abbey to zoo" by Xiao et al. (2010). CC BY 4.0

4.2.4 Data analysis

Same as Study III, the paradigm used in Study IV enabled us to measure continuous PEs. Again, we calculated both types of PEs, that is, uPEs and sPEs. Our main analyses focus on sPEs. The detailed analyses of uPE effects are presented in the Supplemental Material.

Again, SCRs were analyzed using continuous decomposition analysis in Ledalab Version 3.4.9 (Benedek & Kaernbach, 2010) and derived in line with Studies I-III. For the anticipatory SCR, the response window was set from 0.5 after CS onset until CS offset. Additionally, we set a response window for UI-related SCRs which spanned from CS stimulus offset until the onset of the outcome (shock vs. no-shock). Additionally, we defined the outcome-related SCR to occur between 0.5 s and 4.5 s after outcome onset. Resulting estimates of the average phasic driver within each response window were returned in μS. Post-hoc comparisons of ANOVAs were always Bonferroni-corrected.

Heart rate data were analyzed using the PsychoPhysiological Modelling Toolbox in MATLAB 4.2.1 (Bach et al., 2016) and were first segmented into trial-wise epochs (spanning from CS onset until outcome offset) and then filtered with an antialias Butterworth low-pass filter (second-order, cutoff 100 Hz) and down sampled to 200 Hz. A modified offline implementation of the Pan and Tompkins (1985) real-time QRS detection algorithm was then used to identify QRS complexes. A visual correction of all interbeat intervals (IBIs) longer or shorter than the average IBI $\pm$ 2 *SD* per dataset was performed to further increase detection accuracy. The time series was then linearly interpolated to achieve a sampling rate of 10 Hz. To remove slow drifts, smooth the angles introduced by the interpolation, and reduce the influence of potentially remaining misdetections, the time series was filtered with a second-order Butterworth band-pass filter with cutoff frequencies of 0.01 and 2 Hz, respectively. For the analysis of respiration data, we also used the PsychoPhysiological Modelling Toolbox (Bach et al., 2016). Respiration data was segmented into single-trial responses starting from CS onset until outcome offset. Raw respiratory traces were converted to interpolated respiration amplitude time series with a respiratory cycle detection algorithm (Bach et al., 2016). Then, epochs were filtered offline with an anti-aliasing first-order Butterworth low-pass filter (cutoff 5 Hz) and downsampled to 10 Hz. Respiration amplitude time series were then band-pass filtered with a bidirectional Butterworth filter, with low-pass and high-pass cutoffs of 2 Hz and 0.01 Hz, respectively to remove high-frequency noise and the effects of possible slow movements of the recording device.

Again, we fitted GLMMs with a logit link function using the lme4 R package (Bates et al., 2015) and treated subjects as random effects for both the intercept and all slopes of the fixed

effects included in the model (Barr et al., 2013). The recognition of an individual item was treated as the binary dependent variable, coded "0" for misses and "1" for confident hits. We fitted models using different sets of independent variables including PEs, anticipatory as well as UI and outcome-related arousal, explicit shock prediction and block. In line with Studies I-III, we also distinguished between subsequent and previous PEs with the former referring to PE effects on the recognition of the preceding stimulus and the latter referring to PE effects on the recognition of the following stimulus.

4.2.5 Transparency and openness

The materials, data, and R analysis scripts are publicly available on the Research Data Management System of University of Hamburg and can be accessed at https://www.fdr.uni-hamburg.de/record/14147 (Loock et al., 2024). This study was not preregistered.

4.3 Results

4.3.1 Successful fear conditioning

One participant had to be excluded from the SCR analysis due to technical failure, resulting in sample of $n = 77$ for this analysis. Again, SCR data confirmed the expected fear learning process. Specifically, anticipatory SCR differed significantly between CS categories, $F(2, 152) = 4.08$, $p = .019$, partial $\eta^2 = 0.004$. Post hoc paired t-tests showed that participants showed higher anticipatory SCRs to CS^{a+} items ($M = 0.09$, $SD = 0.02$) compared to CS^{-} items ($M = 0.08$, $SD = 0.02$), $t(76) = 2.32$, $p = .023$, $d = 0.15$ (Figure 6A and 6E). Anticipatory SCRs did not differ significantly between CS^{b+} items ($M = 0.08$, $SD = 0.01$) and CS^{-} items, $t(76) = 1.74$, $p = .085$, $d = 0.04$. Notably, the SCR to the different CS types did not differ between the UI und no-UI blocks, $F(2, 152) = 1.46$, $p = .234$, partial $\eta^2 = 0.003$. In addition, the SCR in response to the UI stimuli ($M = 0.08$, $SD = 0.04$) did not differ from the anticipatory SCR for the CS^{-} items, $t(76) = -0.76$, $p = .452$, $d = 0.06$. Outcome-related SCRs were significantly higher for shocked items ($M = 0.47$, $SD = 0.27$) compared to unshocked items ($M = 0.36$, $SD = 0.27$), $t(76) = 2.44$, $p = .017$, $d = 0.27$. Explicit shock ratings further showed that participants learned the shock contingencies over the task very well. Overall, participants had a significantly higher shock expectancy for CS^{a+}($M = 0.72$, $SD = 0.17$) compared to CS^{b+} ($M = 0.50$, $SD = 0.12$), $t(76) = 10.10$, $p < .001$, $d = 1.31$, and for CS^{b+} compared to CS^{-} ($M = 0.14$, $SD = 0.19$); $t(76) = 14.87$, $p < .001$, $d = 2.20$. Importantly, shock expectancies did not differ between UI and no-UI blocks, $F(1, 76) = 3.29$, $p = .074$, partial $\eta^2 = 0.00$ (see Figure 6B and 6F). In addition, PEs were equally distributed around

zero (Figure 4B) suggesting a sufficient number of positive and negative PEs that could be analyzed.

4.3.2 General memory performance

On average, participants were moderately surprised by the recognition test (see Supplemental Material). Overall, participants showed a lower recognition performance than in the other studies ($M = 0.27$, $SD = 0.19$), which might be due to the higher number of encoded stimuli in total and the UI stimuli in particular, for which memory was poor (see below). The hit rate for CS items ($M = 0.29$, $SD = 0.19$) was significantly higher than the category-based false alarm rate ($M = 0.25$, $SD = 0.18$) for these items, $t(311) = 1.18$, $p = .024$, $d = 0.20$, thus demonstrating intact memory for the CS items.

Importantly, while false alarm rates did not differ between all four stimulus categories, $F(3, 231) = 0.59$, $p = .623$, partial $\eta^2 = 0.003$, hit rates differed significantly between categories, $F(3, 231) = 12.71$, $p < .001$, partial $\eta^2 = 0.05$ (see Figures 5C and 4G), suggesting that memory but not the response bias differed between categories. Post hoc paired *t*-tests revealed that the average hit rate for CS^{a+} items ($M = 0.31$, $SD = 0.12$) was significantly higher than for CS^{b+} items ($M = 0.27$, $SD = 0.12$), $t(77) = 2.07$, $p = .042$, $d = 0.21$, but did not differ significantly from CS^{-} items ($M = 0.29$, $SD = 0.11$); $t(77) = 1.23$, $p = .223$, $d = 0.11$. The average hit rate for CS^{b+} items did not differ from the hit rate for CS^{-} items, $t(77) = 0.87$, $p = .388$, $d = 0.09$. As expected, recognition memory performance was significantly lower for the UI items ($M = 0.20$, $SD = 0.12$) compared to all CS items, all: $t(77) > 3.97$, $p < .001$, $d > 0.43$, indicating that UI stimuli were considered irrelevant by participants and that memory was overall significantly enhanced for predictive stimuli. *d*' differed significantly between categories ($F(3,216) = 5.25$, $p = .002$, partial $\eta 2 = 0.035$). While recognition memory did not differ between both the CS^{a+} items ($M = 0.14$, $SD = 0.63$) and CS^{b+} ($M = 0.12$, $SD = 0.64$; $t(72) = 0.16$, $p = .876$, $d = 0.02$), items differed compared to CS^{-} items ($M = 0.15$, $SD = 0.62$; vs. CS^{a+}: $t(72) = -0.15$, $p = .883$, $d = 0.02$; vs. CS^{b+}: $t(72) = -0.28$, $p = .784$, $d = 0.04$), we found a significantly lower *d*' for UI items ($M = -0.20$, $SD = 0.66$) compared to all CS items (all $t(72) > 3.01$, all $p < .004$, all $d > 0.43$).

Moreover, recognition memory performance differed significantly between blocks with UI and without UI items. Recognition memory for CS items was significantly better in no-UI blocks (hit rate: $M = 0.30$, $SD = 0.07$) compared to UI blocks, hit rate: ($M = 0.27$, $SD = 0.07$), $t(155) = 4.95$, $p < .001$, $d = 0.17$, suggesting that the appearance of an UI stimulus affected memory formation for the predictive stimuli.

Again, participants rated their level of surprise related to the recognition test on a scale ranging from 1 (not surprised at all) to 5 (very surprised). On average, participants were moderately surprised by the recognition test ($M = 2.96$, $SD = 1.17$). Nine participants chose the 'not surprised at all' option. Because excluding them did not affect the results, they were included in all analyses.

4.3.3 Modelling recognition performance at item level

To elucidate the mechanisms of episodic memory formation, we again fitted GLMMs with the recognition of an item as the binary-dependent variable and added the relevant independent predictors in a step-wise manner, in line with the previous studies.

We started with a first minimal model in which we tested whether sPEs contribute to later recognition. Therefore, we treated the subsequent sPE (ranging from −1 to 1) following a CS as the sole independent variable to predict the binary recognition of this CS item, irrespective of the appearance of an UI stimulus. Estimates obtained revealed that (subsequent) sPEs ($z = 2.21$, $p = .027$, $\beta = 0.09$) showed a positive relationship with later memory. To rule out that the PE-effects were confounded with the shock prediction, we also computed a model where we added the explicit shock prediction as a predictor to the previous model. When controlling for the explicit shock prediction, the memory enhancing effect of the subsequent sPEs remained significant ($z = 2.52$, $p = .011$, $\beta = 0.12$).

In addition, we also set up a model that tested whether subsequent sPEs contribute to the recognition of UI stimuli treating subsequent sPEs as the sole independent variable to predict the binary recognition of an UI item. Crucially, estimates revealed that subsequent sPEs ($z = -0.66$, $p = .510$, $\beta = -0.04$) showed no significant relationship with later UI recognition, suggesting that the effect of subsequent sPEs is specific to the predictive stimulus and not found for UI stimuli presented between CS and outcome.

In a follow-up model, we added anticipatory arousal, UI-related arousal and outcome-related arousal as predictors to the minimal subsequent sPE-model for the binary recognition of a CS item. Anticipatory SCRs ($z = 0.63$, $p = .532$, $\beta = 0.09$), UI-related SCRs ($z = 0.72$, $p = .474$, $\beta = 0.10$) and outcome-related SCRs ($z = -0.03$, $p = .974$, $\beta = -0.00$) did not influence item recognition significantly. Even after controlling for arousal effects on memory, we still obtained a significant effect of subsequent sPE on CS memory ($z = 2.25$, $p = .024$, $\beta = 0.10$). To further elucidate whether heart rate and respiration amplitude as additional arousal metrics contribute to recognition memory, we also set up additional models separately for both predictors. Estimated showed that neither respiration amplitude ($z = 0.34$, $p = .732$, $\beta = 0.01$) nor heart rate ($z = 0.03$, $p = .977$, $\beta = 0.00$), predicted CS memory significantly. Following up

on that, we added the subsequent sPE to each model as a predictor separately. When controlling for respiration amplitude, the subsequent sPE effect on memory of the predictive item remained significant ($z = 2.21$, $p = .027$, $\beta = 0.09$). A model including heart rate and subsequent sPE also yielded a significant subsequent sPE effect on item recognition ($z = 2.31$, $p = .021$, $\beta = 0.10$), when controlling for heart rate.

To examine whether sPE effects on memory interfere with the appearance of UI stimuli, we included block (no-UI block vs. UI block) as a predictor in an additional set of models. First, we set up a model that treated the subsequent sPE and block including their interaction as independent variables to predict the binary recognition of an item. We obtained a significant *Subsequent sPE* × *Block* interaction effect on memory ($z = -2.45$, $p = .014$, $\beta = -0.15$), suggesting that the "retrospective" PE effect on memory is influenced by the appearance of an UI stimulus. Accordingly, we set up separate models for blocks that contained UI items and for blocks that did not contain UI items. We treated the subsequent sPE as the sole independent variable to predict the binary recognition of an item. While this revealed that episodic memorywas significantly increased by subsequent sPEs ($z = 2.90$, $p = .004$, $\beta = 0.16$) in no-UI blocks (see Figure 6D), we obtained a nonsignificant effect of subsequent sPEs on memory in UI blocks ($z = 0.63$, $p = .53$, $\beta = 0.13$; see Figure 6H), suggesting that those retrospective PE effects on memory disappear when an UI stimulus is presented between CS and outcome (i.e., PE). Even when controlling for anticipatory, UI- and outcome-related arousal, the pattern of results remained unchanged indicating a memory boost induced by subsequent sPE in no-UI blocks ($z = 2.87$, $p = .004$, $\beta = 0.16$), whereas there was no effect in UI blocks ($z = 0.58$, $p = .564$, $\beta = 0.03$).

Next, we performed additional models in which we treated the previous sPE as the sole independent variable to predict the binary recognition of the following item. This revealed a significant negative effect of sPEs on memory for items following the PE ($z = -2.05$, $p = .040$, $\beta = -0.07$), that is, previous sPEs appeared to be associated with a memory impairment. To rule out that the previous PE effects were confounded with the shock prediction, we also computed a model where we added the explicit shock prediction as a predictor to the previous model. When controlling for the explicit shock prediction, the effect of the previous sPEs remained significant ($z = -2.20$, $p = .028$, $\beta = -0.07$).

Again, we added anticipatory, UI- and outcome-related SCRs to the former model to investigate confounds with physiological arousal. This revealed nonsignificant effects of anticipatory SCRs ($z = 0.73$, $p = .467$, $\beta = 0.10$), UI-related SCRs ($z = 0.82$, $p = .413$, $\beta = 0.12$), and outcome-related SCRs on memory formation ($z = 0.01$, $p = .992$, $\beta = 0.00$). The previous

negative sPE effect on later memory remained significant ($z = -2.05$, $p = .041$, $\beta = -0.07$). In addition, we set up a model including respiration amplitude and previous sPE as variables to predict recognition of the following item. When controlling for respiration amplitude, the previous sPE effect on memory of the following item remained significant ($z = -1.99$, $p = .047$, $\beta = -0.07$). A model including heart rate and previous sPE yielded a trending previous sPE effect on recognition of the following item ($z = -1.76$, $p = .078$, $\beta = -0.06$).

In addition, we also set up a model that tested whether previous sPEs contribute to the recognition of UI stimuli treating sPEs as the sole independent variable to predict the binary recognition of the following UI item. Again, the model estimates revealed that previous sPEs ($z = -1.12$, $p = .261$, $\beta = -0.07$), showed no significant relationship with item recognition suggesting that the recognition of UI is independent of previous PEs.

In a next step, we tested whether the previously observed effect of (previous) PEs on the memory for items following the PE would be influenced by the appearance of an uninformative stimulus. Estimates obtained showed no significant *Previous sPE* × *Block* interaction effect on recognition of the following item ($z = -0.11$, $p = .915$, $\beta = -0.01$), suggesting that previous sPE effects on memory might be irrespective of the appearance of uninfomative stimuli. Even though the critical interaction effect was nonsignificant, in an explorative analysis, we set up separate models for blocks that contained UI items and for blocks that did not contain UI items. We treated the previous sPE as the sole independent variable to predict the binary recognition of the following item. Notably, this revealed that episodic memory was not influenced by previous sPEs in no-UI blocks ($z = -1.35$, $p = .177$, $\beta = -0.06$) nor in UI blocks ($z = -1.42$, $p = .155$, $\beta = -0.07$). When controlling for anticipatory arousal, UI-related, and outcome-related arousal, the previous sPE effect on memory of the following item remained nonsignificant in no-UI blocks ($z = -1.36$, $p = .174$, $\beta = -0.07$) and UI blocks ($z = -1.42$, $p = .155$, $\beta = -0.07$).

To investigate whether subsequent sPEs and previous sPEs reflect distinct mechanisms, we added both types of sPEs as independent variables to an additional model. Estimates obtained revealed that subsequent sPEs showed a positive relationship with item recognition, $z = 2.17$, $p = .030$, $\beta = 0.09$, while previous sPEs, $z = -2.03$, $p = .043$, $\beta = -0.07$, showed a negative relationship with item recognition. In a follow-up model, we added anticipatory, UI-related and outcome-related arousal as predictors. In this model, neither anticipatory arousal, $z = 0.61$, $p = .545$, $\beta = 0.08$, nor UI-related arousal, $z = 0.83$, $p = .405$, $\beta = 0.12$, nor outcome-related arousal, $z = -0.04$, $p = .972$, $\beta = -0.00$, influenced item recognition significantly. After controlling for arousal, we obtained a significant memory boost associated with subsequent PEs, $z = 2.23$, $p =$

.026 , β = 0.10, and a significant memory impairment associated with previous sPEs, z = -2.05, p = .041 , β = -0.07.

4.3.4 Effects of uPEs

We started with a minimal model in which we tested whether uPEs contribute to later recognition. Therefore, we treated the subsequent uPE (ranging from 0 to 1) following on a CS as the sole independent variable to predict the binary recognition of this CS item, irrespective of the appearance of an uninformative stimulus. Estimates obtained revealed no effect of subsequent uPEs on item recognition, z = 0.13, p = .90, β = 0.01. Additionally, we also set up a model that tested whether subsequent uPEs contribute to the recognition of uninformative stimuli treating uPEs as the sole independent variable to predict the binary recognition of an UI item. Estimates obtained revealed that subsequent uPEs, z = -0.93, p = .353 , β = -0.08, showed no significant relationship with item recognition suggesting that the recognition of UI is independent of uPEs.

In a follow-up model, we added anticipatory arousal, UI-related arousal and outcome-related arousal as predictors to the subsequent uPE-model to the binary recognition of a CS item. Anticipatory SCRs (z = 0.73, p = .464 , β = 0.10), UI-related SCRs (z = 0.69, p = .491 , β = 0.10) and outcome-related SCRs (z = -0.04, p = .973 , β = -0.00) did not influence item recognition significantly. When controlling for arousal effects on memory, the effect of subsequent uPE on CS memory remained non-significant, z = 0.06, p = .95, β = 0.01. To rule out that the subsequent uPE-effects were confounded with the shock prediction, we also computed a model where we added the explicit shock prediction and the subsequent uPE as predictors. When controlling for the explicit shock prediction, the effect of the subsequent uPE remained non-significant, z = -0.73, p = .463 , β = -0.06.

To examine whether uPE effects on memory interfere with the appearance of uninformative stimuli, we included the factor block as a predictor in an additional set of models. We set up a model that treated the subsequent uPE and block including their interaction as independent variables to predict the binary recognition of an item. We obtained a non-significant *subsequent uPE* × *block* interaction effect of subsequent uPEs and block on memory, z = 0.24, p = .811 , β = 0.02. Following up on that, we set up separate models for blocks that contained UI items and for blocks that did not contain UI items. We treated the subsequent uPE as the sole independent variable to predict the binary recognition of the predictive item. This revealed that episodic memory was not influenced by subsequent uPEs in no-UI blocks, z = -0.07, p = .947, β = -0.01, and UI blocks, z = 0.21, p = .834, β = 0.02.

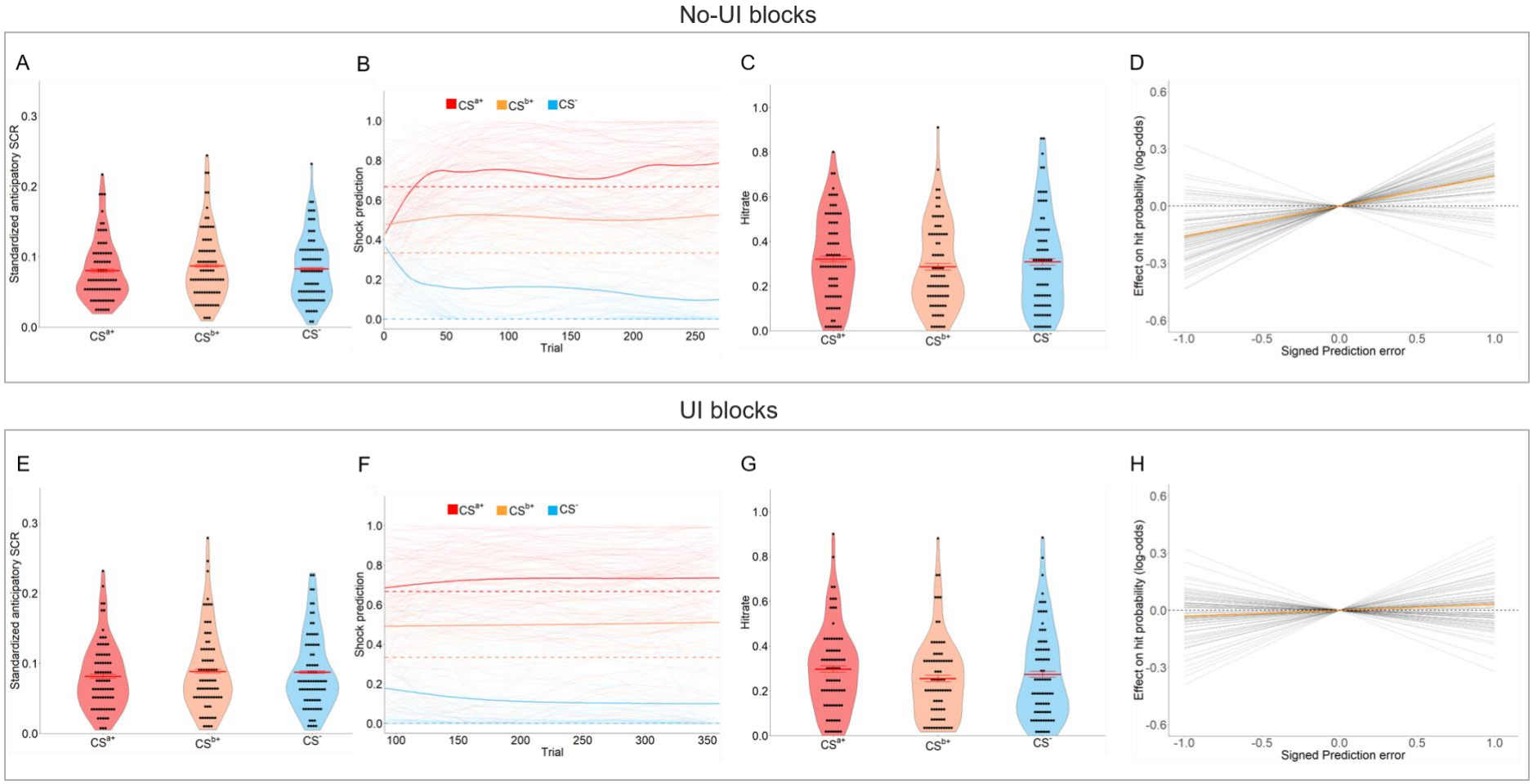
Next, we performed additional models in which we treated the previous uPE as the sole independent variable to predict the binary recognition of the following item. This revealed a non-significant effect of previous uPEs on memory for items following the PE, z = -0.70, p = .487, β = -0.04. To rule out that the PE-effects were confounded with the shock prediction, we also computed a model where we added the explicit shock prediction as a predictor to the previous model. When controlling for the explicit shock prediction, the effect of the previous uPEs was not significant anymore, z = -0.59, p = .554 , β = -0.04. Again, we added anticipatory, UI-related and outcome-related SCRs to theminimal previous uPE-model to investigate confounds with physiological arousal. This revealed no effects of anticipatory SCRs, z = 0.63, p = .531 , β = 0.09, UI-related SCRs, z = 0.68, p = .495, β = 0.10, and outcome-related SCRs on memory formation, z = 0.11, p = .912 , β = 0.01. The previous uPE effect remained non-significant, z = -0.80, p = .424 , β = -0.05. Additionally, we also set up a model that tested whether previous uPEs contribute to the recognition of UI stimuli treating uPEs as the sole independent variable to predict the binary recognition of the following UI item. Again, estimates obtained revealed that previous uPEs, z = -0.49, p = .621, β = -0.04, showed no significant relationship with item recognition suggesting that the recognition of UI is independent of previous uPEs.

In a next step, we tested whether the previously observed effect of PEs on the memory for items following the PE would be influenced by the appearance of an uniformative stimulus. Again, estimates obtained showed no *previous uPE* × *block* interaction effect on recognition of the following item, z = 0.53, p = .600 , β = 0.05. Following up on that, we set up separate models for blocks that contained UI items and for blocks that did not contain UI items. We treated the previous uPE as the sole independent variable to predict the binary recognition of the following item. This revealed that episodic memory was not influenced by previous uPEs in no-UI blocks, z = -0.74, p = .457, β = -0.06, and UI blocks, z = -0.54, p = .586, β = -0.04.

Next, we added both types of uPEs as independent variables to an additional model. Estimates obtained revealed that neither subsequent uPEs, z = 0.15, p = .880 , β = 0.01, nor previous uPEs, z = -0.61, p = .545, β = -0.04, showed a significant relationship with item recognition. In a follow-up model, we added anticipatory, UI-related and outcome-related arousal as predictors. In this model, neither anticipatory arousal, z = 0.60, p = .551 , β = 0.08, nor UI-related arousal, z = 0.65, p = .513, β = 0.10, nor outcome-related arousal, z = 0.11, p = .914 , β = 0.01, influenced item recognition significantly. After controlling for arousal, we obtained non-significant effects of subsequent uPEs, z = 0.09, p = .93 , β = 0.01, and previous uPEs on recognition memory, z = -0.72, p = .471 , β = -0.04.

Figure 6

Results of Study IV



Note. Mean standardized anticipatory SCR (A,E) and hitrates (C,G) confirmed successful fear conditioning, as reflected in significantly elevated SCR and increased hitrates of CS^{a+} compared with CS^{-} items in no-UI and UI blocks. Black dots show data from individual participants. Thick red bar represents group mean, while thin red bars show ±1 standard error of the mean. (B,*F*) Participants' mean shock expectancy ratings (thick lines) approached the true shock probabilities (dotted lines) relatively fast in no-UI- and UI blocks. Subsequent sPEs boosted item memory only in no-UI blocks (G) while the effect was abolished in UI-blocks (H).

4.4 Conclusion

In the present study, we aimed to elucidate the cognitive mechanisms underlying aversive PE-driven memory enhancements for inherently neutral events. While we obtained evidence indicating that the PE-induced memory boost also extends to stimuli presented after the PE, i.e., prospective PE effects (see Studies I and II), the results from Study IV emphasize the specificity of this memory enhancement. It is exclusively linked to predictive stimuli, with uninformative stimuli even shown to interfere with the PE-driven memory enhancement.

In Studies III and IV, we assessed PEs on a continuous scale, allowing us to differentiate between positive and negative sPEs which have been shown to exert differential effects on memory formation (Kalbe & Schwabe, 2022b; Rouhani & Niv, 2021). Our results of Studies III and IV show consistently that negative PEs were associated with impairing effects on subsequent memory, whereas positive PEs were linked to enhanced memory. This pattern of results is in stark contrast to our previous findings (Kalbe & Schwabe, 2022b), which showed the exact opposite pattern. However, a notable distinction between these studies is the testing environment and the number of experimental trials. Our previous study took place in a magnetic resonance imaging scanner, which may have resulted in higher state anxiety levels known to modulate PE processing (Hein & Herrojo Ruiz, 2022). Indeed, participants that volunteer for magnetic resonance imaging studies have been shown to be characterized by reduced trait anxiety levels compared to participants in behavioral experiments (Charpentier et al., 2021). Trait anxiety is typically correlated with depressive mood known to affect PE effects on memory formation (Rouhani & Niv, 2019). However, the potential modulation of PE effects on memory by trait anxiety remains speculative and needs to be tested explicitly in future studies. In addition, we increased the trial number significantly (360 trials vs. 120 trials in our previous study), underlining the high validity of our empirical findings. Because of the increased trial number, participants also received substantially more electric shocks, potentially resulting in higher sensitivity to the aversive outcome (Chen et al., 2000; Lonsdorf et al., 2017), which may have rendered the shock experience even more aversive and hence positive PEs more intense.

Moreover, Study IV demonstrates no memory enhancement for uninformative stimuli presented between the predictive stimulus and PE. How can we reconcile the absence of memory enhancement for these uninformative stimuli with the prospective PE effects, which imply memory enhancement for subsequent stimuli not informative for the current PE event?

The answer to this question may relate to the predictive value of the stimuli per se. Participants presumably quickly learned the irrelevance of the uninformative stimuli, resulting in shallow processing as also reflected in the overall low-memory performance for uninformative stimuli. In contrast to these uninformative stimuli, the CS following the PE does carry informative value, namely for the subsequent PE event. Thus, PEs enhanced memory for predictive stimuli around the time of the PE event but not for entirely unpredictive stimuli, refuting the idea of a PE-induced window of unselective memory enhancement.

Intriguingly, the presentation of an uninformative stimulus between the CS and outcome (i.e., PE) even abolished the PE-induced memory enhancement for the predictive CS. This finding is remarkable, suggesting that uninformative information interferes with the association between the predictive stimulus and PE. Specifically, it highlights the necessary active maintenance of the predictive stimulus until the PE, which the UI stimulus interfered with. The predictive stimulus may be stored in working memory (Baddeley, 1992; Oberauer et al., 2003) and thus be highly vulnerable to competing stimuli appearing during the maintenance phase.
Because previous research used SCR as the only measure of arousal, we added heart rate and respiration amplitude as additional arousal measures in Study IV, to further disentangle arousal- and PE-related effects on memory. When controlling for these measures, the memory-enhancing effect of subsequent PEs remained. These findings suggest that physiological arousal (beyond SCR) cannot account for the memory boost alone indicating that retrospective PE-induced memory enhancements presumably go at least partly beyond the mere effects of arousal on memory formation.

In sum, the findings from Study IV provide insights into the cognitive mechanisms involved in aversive PE-driven enhancements of memory for surrounding neutral events. Notably, these PE effects are not unspecific, as reflected in the absence of any memory boost for uninformative stimuli presented between CS and PE. Rather, PEs appear to enhance memory for predictive stimuli encountered around a PE event. Importantly, these PE effects are sensitive to interference, pointing to an involvement of working memory maintenance. Our findings provide insights into the fundamental question of which of the many stimuli that we are continuously presented with are stored in long-term memory: those that bear predictive value for unexpected emotional events.

5 Does the PE-effect on memory depend on the neural states surrounding the PE event?

This chapter is currently under peer-review in modified form: Loock, K., Heinbockel, H., Kalbe, F., & Schwabe, L. Prediction error-related memory enhancement depends on the neural state surrounding the prediction error event.

5.1 Background

Adaptive memory enables organisms to leverage past experiences to guide actions and choices (Shohamy & Adcock, 2010). However, not all events are stored equally well in memory, preference is rather given to information being crucial for predicting relevant outcomes. In support of this notion, research shows that PEs – mismatches between expected and actual outcomes – associated with rewarding or aversive events enhance memory for preceding stimuli (Ergo et al., 2020; Kalbe & Schwabe, 2020; Rouhani & Niv, 2021; Rouhani et al., 2023). Although these PE effects are fundamental to our understanding of adaptive memory and may have significant implications for educational contexts and psychopathology, the brain mechanisms underlying the impact of PEs on memory for preceding events remain poorly understood.

Specifically, two short-lived neural mechanisms might drive PE effects on memory. First, PEs could evoke a transient reactivation of the preceding predictive stimulus, promoting its memory storage. This mechanism aligns with evidence indicating that post-encoding reactivation is essential for subsequent recall (Staresina et al., 2013; Tambini et al., 2020). A second mechanism may involve the neural state just before the PE. PEs might strengthen memory for preceding events, if these events are still neurally maintained when the PE occurs. This is in line with synaptic or behavioral tagging models proposing that pre-activated representations can be enhanced by a subsequent salient event, such as a PE (Moncada et al., 2015). Potential candidate mechanisms for neural maintenance include alpha oscillations, implicated in attention to task-relevant stimuli (Payne & Sekuler, 2014), theta oscillations, related to the reinstatement of memory representations or the binding of associative memory (Kota et al., 2020; Nyhus & Curran, 2010; Staudigl & Hanslmayr, 2013), and the neural reactivation of the representation of the preceding stimulus.

If PE effects on memory require the reactivation or maintenance of the preceding stimulus around the PE event, this raises the question of whether interference with these mechanisms could reduce or even abolish PE effects on memory. The superior parietal cortex has been repeatedly shown to be crucial for stimulus maintenance, working memory processes, and top-down attentional updating (Corbetta et al., 1995; D'Esposito & Postle, 2015; Ester et al., 2015;

Koenigs et al., 2009; Wager & Smith, 2003), making it a promising candidate for the maintenance of the predictive stimulus. Thus, we hypothesized that inhibiting superior parietal cortex functioning would reduce PE effects on memory for preceding stimuli.

In this pre-registered study, we combined 'neuro-navigated' transcranial magnetic stimulation (TMS) with electroencephalography (EEG) and multivariate pattern analysis to elucidate the brain mechanisms underlying PE effects on memory. Specifically, we tested whether (i) PEs induce a neural reactivation of the preceding stimulus, (ii) PE effects on memory require a specific neural state, associated with alpha or theta oscillations or a neural representation of the preceding stimulus shortly before the PE, and (iii) inhibitory stimulation over the superior parietal cortex reduces PE effects on memory. To these ends, we applied continuous theta-burst stimulation (cTBS) over the superior parietal cortex before participants completed a combined incidental encoding-fear learning task, while EEG was recorded. During this task, participants encoded trial-unique stimuli and predicted whether these would be followed by an electric shock. Memory was tested 24 hours later. We hypothesized that (signed) PEs would enhance subsequent memory and that these effects would be dependent on the neural state and representation around the PE. Additionally, we predicted that cTBS over the superior parietal cortex would generally reduce PE effects on memory.

5.2 Methods

5.2.1 Participants

One hundred twenty-two healthy right-handed volunteers participated in this study (69 female; age: M = 25.55 years, SD = 3.63 years, range = 19-33 years). Exclusion criteria were screened in a standardized interview and comprised: insufficient command of German, life-time history of any neurological, cardiovascular or psychiatric diseases, medication intake or substance abuse, and contraindications for MRI measurements or TMS. All participants provided written informed consent before participation and received a monetary reimbursement. The ethics committee of the Faculty of Psychology and Human Movement Science at the University of Hamburg approved the study (2022_055_Loock_Schwabe), which was carried out in line with the Declaration of Helsinki.

The target sample size was based on a previous behavioral study from our lab that showed an effect of aversive signed PE on memory formation using the same task in n = 120 participants with a power of .92 (Loock, Kalbe & Schwabe, 2025). In the present study, a post-hoc power simulation using the R-package simR (Green & MacLeod, 2016) for the observed effect of aversive signed PE on memory formation and our final sample size of n = 118

participants (due to exclusions in EEG data analysis) yielded a power of .98 based on 1000 simulations.

We employed a between subjects-design with the factor stimulation group (TMS vs. sham). Participants were randomly assigned to the TMS group (n = 62, 31 female) and the sham group (n = 60, 38 female).

5.2.2 Procedure

The experiment consisted of one MRI session – during which we acquired an anatomical brain image required for 'neuro-navigated' TMS – and two experimental sessions (Day 1 and Day 2) that took place on two consecutive days (see Figure 7).

Before experimental Day 1, we acquired T1-weighted structural Magnetic Resonance (MR) images of each participant using a 3T Siemens PRISMA scanner located at the University Medical Center Hamburg-Eppendorf. We utilized a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence to collect the anatomical images that had a voxel size of 0.8 x 0.8 x 0.9 mm^3 and consisted of 256 slices. The imaging parameters for the MPRAGE sequence were a repetition time (TR) of 2.5 s and an echo time (TE) of 2.12 ms. These structural brain images were used for 'neuro-navigating' the TMS or sham stimulation.

Upon arrival on experimental Day 1, participants provided written informed consent and filled out questionnaires assessing depressive symptoms (BDI-II; Beck et al., 1961), sleep quality (PSQI; Buysse et al., 1989), state and trait anxiety (STAI-S and -T; Spielberger et al., 1970), and chronic stress (TICS; Schulz et al., 2004). While completing the questionnaires, the EEG cap and electrodes were set up. A stimulation electrode for applying electric shocks during the incidental encoding-fear learning task was placed on the participant's right lower leg, approximately 20 cm above the heel. Shock intensity was adjusted individually to be unpleasant but not painful in a stepwise-manner. More specifically, 200ms single pulse shocks were administered consecutively with an initial intensity of 15V until they were perceived as unpleasant but not painful. For electrical stimulation, we used the STM-200 stimulation module connected to the MP-160 data acquisition and analysis system (BIOPAC systems, Goleta, California, United States). To assess skin conductance responses (SCRs) as an indicator of physiological arousal, electrodes were placed on the distal phalanx of the index finger and the third finger of the left hand. Next, participants individual motor-thresholds for TMS were determined. Thereafter, participants performed the first session of a Delayed-matching-to-sample (DMS) task (see below), which served to later train a classifier based on L2-penalized logistic regression for EEG-based decoding, before they underwent either the sham or TMS stimulation targeting the right superior parietal cortex. Immediately after the TMS or sham

stimulation, participants performed a combined incidental encoding-fear learning task in which they were asked to predict whether a stimulus presented on screen would be followed by an electric shock (see below). After half of the task, we administered a second TMS or sham stimulation to maintain the effects of the stimulation throughout the task. After finishing the encoding task, participants performed a second session of the DMS task. In total, Day 1 took about 4.5 hours per participant. Approximately twenty-four hours later, participants returned for a surprise recognition test in which we assessed their memory for the stimuli encoded on Day 1.

5.2.3 Day 1: Delayed-matching-to-sample (DMS) task

In order to decode neural stimulus representations before and after PE, we trained a classifier based on the EEG data from a DMS task (Meier et al., 2022; see Figure 7). The DMS is common for examining working memory processes (Anderson & Colombo, 2019). This task was performed twice, once before TMS/sham stimulation and once after the combined incidental encoding-fear learning task to rule out any time- or TMS-related biases in the classifier. In each of the two sessions, participants completed 150 trials. During each session, participants saw images of three different stimulus categories (animals, scenes, tools). Stimuli were taken from available image databases, i.e., Bank of Standardized Stimuli (Brodeur et al., 2010; Brodeur et al., 2014), SUN database (Xiao et al., 2010), Konklab (Konkle et al., 2010), and open online sources. In total, the stimulus set consisted of 300 unique pictures of animals, scenes and tools respectively, with 100 images per category isolated on white background. We pre-allocated two different stimulus sets of 150 pictures each (50 animals, 50 scenes, 50 tools) of which one was used in the first DMS session and the other in the second DMS session. All stimuli were assumed to be emotionally neutral and represented an unique exemplar of its category. On each trial, a target stimlus was presented for 2 s in the center of a grey screen. Participants were instructed to keep this trial-specific target in mind for a 2s delay period during which a black fixation cross was presented on the screen. After the delay period, the target stimulus and two distractor stimuli appeared on the screen simultaneously and participants were required to select via button press within 2 s which stimulus they had seen before. The position of the target stimulus on the screen (left, center, right) was randomized across trials and response buttons corresponded to the numbers ‘1’ (left), ‘2’ (center) and ‘3’ (right) on the keyboard. The distractors were either drawn randomly from the same category as the target or from the two left-over stimulus categories. Trials were pseudo-randomized with the restriction that successive trials did not include targets from the same category for more than three consecutive times. Between trials, there was a fixed interval of 2 s. An example trial of the

DMS task is presented in Figure 7. Importantly, the stimuli used in the DMS task did not overlap with those used in the incidental encoding-fear learning task.

5.2.4 Day 1: TMS and sham stimulation

In order to examine the functional role of the superior parietal cortex in PE-induced memory enhancements, we used neuro-navigated TMS over the right superior parietal cortex before participants underwent the incidental encoding- fear learning task. For stimulation, we used a PowerMag Research 100 stimulator (MAG & More GmbH, Munich, Germany) which applies repetitive transcranial magnetic stimulation (rTMS). Depending on the experimental group (TMS vs. sham), two different figure-eight TMS coils were used: a PMD70-pCool coil (MAG & More GmbH, Munich, Germany; max. magnetic field strength of 2T) was used for continous Theta Burst stimulation (cTBS) in the TMS condition, whereas the PMD70-pCool-SHAM (MAG & More GmbH, Munich, Germany; minimal magnetic field strength) was used in the sham condition. Importantly, the sham condition induced a similar sensory but widespread experience on the scalp not pervading the skull. We used a double-blind protocol, in which neither the participant nor the experimenter was aware of the stimulation condition. Participants were asked to guess which treatment they had received at the end of the experiment.

5.2.4.1 Motor threshold determination

The motor threshold (MT) was determined before participants started to do any task on Day 1, but were already wearing an EEG cap without electrodes attached to it (see Grob et al., 2024). The MT determination was used to determine the appropriate magnetic field strength per individual. Disposable, pre-gelled Ag/ACL surface electromyography (EMG) electrodes were attached to the participant's right hand: An active electrode was placed on the abductor pollicis brevis muscle, with a reference electrode on the bony landmark of the index finger and a ground electrode on the tip of the ulna bone. In order to locate the motor hotspot (MH), we located the center of the head, moved 5 cm leftward and 3.5 cm to the forehead at an angle of 45° and marked this area as the center of a 3×3 point-grid area. Each point was 1cm apart from its neighbours. Starting at 40 % maximum stimulator output, we gradually increased the output intensity of the stimulation (with a step size of 5%) while adjusting the TMS coil to an angle of 45° on the *z*-axis. Then, we screened the 3×3 search grid for the motor hotspot delivering single 10 Hz pulses. As soon as the MH was found, the MT was determined at that certain location. The MT was defined as the minimum percentage of maximum stimulator output over the left motor cortex (area: M1) necessary to elicit motor evoked potentials (MEPs) with a peak-to-peak amplitude of 50 μV in response to at least eight out of 16 consecutive single pulses.

5.2.4.2 Neuro-navigation

Individual T1-weighted anatomical MR images of the participants were used for neuro-navigation with the PowerMag System (MAG & More GmbH, Munich, Germany). This procedure ensured a precise and individually tailored coil placement being aligned with the superior parietal cortex as target area. An infrared camera (Polaris Spectra) was used to locate and track the participant's head and the TMS or sham coil in space. Based on the information we gained from the T1-weighted MR images, we created 3D-models of the participant's head which allowed us to precisely locate the right superior parietal cortex individually based on Talairach (TAL) coordinates from previous work (TAL: 21, -54, 51; Ester et al., 2015) which we transformed into system-compatible MNI coordinates (20, -58, 57). As we assume that memory maintenance processes might be involved in PE-related memory enhancements, we decided to target the superior parietal cortex which has been repeatedly associated with working memory and stimulus representation (D'Esposito & Postle, 2015; Ester et al., 2015; Koenigs et al., 2009). After entering the target MNI coordinates (20, -58, 57), the coil was positioned in alignment with the neuro-navigation system. We aimed for a brain-to-target distance of less than 3 cm to ensure the shortest distance to the cortex.

5.2.4.3 Stimulation protocol

We applied cTBS using the active coil for the TMS group or the sham coil for the sham group. It is assumed that cTBS leads to an inhibitory effect on the target brain region under stimulation (Grob et al., 2024; Huang et al., 2005; Jannati et al., 2023). Based on a cTBS protocol by Grob et al. (2024), participants received a series of theta bursts with three magnetic pulses (triplets) at a frequency of 50 Hz, with the triplets being repeated at a rate of five Hz, i.e., five triplets per second. In total, 600 magnetic pulses over 40 sec per participants were administered to the target area. We fixed the coil using a tripod placed behind the participant to maintain a precise (TMS or sham) stimulation of the right superior parietal cortex (MNI: 20, -58, 57) with less than 3cm of brain-to-target distance.

5.2.5 Day 1: Incidental encoding-fear learning task

To examine the neural mechanisms underlying PE-induced memory enhancements, participants completed an incidental encoding-fear learning paradigm immediately after TMS or sham stimulation while EEG was recorded (see Figure 7). Stimuli were taken from existing databases, i.e., Bank of Standardized Stimuli (Brodeur et al., 2010; Brodeur et al., 2014), SUN database (Xiao et al., 2010), Konklab (Konkle et al., 2010), and open online sources. The stimulus set consisted of 540 pictures of animals, scenes and tools, with 180 pictures per category isolated on white background. All stimuli were assumed to be of neutral valence and

represented an unique exemplar of its category. Out of this pool, 360 pictures (120 per category) were randomly drawn and used during encoding on Day 1. The remaining 180 pictures (60 pictures per stimulus category) served as lures in the recognition test on Day 2. The order of item presentation was randomized across participants.

In the incidental encoding-fear learning task, participants were instructed that they would see a stream of pictures (animals, scenes, tools) presented one after another on the screen and that some pictures will be followed by an electric shock. Participants were asked to predict how likely a shock would be to follow the presented picture by adjusting a slider on the screen to a value that corresponded with their prediction of the shock probability (range: 0 to 100%). For each trial, we derived a PE which was calculated as the relative value of the difference between participants' continuous explicit shock expectancy ratings (ranging from 0, corresponding to full confidence that no shock would occur, to 1, corresponding to full confidence that a shock would occur) and the actual binary outcome of the trial (coded '0' if no shock occurred and coded '1' if a shock occurred in the current trial). Importantly, participants were neither informed about the true shock contingencies, nor about the recogniton test on Day 2. They were informed that the shock occurences were not affected by their predictions, but that they should learn by trial-and-error to improve their predictions over the duration of the task. Unbeknownst to the participants, the probabilities of a shock were linked to the three picture categories. One category served as CS^{a+} (67 % shock probability), one as CS^{b+} (33 % shock probability), and one as CS^{-} (0% shock probability). The assignment of image categories (i.e., animals, scenes, tools) to the CS categories (i.e., CS^{a+}, CS^{b+}, CS^{-}) was counterbalanced across participants and groups. Throughout the incidental encoding-fear learning task, SCR was measured as an indicator of physiological arousal by using electrodes on the individual's left hand. Electric shocks were applied via the shock electrode placed at the participant's lower right leg.

In total, the incidental encoding-fear learning task consisted of 360 trials split into four blocks of 90 trials. In each block, 30 pictures of animals, 30 pictures of scenes and 30 pictures of tools were presented in a pseudorandomized order, so that no more than three pictures of the same category appeared in a row. On each trial, a picture was shown in the center of the screen for 4.5 s, during which participants were asked to make their prediction about the probability of an electric shock (Figure 7). A slider was presented underneath each item which could be individually adjusted to any integer value between 0 % and 100 % by using the computer mouse. After stimulus offset, a black dot appeared centrally on the screen which coterminated with the 200ms-outcome (shock vs. no-shock), i.e., in no-shock trials, the transition from dot to fixation

cross indicated that the CS was not followed by a shock. Critically, the duration of the dot's presentation on the screen ranged randomly between 0 and 10 s per trial to vary the critical CS-outcome delay. After the outcome, a black fixation cross centered on grey background was presented for 6.5 ± 1.5 s. Between blocks, there was a short break (1-2 min) during which participants had the chance to recalibrate the shock intensity and rest, if required. Each encoding block lasted approximately 25 min, resulting in a total duration of 100 min for the entire incidental encoding-fear learning task.

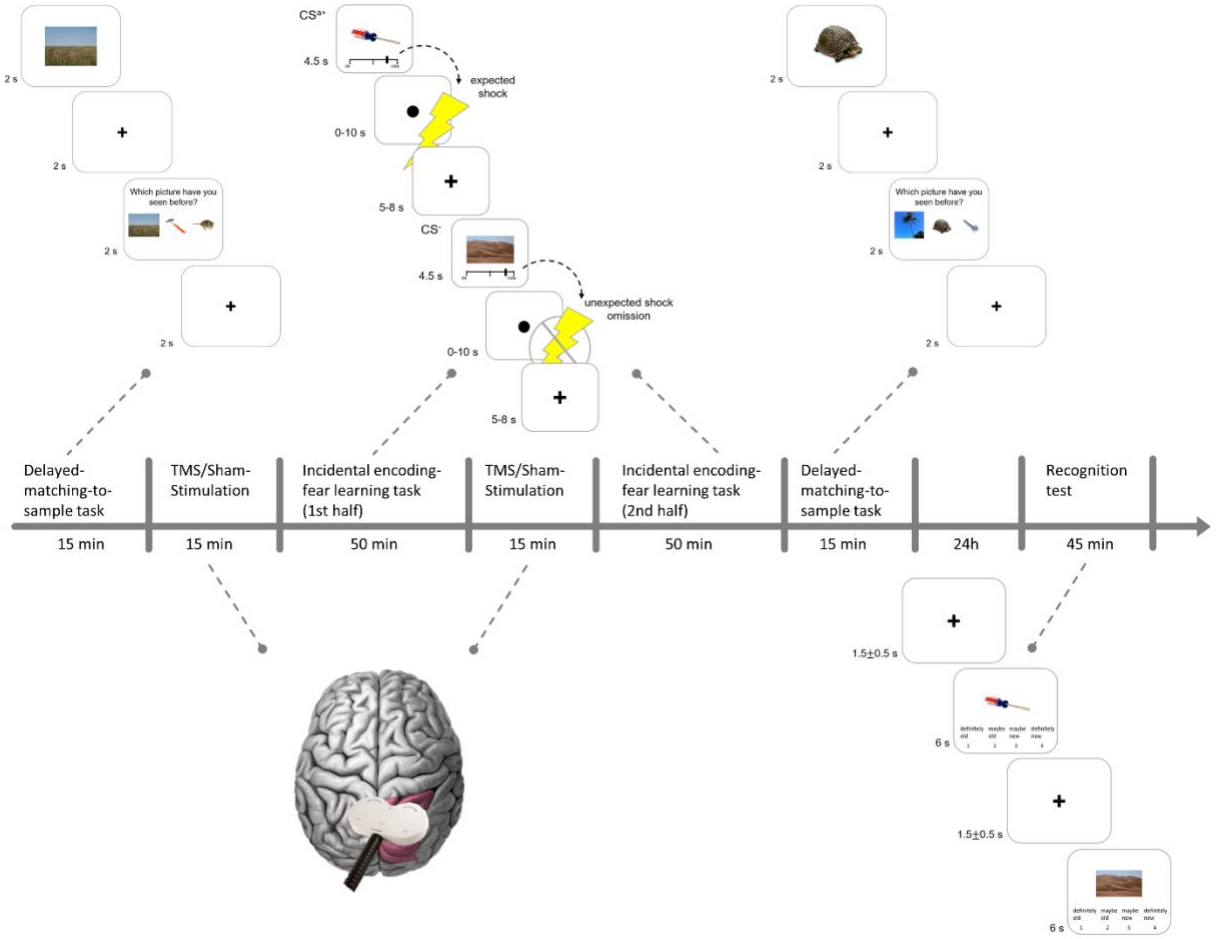
Our experimental design allowed us to capture continuous predictions which resulted in continuous PEs. Based on recent literature on PEs (Loock et al., 2025; Kalbe & Schwabe, 2022b; Rouhani et al., 2023), we calculated signed PEs (sPE) which facilitated the distinction between positive and negative PEs. The sPE were calculated as the relative difference between the binary outcome (shock vs. no-shock) and the explicit shock prediction in the respective trial resulting in a value between -1 and 1. Notably, the sPE's sign also indicated the value of the outcome: Negative sPEs (sPE < 0) could only occur in unshocked trials, i.e., unexpected shock omissions, while positive sPEs (sPE > 0) could only occur in shocked trials, i.e., unexpected shock occurrence.

5.2.6 Day 2: Recognition memory test

On experimental Day 2, approximately 22-26 hours after Day 1, participants returned for a surprise recognition test (see Figure 7). First, they completed a short questionnaire to assess whether they anticipated a memory test and then rated how surprised they were about the recognition test on a scale from 1 (*not surprised at all*) to 5 (*very surprised*). In the recognition test, participants saw all pictures they had seen during the incidental encoding-fear learning task (120 pictures of animals, 120 pictures of scenes and 120 pictures of tools) as well as 180 'new' pictures, i.e., lures (60 pictures of animals, 60 pictures of scenes and 60 pictures of tools) that had not been presented before, in a randomized order. Each trial started with a centred black fixation cross on a grey background for 1.5 ± 0.5 s, followed by an 'old' or 'new' picture presented centrally on the screen for 6 s. For each item, participants were instructed to indicate whether the currently presented picture was *definitely old*, *maybe old*, *maybe new* or *definitely new* by pressing the '1', '2', '3' or '4' button on the keyboard, respectively. Participants had to log in their response while the stimulus was presented on screen (max. 6 s).

Figure 7

Overview of the experimental procedure



Note. In the Delayed-matching-to-sample task, participants were required to keep stimuli from three different categories (animals, scenes, tools) in mind for 2 s and to select which stimulus they had seen before, which was then used for classifier training.TMS/Sham-Stimulation was applied to the right superior parietal cortex (highlighted in red). In the incidental encoding-fear learning task, participants saw a series of trial-unique pictures from three different categories (animals, scenes, tools) linked to fixed probabilities for receiving an aversive electric shock (CS^{a+}=67%, CS^{b+} =33%, and CS^{-}=0%). On each trial, participants indicated their shock expectation on a continuous scale (from 0 to 100 %). The delay with which the outcome (shock vs. no-shock) occurred after stimulus-offset varied between 0 and 10 s. In a surprise recognition test 24 hours later, participants had to indicate whether they had seen the item on the screen before while indicating their certainty (definitely old, maybe old, maybe new or definitely new). Importantly, they were presented with old items from the incidental encoding-fear learning task and unseen, new items. Critically, the TMS/Sham stimulation over the right superior parietal cortex was applied before and during the incidental encoding-fear learning task. All depicted images are licensed under Creative Commons BY-SA license.

5.3 SCR data acquisition and analysis

On experimental Day 1, we recorded SCR as a measure of arousal and conditioned fear during the incidental encoding-fear learning task. SCR was measured using a MP-160 BIOPAC data acquisition system (BIOPAC systems, Goleta, California, United States).

SCRs were analyzed using Continuous Decomposition Analysis in Ledalab Version 3.4.9 (Benedek and Kaernbach, 2010). On each trial, we derived the average phasic driver within a specified response window. First, the skin conductance signal was downsampled to 50 Hz and

optimized applying four cycles of initial values to increase the goodness of the model. For the anticipatory SCR, a response window was set from 0.5 after stimulus onset until the onset of the outcome (shock/no-shock) and could vary depending on the CS-outcome delay. Outcome-related SCR was analysed between 0.5 s and 4.5 s after outcome onset. The minimum amplitude threshold was set to 0.01 μS for both the anticipatory and the outcome-related SCR. Resulting estimates of the average phasic driver within each response window were returned in μS. Notably, these estimates are sensitive to interindividual differences because of physiological factors such as the thickness of the corneum (Figner and Murphy, 2011). We therefore standardized both the anticipatory and the outcome-related SCR by dividing the average phasic driver estimated in each trial by the maximum average phasic driver for each participant observed in every trial. During SCR analysis, we noticed that there were no significant SCRs to the unconditioned stimulus, i.e., an electric shock, which rendered the SCR data unreliable. We therefore decided to not include the SCR data in further analyses.

5.4 EEG data acquisition and analysis

5.4.1 EEG acquisition

On Day 1, EEG was recorded during each of the DMS sessions and the incidental encoding-fear learning task. Participants were seated 80 cm in front of a computer screen in an electrically-shielded and sound-isolated room. A 64-channel BioSemi ActiveTwo system (BioSemi B.V., Amsterdam, The Netherlands), following the international 10-20 system, was used to record EEG at a sampling rate of 1024 Hz. Additional electrodes were placed at the mastoids, above and below the orbital ridge of the right eye and at the outer canthi of both eyes. Electrode impedances were kept between ±30 mV. The EEG data was referenced online to the BioSemi common mode sense (CMS) - driven right leg (DRL) reference electrodes and filtered online with a band-pass filter of 0.03-100 Hz. Due to technical issues, 4 participants had to be excluded from EEG analysis resulting in a sample of $n = 118$ participants for the EEG analysis.

5.4.2 Preprocessing

Preprocessing was performed offline using the FieldTrip toolbox (Version 20200607; Oostenfeld et al., 2011) and custom scripts in Matlab (Version 2020b; TheMathWorks). Trials from the incidental encoding fear-learning task were segmented from -5 to 5s relative to outcome onset and re-referenced to the mean average of all scalp electrodes. Data were demeaned based on the average signal of the entire trial and de-trended. A discrete Fourier-Transform filter (DFT) at 50 Hz was applied to minimize power-line noise. Electrodes that did not record or showed extensive noise (max. one per participant) were removed and interpolated by weighted neighboring electrodes. Noisy trials were removed by visual inspection. On

average, 11.05 (*SD* = 5.57) of the 360 trials were removed in the incidental encoding-fear learning task, corresponding to approximately 3% of all trials. After artifact rejection, the segments were downsampled to 256 Hz. Next, we ran an extended infomax independent component analysis (ICA) using the 'runica' method with a stop criterion of weight change $<10^{-7}$ in order to identify and reject components associated with eye blinks or other sources of noise. Following a two-step procedure, we first correlated the signals from the horizontal and vertical EOG electrodes with each independent component and removed components exhibiting a correlation higher than 0.9. The remaining components were then identified through visual inspection along the time course and corresponding brain topographies. On average, 6.08 (*SD* = 2.67) components per participant were detected and removed.

5.4.3 Event-related potential (ERP) analysis

Based on previous studies on PE-related ERPs (Silvetti et al., 2014; Turan et al., 2025), we analyzed outcome (i.e., shock vs. no-shock) effects on the Feedback Related Negativity component (FRN; Talmi et al., 2013) and on the Positivity 300 component (P3; Ridderinkhof et al., 2009) in the EEG data of the incidental encoding-fear learning task on Day 1. For this ERP analysis, data was segmented into epochs from -2000 ms to 2000ms relative to outcome onset (shock vs. no-shock) and baseline-corrected by subtracting the average 2000 ms interval before outcome onset. The FRN was analysed between 0 and 800 ms after outcome onset at the fronto-central electrode channel FCz independent of the trials outcome (shock vs. no-shock) and defined as the largest negative peak between 0 and 800ms after outcome onset used for later analyses with item recognition and PE effects. The P3 was analysed between 300 and 1000ms at the posterior channel Pz independent of the trials outcome and defined as the largest positive peak between 300 and 1000ms after outcome onset used for later analyses with item recognition and PE effects.

5.4.4 Time-frequency analyses

EEG data from the incidental encoding-fear learning task was decomposed spectrally using sliding Hanning windows (2-30 Hz, 1-Hz steps, five-cycle window, interval: -5 to 5 s relative to outcome onset) averaged over all trials. This enabled us to calculate the time-frequency representations with respect to two temporal windows: pre-outcome (−3 to 0 s relative to outcome onset) and post-outcome (0 to 3 s relative to outcome onset). To obtain a more nuanced picture of what is emerging around the occurrence of a PE, we also computed shorter time windows (pre-outcome: −1 to 0 s relative to outcome onset; post-outcome: 0 to 1 s relative to outcome onset) for which we obtained a similar pattern of results. In each window, single trial power estimates were calculated trial-wise, log-transformed (Grandchamp &

Delorme, 2011; Smulders et al., 2018) and baseline corrected (absolute baseline correction −5 to −3 s relative to stimulus onset). For the whole-brain time–frequency data, spectral power averaged over all trials was tested with a dependent sample cluster-based permutation *t*-test (10.000 permutations to correct for multiple comparisons; Maris & Oostenveld, 2007). This approach allows testing for statistical differences while simultaneously controlling for multiple comparisons without spatial constraints. The samples were clustered at a level of $\alpha_{cluster} = 0.001$. Clusters with a corrected Monte Carlo *p*-value < 0.05 are reported as significant. Additionally, we entered the single power estimates in alpha (8-12 Hz) and theta bands (4-7 Hz) into models assessing item recognition dependent on PEs.

5.4.5 MVPA

Multivariate decoding analysis was performed using the MVPA-light toolbox (Treder, 2020).

5.4.5.1 Classifier training

For decoding, EEG data from the DMS sessions before and after the incidental encoding-fear learning task was pooled to reduce any time-related biases and to make sure that there is a sufficient number of trials for a reliable classifier training. In the DMS task, epochs were defined as -2000 to 2000 ms relative to the delay phase. Then, the EEG data was processed exactly as in the PE-task. On average, 7.76 (*SD* = 3.42) of the 300 trials (pooled over both sessions) were removed from the DMS task, corresponding to < 3% of all trials. During ICA, 4.97 (*SD* = 2.20) components per participant were detected and removed on average. Afterwards, the classifier was trained within-subject, utilizing a logistic regression (L2 penalized) on the preprocessed data of the DMS task (pooled over both DMS sessions) to differentiate between image categories (animals, scenes, tools). On each trial, each target category (e.g., a scene), was contrasted against the two unseen image categories (e.g., an animal and a tool). To account for class imbalances, we applied class weights that incorporated the inverse frequency of each class. All EEG channels were used as features. Prior to classification, we segmented the preprocessed data into one time window that was subject of the classifier training: The investigated window (0 to 2000 ms relative to the delay phase) included the whole delay phase where participants had to keep the target in mind (*stimulus maintenance phase)*. Here, we used a sliding window averaging 100 ms with a step size of 10 ms to identify the optimal time window for individual decoding. To evaluate the classification performance, we implemented a 5-fold cross-validation. The classifiers with the highest performance, i.e., accuracy, were used to decode the neural representations during the incidental encoding-fear learning task.

5.4.5.2 Decoding

The classifier trained during the stimulus maintenance phase per participant was used to decode neural patterns emerging before and after outcome onset in the incidental encoding-fear learning task, i.e., before and after the occurrence of PEs. Specifically, we defined two windows of decoding and segmented the preprocessed data from the incidental encoding-fear learning task accordingly: A *pre-outcome* window during the CS-outcome delay (-2000 to 0 ms relative to outcome onset) and a *post-outcome* window after the outcome was presented (0 to 2000 ms relative to outcome onset). The pattern of results remained unchanged when analyzing smaller time windows (± 1000 ms relative to outcome onset). The pre-outcome window was indicative of the neural stimulus representation (i.e., maintenance) during the CS-outcome delay, shortly before a PE occurred, while the post-outcome window indicated neural patterns that emerge after a PE has occurred (i.e., potential stimulus reactivation). In a trialwise manner, the classifier trained during stimulus maintenance in the DMS task was applied to each of the decoding windows (pre-outcome, post-outcome) utilizing an overlapping sliding window, with a time average of 100ms and a step size of 10 ms. The resulting average decoding accuracy per trial indicated the strength of the neural patterns before and after a PE, respectively, and was used for later analyses of item recognition and PE effects.

5.5 Statistical analysis

5.5.1 Behavioral analyses

Overall, item recognition was treated as the binary dependent variable, coded as '0' for misses and '1' for hits. In line with previous research on episodic memory (Bartlett et al., 1980; Gagnon et al., 2019; Heinbockel et al., 2024; Kalbe and Schwabe, 2022), our analysis focused on high-confidence responses, such that only trials in which participants indicated that they were 'very sure' were considered as hits. Such high-confidence recognitions have been linked to a hippocampus-based recollection rather than only familiarity with an item, which is assumed to depend on the perirhinal cortex (Eichenbaum et al., 2007). Accordingly, we computed hit rates (i.e., recognizing an item as "surely old") and category-based false alarm rates based on stimulus category-level (CS^{a+} vs. CS^{b+} vs. CS^{-}). Additionally, we computed the signal detection theory-based parameter *d',* computed as the difference between *z*-transformed hit rates and *z*-transformed false alarm rates, where *z* represents the inverse of the standard normal distribution. Using mixed-design ANOVAs, we also examined the influence of the between-subjects factor stimulation group (TMS vs. sham) and the within-subject factor conditioning category (CS^{a+} vs. CS^{b+} vs. CS^{-}). For pairwise comparisons, independent-samples t-tests were conducted, with Welch's correction applied when the assumption of equal variances was violated.

5.5.2 Linear and multilevel models

To analyze how PEs impacted subsequent recognition memory, we fitted generalized linear mixed models (GLMMs) with a logit link function using the lme4 R package (Bates et al., 2015) that enabled us to perform trial-wise analyses. To maximize generalizability of the GLMMs, we utilized the maximal random effects structure and treated subjects as random effects for both the intercept and all slopes of the fixed effects included in the model (Barr et al., 2013). The recognition of an individual item was treated as the binary dependent variable, coded '0' for misses and '1' for confident hits. For PEs, we derived a fine-grained measure of PEs, the sPE, ranging between -1 and 1 and allowing to differentiate between negative and positive PEs. We fitted models using different sets of independent variables, including sPEs, anticipatory and outcome-related arousal, the explicit shock prediction, the CS-outcome delay and the stimulation group including their interaction. The best fitting models were selected based on χ^2 tests.

To directly link electrophysiological data with the participant's recognition performance and PE-effects, we computed a set of separate linear mixed models (LMMs) and GLMMs at the trial level. For ERP data, we computed LMMs with the independent variable sPE, the stimulation group, CS-outcome delay and the explicit shock prediction on the FRN amplitude and on the P3 amplitude, respectively. Additionally, we also computed GLMMs that tested whether the P3 amplitude and the FRN amplitude, respectively, predicted item recognition.
For time-frequency data, we fitted LMMs to investigate whether the average spectral power interacted with the PE in the pre-outcome and post-outcome window to affect item recognition. For each window, we computed a set of GLMMs that included the average spectral power, sPE, CS-outcome delay and stimulation group including their interactions to predict item recognition. First, we investigated pre-outcome and post-outcome windows that lasted 3 s.

To analyze if and how the neural stimulus representations are associated with PE-induced memory enhancements, we also computed (G)LMMs in which we included decoding accuracy from the MVPA. To examine to what extent PE effects on subsequent memory require a neural stimulus (category) representation at the time of the outcome (i.e., shortly before a PE occurred), we computed a GLMM with item recognition being predicted by sPE and decoding accuracy (and stimulation group) in the pre-outcome window. In order to assess whether PE magnitude is associated with stimulus (category) reactivation, we set up an LMM in which we predicted the post-outcome decoding accuracy by the sPE and the stimulation group. Additionally, we fitted a GLMM with the post-outcome decoding accuracy, sPE and stimulation

group as independent variables to predict the binary variable item recognition and to further elucidate the mechanisms underlying PE-induced memory enhancement.

All analyses were performed in R Studio (Version 1.2.5033, RStudio Team (2020), PBC, Boston, MA, USA), unless indicated otherwise above, and subjected at a significance level of $\alpha = .05$ and reported *p*-values are two-tailed. In case of sphericity violation, indicated by Mauchly's test, Greenhouse Geisser corrected degrees of freedom and *p*-values are reported. Post-hoc tests following significant main or interaction effects were Bonferroni-corrected for multiple comparisons, if required. Effect sizes were either reported as Cohen's *d* for *t*-values for between-subjects analyses, Cohen's d_z for within-subjects analyses or partial eta squared for *F*-values.

5.6 Transparency and openness

All data, materials, and scripts have been made publicly available and can be accessed at: https://doi.org/10.25592/uhhfdm.17016. This study was pre-registered at the German Clinical Trials Register (DRKS-ID: DRKS00030529; https://drks.de/search/en/trial/DRKS00030529).

5.7 Results

5.7.1 Successful fear learning

PEs were overall equally distributed around zero indicating that a sufficient number of positive and negative PEs could be analyzed (Figure 8A). Participants' explicit shock predictions (ranging from 0 to 100 %) showed that they learned the shock contingencies very well (Figure 8B). For CS^{a+} ($M = 0.68$, $SD = 0.16$), the shock expectancy was significantly higher compared to CS^{b+} ($M = 0.47$, $SD = 0.15$; $t(121) = 9.93$, $p < .001$, $d = 1.36$) and for CS^{b+} compared to CS^{-} ($M = 0.06$, $SD = 0.13$; $t(121) = 23.33$, $p < .001$, $d = 3.22$). Importantly, shock predictions did not differ between stimulation groups ($F(1,120) = 0.06$, $p = .806$, partial $\eta^2 = 0.00$).

5.7.2 General memory performance

Overall, participants performed well in the surprise recognition test on Day 2, as reflected in significantly higher hit rates ($M = 0.59$, $SD = 0.33$) than false alarm rates ($M = 0.41$, $SD = 0.37$; $t(121) = 16.16$, $p < .001$, $d = 0.50$). Hit rates differed significantly between CS conditions ($F(2,242) = 5.63$, $p = .004$, partial $\eta^2 = 0.005$), as did false alarm rates ($F(2,242) = 6.47$, $p = .002$, partial $\eta^2 = 0.002$). Post-hoc comparisons revealed that the average hit rate for CS^{a+} items ($M = 0.62$, $SD = 0.13$) was significantly higher than for CS^{b+} items ($M = 0.58$, $SD = 0.12$; $t(121) = 2.71$, $p = .007$, $d = 0.11$) and CS^{-} items ($M = 0.56$, $SD = 0.15$; $t(121) = 2.91$, $p = .004$, $d = 0.16$), while the hit rates for CS^{b+} and CS^{-} items did not differ ($t(121) = 0.93$, $p = .355$, $d = 0.05$). For the false alarm rate, there was a comparable pattern with the average false alarm rate being significantly increased for CS^{a+} ($M = 0.42$, $SD = 0.09$) items compared to CS^{b+} items

(M = 0.39, SD = 0.09; $t(121) = 2.48$, $p = .014$, $d = 0.08$) but not to CS^- items (M = 0.43, SD = 0.08; $t(121) = -0.66$, $p = .509$, $d = 0.02$). Notably, the false alarm rate was significantly lower for CS^{b+} items compared to CS^- items ($t(121) = -3.65$, $p < .001$, $d = 0.10$). When using the signal detection theory-based parameter d', recognition memory was higher for both CS^{a+} items (M = 0.90, SD = 0.37) and CS^{b+} items (M = 0.84, SD = 0.38) compared to CS^- items (M = 0.65, SD = 0.42; vs. CS^{a+}: $t(121) = 4.81$, $p < .001$, $d = 0.51$; vs. CS^{b+}: for $t(121) = 3.57$, $p = .001$, $d = 0.39$; main effect CS category: $F(2,242) = 13.24$, $p < .001$, partial $\eta^2 = 0.041$), while there was no reliable difference between CS^{a+} and CS^{b+} items ($t(121) = 1.24$, $p = .218$, $d = 0.10$; Figure 8C).

CTBS over the superior parietal cortex had a significant impact on overall recognition memory performance. For d', our analyses revealed significantly increased memory performance for participants of the TMS group (M = 0.87, SD = 0.53) compared to those of the sham group (M = 0.72, SD = 0.41; $t(116.2) = 2.17$, $p = .032$, $d = 0.39$; Figure 8C). To further elucidate where this difference in memory sensitivity is coming from, we analyzed the hit rates and false alarm rates for each experimental group. Participants who received cTBS immediately before the encoding session had a significantly lower hit rate (M = 0.38, SD = 0.20) but also a significantly lower false alarm rate (M = 0.17, SD = 0.22) compared to the participants of the sham group (hit rate: M = 0.81, SD = 0.09, $t(115.0) = -9.57$, $p < .001$, $d = 1.74$; false alarm rate: M = 0.67, SD = 0.33, $t(99.3) = -10.13$, $p < .001$, $d = 1.85$), suggesting that cTBS over the superior parietal cortex led to more conservative mnemonic responses.

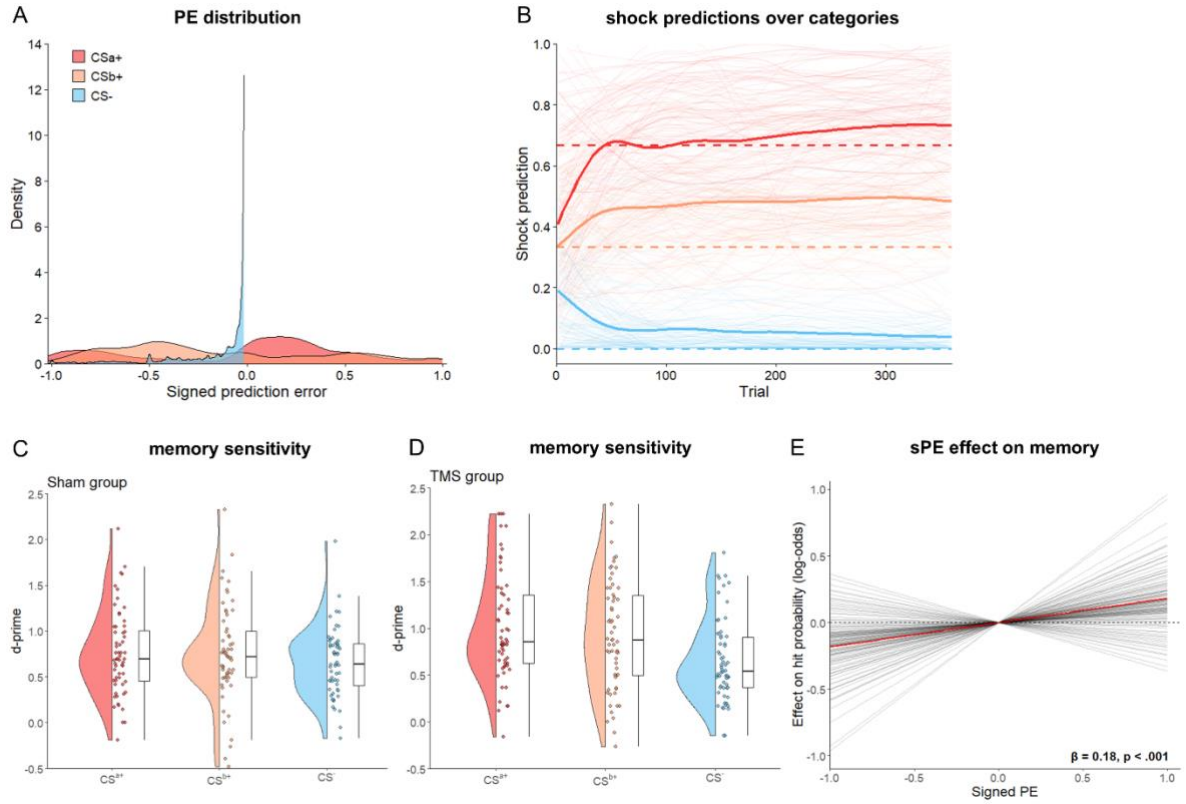
5.7.3 Signed PEs enhance memory for preceding stimuli

To investigate the influence of PEs on memory formation, we fitted GLMMs with recognition of an item as the binary dependent variable and added relevant independent predictors in a step-wise manner.

We started with a minimal model, in which we tested whether trial-wise sPEs contribute to item recognition. We treated the sPE (ranging from -1 to 1) after a CS item, i.e., the predictive stimulus, as the sole independent variable to predict subsequent item recognition. Estimates obtained revealed that sPEs, $z = 2.44$, $p = .015$, $\beta = 0.09$, showed the expected positive relationship with subsequent item recognition. To rule out that the sPE-effects were confounded with the explicit shock prediction or the outcome (shock vs. no-shock), we computed a model in which we added the explicit shock prediction and the outcome, respectively, as additional predictors. Importantly, the memory enhancing effect of the subsequent sPEs remained significant (all $z > 3.95$, $p < .001$, $\beta > 0.18$; Figure 8E) after adding the prediction or the outcome, respectively.

Figure 8

PEs, shock expectations, and memory performance



Note. (A) PEs for CS^{a+} and CS^{b+} were equally distributed around zero. (B) Participants' mean shock predictions (thick lines) approached the underlying shock probabilities (dotted lines) relatively quickly confirming successful fear learning. (C, D) Memory sensitivity, computed as *d*-prime (computed as the difference between *z*-transformed hit rates and *z*-transformed false alarm rates) in the sham group (C) was significantly lower than in the TMS group (D). Dots show data from individual participants. (E) sPEs significantly boosted memory for the preceding stimulus.

To investigate whether sPE effects on memory are dependent on the CS-outcome delay, we set up a model in which we included the sPE, the explicit shock prediction, the CS-outcome delay and the sPE×CS-outcome delay interaction to predict the binary item reocgnition. This model revealed that memory was not influenced by the CS-outcome delay ($z = 0.46$, $p = .648$, $\beta = 0.00$) nor by the sPE×CS-outcome delay ($z = 0.20$, $p = .841$, $\beta = 0.00$) but significantly enhanced by the sPE ($z = 2.95$, $p = .003$, $\beta = 0.19$). These findings suggest that the memory enhancing effect of sPEs is not affected by the delay between CS and outcome, in line with previous findings (Loock et al., 2025).

Next, we set up a model that incoporated the stimulation group (TMS vs. sham), sPE, their interaction and the shock prediction to examine whether cTBS over the superior parietal cortex modulated the sPE effects on memory. Again, we found a memory enhancing effect of sPEs, $z = 3.42$, $p = .001$, $\beta = 0.22$. As expected, the stimulation group also affected item memory ($z = -10.18$, $p < .001$, $\beta = -2.75$) suggesting that memory, expressed as hits, decreased when

cTBS was applied. However, as the above analysis of overall memory performance showed, this effect was most likely due to a more conservative response bias in the TMS group. Importantly, the sPE×stimulation group interaction was not significant (z = -0.64, p = .525, β = -0.05) suggesting that the sPE effect on memory was not modulated by cTBS over the superior parietal cortex.

5.7.4 PEs trigger neural stimulus category reactivation

In line with previous findings (Loock et al., 2025), our behavioral data showed that the sPE enhance subsequent memory for stimuli preceding the PE, irrespective of the time interval between the predictive stimulus and the outcome (i.e., PE). CTBS over the superior parietal cortex appeared to not modulate the sPE effect on memory. In a next step, we investigated the neural responses elicited by a PE.

5.7.4.1 Event related potentials

First, we analyzed electrophysiological responses to the outcome, i.e., shock vs. no-shock, reflected in the FRN and P3. Therefore, we computed LMMs that included the sPE, the stimulation group, CS-outcome delay, the sPE×CS-outcome delay×stimulation group interaction and the explicit shock prediction to predict the post-outcome FRN amplitude and P3 amplitude, respectively. For FRN amplitudes (0-800 ms after outcome onset at Fz), we found a significant effect of sPEs (t(1298) = 2.68, p = .007 , β = 3.53), while there were no significant effects of the stimulation group (t(119.6) = -0.97, p = .333 , β = -1.66), the CS-outcome delay (t(337.5) = -1.83, p = .068, β = -0.20), or the sPE×CS-outcome delay×stimulation group interaction (t(40264.6) = 1.05, p = .293 , β = 0.31). As expected and in line with previous research showing a role of the FRN in error processing (Bellebaum & Daum, 2008; Holroyd & Coles, 2002), these findings show that the FRN is scaled by sPE magnitude, i.e., more pronounced with increasing sPEs, irrespective of CS-outcome delay or stimulation group. For P3 amplitudes (300-1000 ms after outcome onset at Pz), we found neither a significant effect of the sPE (t(2406.4) = 1.21, p = .226, β = 6.35), nor of any other predictor (all p > .222) suggesting that none of them affected the P3 significantly.

5.7.4.2 Time-frequency analyses

In a next step, we assessed whole-brain time-frequency patterns data after an outcome was revealed. We obtained a significant positive cluster emerging at approximately 0.5 until 2 sec after outcome onset at parieto-occipital electrodes in alpha and beta bands (9-18 Hz; electrode: PO8; p < .001, ci-range < 0.01, SD < 0.01). Additionally, we found a significant negative cluster emerging at approximately 1.15 until 3 sec after outcome onset at fronto-central electrodes in theta and alpha bands (7-13 Hz; electrode: C1; p < .001, ci-range < 0.01, SD <

0.01) suggesting that there was an early alpha synchronization and a late theta desynchronization after outcome onset (see Figure 9A).

Next, we investigated whether the oscillatory power in the alpha (8-12 Hz) and theta bands (4-7 Hz), respectively, in a 3s-post-outcome window scaled with PE magnitude. To this end, we computed LMMs that treated the sPE and the CS-outcome delay including their interaction as independent variables to predict oscillatory power in the alpha and theta band separately after the outcome was presented. For alpha, this analysis showed no significant effects of sPEs ($t($20055.8$)$ = 0.30, p = .766, $ß$ = 0.00), nor of CS-outcome delay ($t($118.6$)$ = -0.34, p = .732, $ß$ = 0.00) and no sPE×CS-outcome delay interaction ($t($36776.7$)$ = 0.10, p = .922, $ß$ = 0.00). Likewise, for theta, we obtained no significant effects of sPEs ($t($8408.4$)$ = 0.47, p = .636, $ß$ = 0.00), nor of CS-outcome delay ($t($138.4$)$ = 0.34, p = .738, $ß$ = 0.00) and sPE×CS-outcome delay interaction ($t($37119.8$)$ = -0.43, p = .666, $ß$ = 0.00). These results indicate that the oscillatory post-outcome power in the alpha and theta band seems to be unaffected by the sPE and by the interval between predictive stimulus and outcome.

5.7.4.3 Decoding of category representations

In a next step, we leveraged an EEG-based decoding approach to investigate the neural patterns after the occurrence of a PE, i.e., potential stimulus (category) reactivation. As expected, the average performance of participants in the DMS task (i.e., during classifier training) was very high (M = 95.14% correct, SD = 0.12) and did not reliably differ between stimulation groups (t(70.7)= 1.84, p = .069, d = 0.38). Overall, the decoding accuracy during the *post-outcome* window (averaged over participants' individual peak accuracies) was significantly above chance (M = 0.55, SD = 0.03, t(117) = 23.04, p < .001, d = 4.26; see 8C) and significantly higher in the TMS group (M = 0.56, SD = 0.03) than in the sham group (M = 0.55, SD = 0.02; t(112) = 3.15, p = .002, d = 0.58).

Next, we investigated whether the PE magnitude was associated with stimulus (category) reactivation after the outcome by setting up a LMM that included the sPE to predict the decoding accuracy in the the post-outcome window. Our analyis revealed a significant sPE effect on decoding accuracy (t(117.1)= -2.43, p = .017, $ß$ = -0.02). Interestingly, decoding accuracy was significantly higher following negative PEs (i.e., unexpected shock omissions, M = 0.57, SD = 0.63) compared to positive PEs (i.e., unexpected shock presentations, M = 0.56, SD = 0.64; z = 2.92, p = .004 , $ß$ = 0.02; Figure 9C). This result indicates that post-outcome stimulus (category) reactivation was mainly driven by negative sPE.

To investigate whether cTBS affected the PE-related stimulus (category) reactivation we set up a LMM that included the sPE×stimulation group interaction. Importantly, there was

no significant sPE×stimulation group interaction ($t(115.8) = -1.48$, $p = .143$, $\beta = -0.02$) on post-outcome decoding accuracy, suggesting that the sPE effect on post-outcome neural stimulus (category) reactivation seems to be unaffected by cTBS over the superior parietal lobe.

5.7.5 PE-evoked theta power boost predicts item recognition after negative PEs

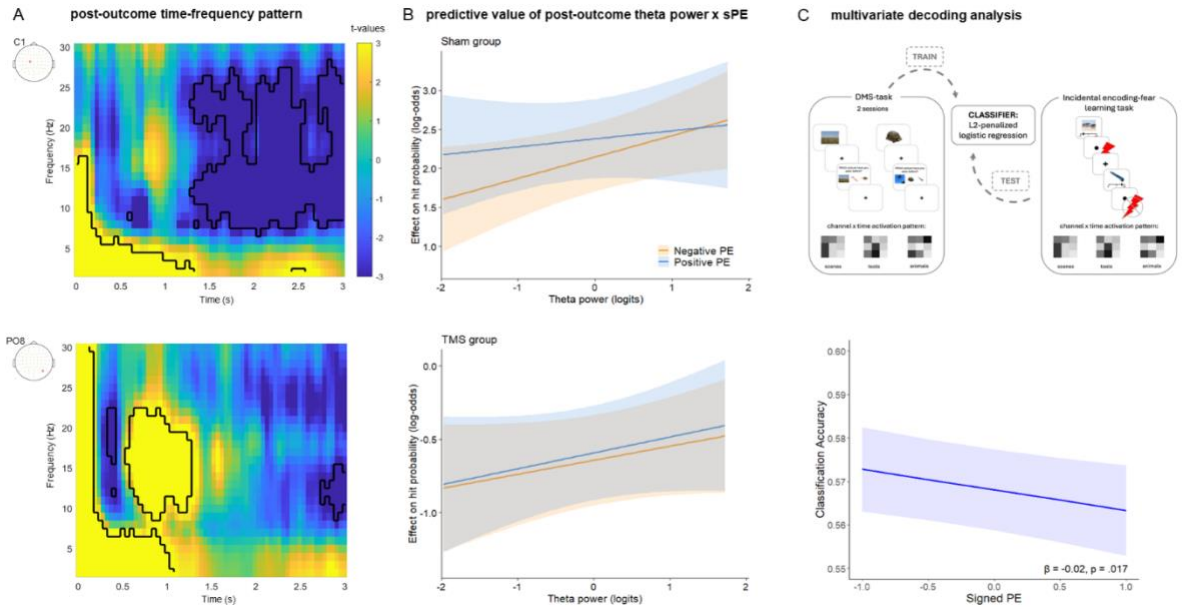
In order to test whether PE-evoked neural changes predicted subsequent memory, we first computed a GLMMs that tested whether the FRN amplitude, in response to a PE, or the FRN amplitude × sPE interaction predicted the binary item recognition. However, we found no significant effect of the FRN amplitude, $z = 1.22$, $p = .224$, $\beta = 0.00$, nor a significant FRN amplitude × sPE interaction, $z = -1.16$, $p = .248$, $\beta = -0.00$, on subsequent item memory.

Then, we set up GLMMs that included the post-outcome oscillatory spectral power in the alpha and theta bands respectively, sPE and stimulation group including their interaction to predict item recognition. When considering alpha power, we obtained no significant sPE×average spectral power×stimulation group interaction ($z = 1.25$, $p = .212$, $\beta = 0.30$) on subsequent memory. Interestingly, when considering theta power in a separate GLMM, we found a significant sPE×average spectral power×stimulation group interaction ($z = 2.96$, $p = .003$, $\beta = 0.68$) on memory. We pursued this effect with follow-up GLMMs for the sham and TMS group separately. In the sham group, we found a significant sPE×theta band power interaction ($z = -2.65$, $p = .008$, $\beta = -0.51$), whereas there was no such effect in the TMS group ($z = 1.33$, $p = .185$, $\beta = 0.17$; see Figure 9B). Although theta power was unaffected by the PE per se, our analysis showed that, in the sham group, the relationship between the post-outcome theta power and subsequent recognition memory appeared to depend on the nature of the PE. Specifically, following negative PEs, increased theta power was significantly associated with enhanced item recognition ($z = 2.12$, $p = .029$, $\beta = 0.30$), whereas for positive PEs, theta power was not significantly associated with item recognition ($z = 0.43$, $p = .670$, $\beta = 0.06$). CTBS over the superior parietal cortex appeared to abolish this association.

Next, we investigated whether the PE-induced changes in the stimulus (category) reactivation in the post-outcome window, as assessed by MVPA-based decoding, predicted subsequent item recognition. A minimal GLMM in which we included post-outcome decoding accuracy to predict item recognition revealed no significant effect of decoding accuracy ($z = 0.81$, $p = .416$, $\beta = 0.03$). In a follow-up GLMM we included the sPE, stimulation group and the sPE×decoding accuracy×stimulation group interaction as additional predictors in the previous model. Our analyses showed no significant interaction effects of sPE×decoding accuracy ($z = 0.07$, $p = .942$, $\beta = 0.01$) and sPE×decoding accuracy×stimulation group ($z = -0.38$, $p = .701$, $\beta = -0.05$) on item recognition.

Figure 9

Post-outcome changes, PEs, and memory



Note. (A) Averaged time-frequency representations of channels C1 (upper panel) and PO8 (lower panel) showing significant clusters as revealed by cluster-based permutation tests (outlined in black). (B) Signed PEs and theta power interactively predicted memory performance in the sham group (upper panel), but not in the TMS group (lower panel). (C) Trialwise category pattern reactivation computed by multivariate pattern analysis of EEG-data. An L2-penalized logistic classifier was trained to classify between neural patterns of three categories from the DMS-task (scenes, animals, tools) and tested on the same categories in the incidental encoding-fear learning task (upper panel). Overall classification accuracy was higher following negative PEs compared to positive PEs (lower panel).

5.7.6 PE effect on subsequent memory depends on the neural state shortly before the PE

In a next step, we tested whether PE effects on subsequent memory depend on the neural state and potential stimulus category representation shortly before the PE.

5.7.6.1 Time-frequency analyses

First, we investigated whether the average spectral power in the 3 s before the outcome interacted with the sPE on subsequent memory. Critically, in this analysis, we disregarded all trials in which the analyzed window could have overlapped with the presentation of the predictive stimulus to ensure temporal separation, i.e., excluding all trials in which the CS-outcome delay was shorter than 3s. Notably, when including all trials, irrespective of the CS-outcome delay duration, we observed a very similar pattern of results.

We set up GLMMs that included pre-outcome alpha power (8-12 Hz) and theta power (4-7 Hz) respectively, sPE and stimulation group including their interaction to predict item recognition. When including alpha power, our analysis yielded a significant sPE×alpha power×stimulation group interaction (z =2.55, p = .011, β = 0.84) on memory. We pursued this effect with separate follow-up GLMMs for the sham and TMS group, respectively. In the sham group, we found a significant sPE×alpha power interaction (z = -2.95, p = .003, β = -0.85),

while there was no such interaction effect in the TMS group (z = 0.06, p = .949, β = 0.01; see Figure 10). Similarly, when including theta power, we obtained a significant sPE×theta power×stimulation group interaction (z =3.18, p = .001, β = 1.03) on memory. Again, we pursued this effect with separate follow-up GLMMs for the sham and TMS group, respectively. In the sham group, we found a significant sPE×theta power interaction (z = -2.92, p = .004, β = -0.79), which was not present in the in the TMS group (z = 1.45, p = .148, β = 0.26; Figure 10). These findings suggest that the impact of sPE on item recognition was modulated by the pre-outcome neural oscillatory activity in the sham group. Specifically, before negative PEs, increased theta and alpha power were significantly associated with enhanced item recognition (theta: z = 2.82, p = .005, β = 0.45; alpha: z = 2.30, p = .021, β = 0.39), whereas for positive PEs, theta power and alpha power (theta: z = 0.46, p = .645, β = 0.08; alpha: z = -0.94, p = .345, β = -0.17) were not significantly associated with item recognition. CTBS over the superior parietal cortex appeared to abolish these associations.

5.7.6.2 Decoding of stimulus category reactivation

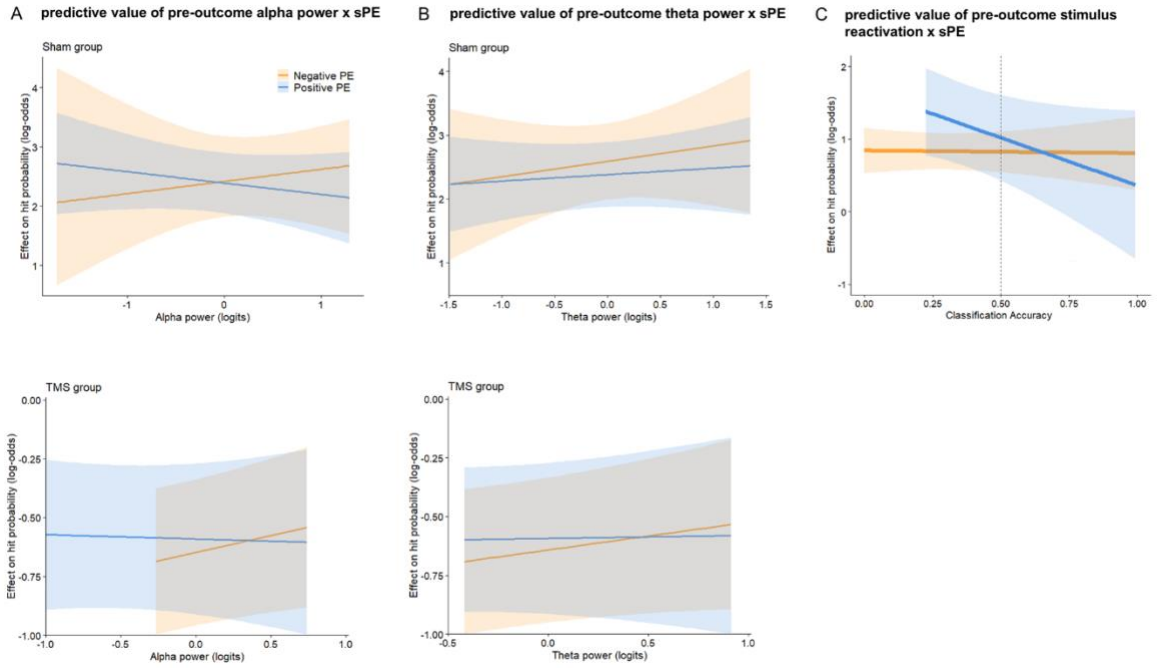
Next, we leveraged an EEG-based decoding approach again to investigate the neural patterns shortly before the occurrence of a PE, i.e., potential stimulus (category) representation. Same as in the time frequency analysis, we disregarded all trials in which the CS-outcome delay was shorter than 2s to make sure that the analyzed window did not overlap with the stimulus presentation. Overall, the decoding accuracy during the *pre-outcome* window (averaged over participants' individual peak accuracies) was significantly above chance (M = 0.55, SD = 0.03, t(117) = 22.84, p < .001, d = 4.22) and significantly higher in the TMS group (M = 0.57, SD = 0.03) than in the sham group (M = 0.55, SD = 0.03; t(115.7) = 4.36, p < .001, d = 0.80). A LMM including the stimulation group and CS-outcome delay to predict pre-outcome decoding accuracy on trial level, confirmed that there was a significant effect of stimulation group (t(116.4) = 3.77, p < .001, β = 0.16, d = 0.70) but no significant effect of CS-outcome delay (t(33663.6) = 1.15, p = .251, β = 0.00, d = 0.01).

Next, we set up a GLMM in which we included the sPE and the sPE×decoding accuracy interaction to predict subsequent item recognition. This analysis yielded a significant effect of sPEs (z =3.31, p = .001, β = 0.16) and a significant sPE×decoding accuracy interaction effect (z = -2.10, p = .035, β = -0.14) on item memory suggesting that the PE effect on memory performance depends on the strength of the stimulus (category) maintenance shortly before the PE (see Figure 10C). For negative PEs, pre-outcome decoding accuracy was not signficantly associated with subsequent item recognition (z = 1.24, p = .216, β = 0.06), whereas for positive PEs, increased decoding accuracy was associated with decreased memory performance (z = -

2.59, p = .010, β = -0.13). These findings suggest that stronger stimulus (category) representation shortly before the PE seems to impair memory encoding in the context of positive PEs, whereas this effect is absent for negative PEs. A follow-up model incorporating stimulation group yielded that the PE×decoding accuracy was not affected by cTBS over the right superior parietal cortex, i.e., a non-significant effect of stimulation group×decoding accuracy (z =0.01, p = .938, β = 0.01) and a non-significant sPE×decoding accuracy×stimulation group interaction on item recognition (z =-0.66, p = .511, β = -0.10).

Figure 10

Pre-outcome changes, PEs and memory



Note. (A) Signed PEs and alpha power interactively predicted item memory in the sham group (upper panel), but not in the TMS group (lower panel). Shaded areas represent 95%-confidence intervals. (B) A similar pattern was obtained for theta power, with a significant sPE×theta power interaction on item memory in the sham group (upper panel) but not in the TMS group (lower panel). (C) Stronger stimulus (category) representations above chance level (dotted line) shortly before the PE impair memory in the context of positive PEs.

5.7.7 Control variables

Importantly, participants were not aware of their actual stimulation condition (sham vs. TMS) as assessed by a treatment guess at the end of the experiment ($\chi^2(1) = 0.278$; p= .599). Moreover, participants were moderately surprised by the recognition test on Day 2 (M = 3.13, SD = 1.27). Participants who underwent the cTBS appeared to be less surprised by the recognition test (M = 2.42, SD = 1.18) than participants of the sham group (M = 3.83, SD = 0.92; $t(109.88)$= 7.26, $p < .001$, d = 1.39). Seventeen participants chose the 'not surprised at all'

option. Because excluding them did not affect the results, we included them in all analyses. Additionally, we included participants suprise into the model to analyse overall memory performance, i.e., *d'*, but the group difference remained ($F(1,115) = 5.87$, $p = .017$, partial $\eta^2 = 0.049$), suggesting that it cannot be explained by the difference in the surprise related to the recognition test.

State and trait anxiety levels were comparable between stimulation groups (STAI-T: $t(118) = 0.08$, $p = .935$, $d = 0.02$; STAI-S: $t(118) = 0.95$, $p = .346$, $d = 0.18$). Notably, there were significant differences between groups in depressive mood (BDI-II: $t(118) = 6.12$, $p < .001$, $d = 1.13$), chronic stress (TICS: $t(114) = 5.41$, $p < .001$, $d = 1.01$) and sleep quality *(*$t(86) = 5.55$, $p < .001$, $d = 1.20$) suggesting that those characteristics were increased in the TMS group before cTBS was applied. Importantly, adding those measures as covariates in our analyses did not change the pattern of results, suggesting that these variables could not explain the group differences we obtained.

5.8 Conclusion

Recent evidence indicates that PEs play a pivotal role in memory formation (Kalbe & Schwabe, 2020, 2022b; Loock et al., 2025; Rouhani et al., 2023). While these effects are essential for understanding adaptive memory processes, the underlying mechanisms of these effects remain poorly understood. Here, we employed EEG and MVPA to investigate the neural dynamics surrounding PEs in the context of memory. Our findings demonstrate that the impact of the PEs on adaptive memory formation is driven by the neural state immediately before the PE as well as neural changes evoked by the PE.

Our results replicate the previously reported beneficial effect of PEs on episodic memory for preceding inherently neutral stimuli. Previous studies have often examined unsigned PEs (Kalbe & Schwabe, 2020), which reflect the overall PE magnitude regardless of its direction. However, more recent evidence suggests that positive and negative PEs differentially affect memory formation (Kalbe & Schwabe, 2022b; Loock et al., 2025; Rouhani & Niv, 2021). Therefore, we focused on sPEs on a continuous scale (ranging from -1 to +1) to distinguish between positive and negative PEs. Our behavioral data corroborate previous findings (Loock et al., 2025) showing that negative PEs were associated with impaired subsequent memory, whereas positive PEs were linked to memory enhancement. While initial fMRI evidence suggests that PE-induced effects on memory are associated with decreased activation of the medial temporal lobe and enhanced interaction between the salience and a frontoparietal network (Kalbe & Schwabe, 2022b), we focused on neural processes surrounding a PE event. We predicted that the sPE effects on memory may require a neural representation

of the predictive stimulus shortly before the PE event. Remarkably, our findings provide evidence that the sPE effect on memory depends on the strength of the stimulus (category) representation shortly before the PE. Interestingly, our findings show for positive PEs, i.e., unexpected shock presentations, that increased stimulus category representation before the PE was associated with impaired memory performance. This result could suggest that strong reactivation of the predictive stimulus category increases the interference with the attention-grabbing unexpected shock occurrence, thereby impairing memory storage. In contrast, negative PEs, i.e., unexpected shock omissions, did not show this effect, presumably because they are less emotionally salient than positive PEs.

Beyond the neural stimulus category reactivation, we also analyzed the impact of neural oscillations before the PE event on PE-related memory effects. Interestingly, both alpha and theta activity before the PE were associated with sPE-induced memory changes (in the sham group). Again, these effects were dependent on the sign of the PE. For negative PEs, increased alpha and theta activity before the PE event were linked to improved subsequent memory, potentially due to enhanced attentional processing and associative binding (Payne & Sekuler, 2014; Staudigl & Hanslmayr, 2013). For positive PEs, however, increased alpha and theta were associated with reduced subsequent memory, which could be, same as the neural stimulus category reactivation, related to increased interference and attentional competition.

In addition to the relevance of the neural state shortly before the PE event, we proposed an alternative mechanism suggesting that the PE may induce neural changes that facilitate the memory storage of preceding stimuli. As expected, sPEs led to an increased FRN, consistent with findings linking the FRN with error processing (Bellebaum & Daum, 2008; Holroyd & Coles, 2002). Intriguingly, sPEs were also associated with increased stimulus category reactivation. Specifically, we found increased stimulus category reactivation after negative PEs compared to positive PEs suggesting that negative PEs elicit stronger stimulus (category) reactivations after the outcome, presumably reflecting an adaptive mechanism for updating predictive models in response to an unexpected safety signal, i.e., an unexpected shock omission. This aligns with research demonstrating that unexpected omissions of aversive events engage mnemonic processes related to model updating and learning (Iglesias et al., 2013), also suggesting that the increased stimulus reactivation following negative PEs may reflect increased attentional processing due to the unexpected relief from an anticipated aversive event (Li et al., 2011; Kalbe & Schwabe, 2022b). In contrast, positive PEs, representing an unexpected aversive outcome, presumably impair stimulus (category) reactivation due to a more pronounced processing of the unexpected shock. This might lead to arousal-biased

competition of attention where attentional resources are preferentially allocated to the aversive event itself rather than to the preceding predictive stimulus (Kalbe & Schwabe, 2022b; Mather & Sutherland, 2011). Surprisingly, stimulus (category) reactivation after a PE was not related to item recognition suggesting that memory memory formation does not depend on reactivation after a PE. Stimulus (category) reactivation may primarily support model updating rather than direct memory consolidation (Iglesias et al., 2013). Particularly in the context of aversive learning, neural responses after outcome representation may reflect an adaptive updating of future expectations rather than strengthening individual item memory (Schwiedrzik & Freiwald, 2017). Furthermore, physiological arousal induced by the PE may interact with stimulus (category) reactivation, such that the relevance of stimulus information at this stage depends on whether attentional resources are directed toward updating predictive models rather than encoding specific stimulus details (Mather & Sutherland, 2011).

Although theta and alpha power were not directly modulated by sPEs, our findings indicate an interplay of sPEs and post-PE theta power on recognition, depending on the sign of the PE. Following negative PEs, theta oscillations were associated with better item recognition implying that theta waves may support memory consolidation when an expected aversive event does not occur. This could reflect enhanced memory storage under conditions of surprise by attenuating processes in the default mode network which may otherwise impair memory formation (Klimesch, 1999; Kota et al., 2020; White et al., 2013), which might be due to theta oscillation-induced binding of an item to its spatiotemporal context in the medial temporal lobe and in hippocampo-cortical feedback loops (Klimesch, 1999; Hanslmayr et al., 2011). However, for positive PEs, there was no such relationship indicating that theta oscillations may be more pronounced and relevant for encoding unexpected safety signals rather than unexpected threat.

We also investigated whether and how the effects of PEs on memory and their underlying neural mechanisms could be modulated using brain stimulation. Here, we focused on the superior parietal cortex given its critical role in working memory processes and top-down attentional updating (Corbetta et al., 1995; D'Esposito & Postle, 2015; Koenigs et al., 2009). We applied cTBS over the right superior parietal cortex to modulate PE effects on memory. Overall, we found an increase in memory performance in the cTBS group primarily driven by a more conservative response bias. This finding resonates with the observed increase in neural (category) representation and reactivation of the predictive stimulus following cTBS. Inhibiting the superior parietal cortex may have reduced competing attentional processes, thereby facilitating more targeted reactivation of relevant memory traces. Given its role in top-

down attentional control (D'Esposito & Postle, 2015), an inhibition might have limited the influence of irrelevant information during retrieval, resulting in a more coherent and robust reactivation of task-relevant stimulus representations, i.e., stronger decoding signals. This assumption is supported by findings that cortical reinstatement enhances memory retrieval by reactivating content-specific encoding patterns (Gordon et al., 2014). Importantly, the behavioral PE effect on memory was not significantly altered by cTBS suggesting that intact functioning of the superior parietal cortex region is not essential for the PE-effects on general memory performance. Instead, the beneficial effect of PEs on memory might derive from more distributed neural mechanisms, potentially involving medial temporal and prefrontal regions (Kalbe & Schwabe, 2022b) involved in the detection of expectancy violations and the prioritization of salient event information for long-term storage. However, at the neural level, the mechanisms underlying the PE effect on memory were altered by cTBS. Specifically, the neural signatures of PE-related memory effects, which were evident in the sham group, were diminished in the cTBS group. The superior parietal cortex is assumed to be critical for pre-outcome neural states that then shape PE-related memory effects, presumably by facilitating attentional and predictive processes (Cabeza et al., 2008). However, the absence of this effect in the TMS group suggests that cTBS may have disrupted the anticipatory modulation of encoding processes by pre-outcome alpha and theta activity, thereby weakening the interplay between preparatory neural states and PEs.

In summary, we demonstrate that the effects of sPE on memory formation for preceding events are influenced by neural states and representations surrounding the PE. Specifically, the impact of PEs on memory depends on theta and alpha oscillations shortly before the PE occurs, which may provide attentional and mnemonic binding ressources required for PE-induced modulation of memory formation. More generally, these results contribute to our understanding of the mechanisms underlying adaptive memory, where stimuli predicting emotionally relevant events are particularly well stored in long-term memory.

6 General discussion

Rather than being just a signal for learning, surprise is a gateway for restructuring memory (Antony et al., 2023; Frank et al., 2022; Sinclair & Barense, 2018; Sinclair et al., 2021). Across moments of expectancy violations, the brain not only adapts future predictions (Henson & Gagnepain, 2010; Rescorla & Wagner, 1972; Schultz et al., 1997; Shohamy & Adcock, 2010) but also selects which elements of an experience are stored in long-term memory or forgotten subsequently (Kalbe et al., 2020; Nairne & Pandeirada, 2008; Rouhani et al., 2023). This thesis set out to explore this process by investigating how PEs related to aversive outcomes affect episodic memory formation. Specifically, their underlying cognitive and neural mechanisms have remained elusive although these effects are essential for understanding adaptive memory processes. The current thesis aimed to address this gap by systematically investigating the selectivity, the temporal dynamics and, at its heart, the neural underpinnings of PE-related memory modulations across a series of five studies.

Across all five studies, we consistently replicated the beneficial effects of PEs related to aversive events on episodic memory for preceding inherently neutral stimuli. In line with findings on uPE-effects (Kalbe & Schwabe, 2020), Studies I and II showed that uPEs enhanced memory retrospectively, and also prospectively (see Figure 11). However, as more recent evidence suggests that positive and negative PEs differentially affect memory formation (Kalbe & Schwabe, 2022b; Rouhani & Niv, 2021), our behavioral data from Studies III to V revealed that the memory enhancing effects of the PE depended on its valence (see Figure 11). Furthermore, Study III aimed to shed light on the time-dependency of these PE-induced memory-enhancing effects demonstrating that the retrospective effect of PEs persists for at least 10 seconds, and also indicating that this effect emerges already at encoding. While Studies I and II suggested that PE effects may also extend prospectively to stimuli following the PE event, a critical question concerned whether this effect emerged due to contingency or temporal contiguity (Schultz, 2006). Therefore, Study IV provided critical evidence for the selectivity of the PE-effects on memory. Specifically, when uninformative stimuli were inserted between the predictive stimulus and the outcome, the PE-related retrospective memory enhancement was abolished, indicating that the memory enhancement was bound to predictive value indicating contingency-dependence, rather than reflecting a general window of heightened encoding performance. To investigate underlying neural mechanisms, Study V examined oscillatory dynamics and neural states surrounding the PE event. Our findings showed that theta and alpha oscillations, along with pre-PE activation of stimulus-specific neural patterns, modulated the strength of PE-related memory enhancements. Moreover, we also tested the causal role of the

superior parietal cortex in PE-induced memory modulations, given its critical role in working memory processes and top-down attentional updating (Corbetta et al., 1995; D'Esposito & Postle, 2015; Koenigs et al., 2009), using inhibitory brain stimulation. Behaviorally, cTBS did not significantly alter the retrospective PE-effects on memory suggesting that this region is not essential for the PE-effects on memory. However, at the neural level, neural states linked to PE-effects were diminished following cTBS, indicating that an inhibition of the superior parietal cortex may weaken the neural states supporting PE-related memory enhancements. Understanding how PEs affect memory formation provides a non-invasive mechanism for modifying maladaptive beliefs, making it highly relevant for therapeutic interventions in psychological disorders, e.g., fear-related disorders or depression (Gagné et al., 2018; Hein & Ruiz, 2022; Papalini et al., 2020; Queirazza et al., 2019).

6.1 PEs enhance episodic memory formation

6.1.1 PEs boost memory retrospectively

Studies I and II revealed that the PE magnitude, i.e., uPE, affected memory formation. Dovetailing with previous research (Kalbe & Schwabe, 2020; Rosenbaum et al., 2022; Rouhani & Niv, 2021), our findings demonstrated better memory performance if there was a deviation between expectation and actual outcome, regardless of its valence. This effect might also be due to the PE's saliency in general, rather than its valence, also indicating that the deviation between outcome and expectation was more relevant rather than the stimulus' novelty (Bromberg-Martin et al., 2010; Hayden et al., 2011; Sambrook & Goslin, 2014). Specifically, the uPE-effect on memory for aversive events might be supported by the assumption that the LC mediates the influence of uPEs on memory by showing rapid, transient responses to unexpected outcomes, irrespective of their direction, during reward and fear learning (Nassar et al., 2012). This system may serve as an alternative source of dopamine to the HC by co-releasing dopamine along with norepinephrine, thereby supporting dopamine-dependent plasticity in the hippocampus (Steinberg et al., 2013; Takeuchi et al., 2016; Wagatsuma et al., 2018).

When considering the direction of the PE, i.e., sPEs, our results from Studies III to V revealed that negative PEs were associated with impairing retrospective effects on subsequent memory, while positive PEs were linked to enhanced memory retrospectively. At first sight, our findings seem to show the exact opposite pattern of results from Kalbe & Schwabe (2022b) where negative PEs were linked to memory enhancement. However, it is important to note that our pattern of results was consistently found across three studies employing more than 300 participants, underlining the validity and reliability of our results. Further, participants

completed triple the amount of trials (360 vs. 120 trials) compared to Kalbe & Schwabe (2022b), and thus received significantly more electric shocks which may have heightened the participants' sensitivity to the aversive outcome (Chen, Ho, & Liang, 2000; Lonsdorf et al., 2017). Ultimately, this could have intensified the impact of positive PEs, i.e., unexpected shock occurrences. Notably, Studies III and V demonstrate that this retrospective memory-enhancing effect of PEs persists for at least 10 s. Crucially, this effect is evident immediately after encoding, indicating that it does not depend on post-encoding consolidation process.

Interestingly, we also observed that both uPEs and sPEs related to aversive events consistently enhanced memory in a linear fashion, despite prior evidence suggesting a U-shaped relationship between PE magnitude and encoding strength in which both low (i.e., consistent) and high (i.e., surprising) PEs are assumed to lead to facilitated learning (Frank et al., 2018; Greve et al., 2019; Van Kesteren et al., 2012; Ortiz-Tudela et al., 2023). While our findings may initially appear to contradict this framework, they provide a critical refinement by demonstrating that not only the magnitude but also the direction of expectation violations, i.e., worse or better than expected, plays a distinct role in modulating memory formation.

Speculatively, these findings may also be interpreted in light of the event segmentation theory, which assumes that surprising or salient events act as segmentation points in ongoing events by delineating discrete memory episodes (Kurby & Zacks, 2008; Zacks & Swallow, 2007). Although our paradigm did not explicitly probe event boundaries in a continuous stimulus stream, but instead presented discrete trials embedded within a Pavlovian learning task, the observed effects of PEs may functionally resemble such segmentation processes. PEs may serve as internal event boundaries, i.e., subjective inflection points signalling the brain that something important has changed. Even though our task structure involved separate trials, participants continuously formed and updated expectations. Speculatively, when a PE occurs, it may speculatively interrupt this predictive flow, triggering a boundary-like response that reorganizes memory encoding around that moment. Thus, PEs may initiate mnemonic segmentation and enhance the encoding of information surrounding these internal boundaries (Gershman et al., 2014; Rouhani et al., 2020). Our observation that beneficial effects of PEs on memory persist (retrospectively) for several seconds additionally supports the assumption that PEs disrupt the sequential encoding of ongoing events, thereby leading to increased mnemonic salience for stimuli presented before the PE event and even across trial boundaries. Thus, even in structured, trial-based learning paradigms, PEs may segment experiences into meaningful units, broadening the applicability of event segmentation theory to punctuated, rather than continuous, contexts.

6.1.2 uPEs affect memory formation prospectively

Interestingly, we present initial evidence for a prospective effect of uPEs on subsequent memory. In Studies I and II, uPEs not only enhanced memory for events preceding the PE, but also for events following the PE. How do we reconcile that this prospective enhancement might appear to conflict with event boundary theory, which postulates that PEs highlight transitions between discrete memory episodes and may weaken associations across such boundaries (Gershman et al., 2014; Rouhani et al., 2020)? Specifically, PEs may trigger physiological arousal and neuromodulatory activity in the amygdala which could then create a brief window of heightened encoding, during which memory for adjacent stimuli, including those following the PE, is enhanced (Buchanan, 2007; Cahill & McGaugh, 1995; Mather & Sutherland, 2011). In this sense, PEs may act as event boundaries while simultaneously inducing a transient boost in encoding efficiency. These boundaries are also assumed to direct attention to novel information, resulting in enhanced memory for items situated near the boundary (Gold et al., 2017; Heusser et al., 2018; Rubínová & Kontogianni, 2023).

Although these prospective PE effects were generally weaker and less consistently found than retrospective PE effects, they suggest that memory enhancements may extend to stimuli not directly linked to the PE itself. While the prospective effects of PEs on subsequent memory could be interpreted as evidence for a transient window of enhanced memory formation for all stimuli, regardless of their relevance for the PE event, the findings of Study IV argue against such an unselective mechanism. Instead, the results of Study IV underscore the specificity of PE-related memory enhancement, demonstrating that it is restricted to predictive stimuli. Notably, uninformative stimuli not only failed to benefit from this PE-induced memory-promoting effects but even appeared to interfere with them. This specificity may be due to the predictive stimuli's informational value in signaling the outcome which is absent in uninformative cues but present in (predictive) stimuli following the PE. This suggests that prospective PE effects on memory depend on the relevance of the stimulus for outcome prediction, rather than reflecting a general window of enhanced memory encoding.

6.1.3 Dissociation of PE- and arousal effects on memory

Although the observed memory-enhancing effects of PEs partially overlap with arousal-induced memory modulation they are not fully explained by arousal alone (Rouhani et al., 2023). In Studies I, II and IV, we demonstrate that the retrospective memory benefits of PEs persist even after controlling for physiological arousal measured using SCR, respiratory responses and heart rates, dovetailing with previous evidence that PE effects extend beyond arousal-based mechanisms (Kalbe & Schwabe, 2020, 2022b). We were not able to replicate this

effect in Study III where the PE-effect was reduced to a non-significant trend after controlling for arousal. This attenuation may be related to the varying CS-outcome delay in Study III, which could have increased the influence of arousal on memory formation superimposing the PE effects. In contrast, the role of arousal in prospective PE effects remains less clear. While Study II showed that memory for stimuli following a PE is still enhanced after controlling for arousal, this pattern did not reach significance in Study I. Given that prospective PE effects extend across a longer temporal window including stimuli from the subsequent trial, the maintenance of the CS-outcome association may weaken over time, increasing the relative contribution of arousal to enhanced memory. Indeed, recent findings suggest that arousal effects linger in time to bind episodes in memory, helping to shape the temporal organization of events (Clewett & McClay, 2024). Thus, our findings suggest that retrospective PE effects reflect a mechanism at least partially independent of arousal whereas prospective PE-effects appear more susceptible to arousal-driven influences.

6.2 PE-induced memory enhancement is sensitive to interference

Intriguingly, Study IV demonstrated that the presentation of an uninformative stimulus between the CS and the outcome, i.e., the PE-event, abolished the PE-induced memory enhancement for the predictive stimulus, i.e., the CS. This finding suggests that uninformative stimuli may disrupt the association between the predictive stimulus and the PE, most likely by interfering with the active maintenance of the predictive stimulus across the delay. Given that the predictive stimulus needs to be held in working memory until the outcome is revealed, the presence of an intervening UI stimulus may overwrite or interfere with this association. This interpretation aligns with established frameworks of working memory which emphasize its limited capacity and high susceptibility to interference (Baddeley, 1992; Oberauer, 2002; Dosher & Ma, 1998).

Yet, the finding that PE-induced memory enhancements persisted across delays of up to 10 s in Studies III and V suggests that temporal decay alone is insufficient to disrupt the PE-effect on memory. Instead, memory benefits appear robust as long as task-relevant information, i.e., the CS, can be held in mind without competing interference. Presumably, this finding indicates that the working memory load in our paradigm remained low when exactly only a single item had to be maintained, i.e., the CS, thus mitigating the impact of time-based decay. Such a finding is consistent with theories proposing that working memory is more vulnerable to interference than to delay, particularly under minimal working memory load (Dosher 1999; Gresch et al., 2021).

Hence, the PE-effect appears to be not only exclusively linked to stimuli that bear predictive value, but is also modulated by their susceptibility to interference or competing information, as induced by uninformative cues.

To further probe the neural mechanisms underlying this process, Study V used cTBS to inhibit the rSPC, as this area is broadly implicated in cognitive control and working memory updating (Cabeza et al., 2008; Corbetta et al., 1995; D'Esposito & Postle, 2015; Koenigs et al., 2009). Interestingly, rSPC inhibition enhanced general memory performance, primarily via a more conservative response bias, which may reflect altered mnemonic decision processes or facilitated reactivation of relevant information. Although this finding seems counterintuitive, it aligns with prior evidence showing that inhibitory stimulation of lateral parietal regions could enhance associative memory and confidence, presumably by modulating the encoding of new associations in the HC into memory without altering retrieval processes (Tambini et al., 2018). In our case, inhibition of the rSPC may have reduced interference from competing stimulus or outcome representations, thereby supporting a more stable reactivation of the task-relevant predictive stimulus.

Consistent with this idea, multivariate EEG decoding showed stronger stimulus reactivation following rSPC inhibition. One might speculate that these findings suggest that PE-related memory enhancements are supported by the strength and accessibility of task-relevant memory traces, rather than being directly driven by parietal attentional mechanisms. Furthermore, these results align with recent neuroimaging findings linking PE-induced encoding benefits to increased activity in the salience network (Kalbe & Schwabe, 2022b). The salience network serves a crucial gatekeeping function by evaluating the significance of incoming information (Schimmelpfennig et al., 2023). Core regions of this network, including the ACC and the insula, are essential in modulating attention and prioritizing salient information (Menon & Uddin, 2010; Vogt et al., 1992; Weissman et al., 2005). By dynamically allocating attentional resources, the salience network presumably facilitates the selection of behaviorally relevant stimuli and orchestrates neural and behavioral responses. Dysregulation of this system has been associated with deficits in prioritizing information (Green et al., 2016; Schimmelpfennig et al., 2023).

Intricately, while rSPC inhibition modulated general memory outcomes, it did not abolish the PE-induced memory enhancement, indicating that, while cortical modulation can influence memory selectivity, it is not strictly essential for the occurrence of these effects. Rather, these effects likely seem to depend on broader neural dynamics including medial temporal and prefrontal regions, e.g., ventromedial PFC and orbitofrontal cortex, that support the detection

of expectancy violations and the updating of memory based on behaviorally significant outcomes (Kalbe & Schwabe, 2022b; Möhring & Gläscher, 2023). Thus, Study V extends the mechanistic understanding of PE effects on memory by providing causal evidence that cognitive interference and not just attentional filtering can modulate PE-related memory effects.

6.3 Neural states surrounding PEs modulate memory formation

Intriguingly, our behavioral results showed that only stimuli that bear predictive value boost episodic memory formation, in both a retrospective and a prospective manner. But what are the neural mechanisms that underly these PE-induced memory modulations? Study V revealed compelling evidence in favor of the two proposed accounts of PE-induced retrospective memory enhancement: First, stronger stimulus reactivation following PEs suggests that salient events may trigger rapid reactivation of the preceding predictive stimulus, thereby boosting its consolidation. Second, if the predictive stimulus is still active in working memory when the PE occurs, it may benefit from retroactive stabilization through a tagging mechanism induced by the PE, i.e., the salient event.

6.3.1 Pre-PE neural states reflect anticipatory processing

Intriguingly, we found evidence that the neural states before the PE, i.e., oscillatory dynamics and stimulus (category) reactivation, predicted memory changes but we further revealed that the effects critically depended on the sign of the PE (see Figure 11). Our results suggest that for positive PEs, i.e., unexpected shocks, stronger pre-PE stimulus reactivation was linked to impaired memory, presumably due to interference with the highly salient shock that leads to competing attentional resources at the time of encoding where the salient stimulus captures attention (Kerzel & Schönhammer, 2013; Mather & Sutherland, 2011). Additionally, this effect seems to be increased if affective salience, as induced by the aversive shock, is taken into account (Biggs et al., 2012). This effect was not observed for negative PEs, i.e., unexpected shock omissions, likely due to the outcomes' lower emotional salience. At first glance, this finding seems to contradict the behavioral tagging account where a memory enhancement would be postulated for the stimulus preceding a salient event, e.g., a PE (Moncada & Viola, 2007; Moncada et al., 2015). However, this discrepancy may be explained by considering the neural dynamics surrounding highly salient events. In the case of an unexpected shock, the attentional and neural resources required to process the sudden, emotionally intense outcome may interfere with the ongoing maintenance of the preceding stimulus trace, thereby disrupting its consolidation. In other words, while the shock has the potential to act as a reinforcing salient tag, its intensity may disrupt the integration of the preceding memory trace unless that trace has been optimally maintained or encoded in a way that resists interference. It is tempting to

speculate that the PE may possess the capacity to stabilize prior memory traces while this effect likely depends on whether attentional resources are sufficiently available to maintain the stimulus representation before the PE. This interpretation could also propose a limitation of the behavioral tagging account: The successful stabilization of weak memory traces by salient events may require not only a temporal overlap but also a balanced attentional resource allocation.

Beyond stimulus-specific reactivation, our results show that neural oscillatory states preceding the PE also modulate memory formation depending on the direction of the PE (see Figure 11). Specifically, for negative PEs, increased alpha and theta activity were associated with enhanced subsequent memory. This pattern presumably reflects enhanced attentional engagement and associative binding processes, given that alpha oscillations have been linked to the suppression of irrelevant information and sustained attention (Bonnefond & Jensen, 2012; Khader et al., 2010; Payne & Sekuler, 2014), while theta oscillations have been associated with memory encoding and the binding of contextual associations (Roux et al., 2022; Staudigl & Hanslmayr, 2013). Contrarily, for positive PEs, enhanced alpha and theta oscillations were linked to impaired memory, in line with an increased reactivation of the predictive stimulus. These converging effects suggest that a heightened preparatory neural state may render the system more vulnerable to interference when a highly salient outcome occurs, i.e., a shock, consistent with accounts of attentional competition and limited resource allocation (Asgeirsson & Nieuwenhuis, 2019; Mather & Sutherland, 2011). Together, these findings again indicate that PE-induced memory modulation is not determined solely by the PE itself but is critically shaped by the neural state immediately preceding the outcome and its salience.

6.3.2 PE-induced changes in neural dynamics drive memory formation

Beyond the importance of the neural state immediately preceding the PE event, we also considered the alternative mechanism of PE-induced memory enhancement in which the PE triggers neural changes that promote the consolidation of preceding stimuli. Intriguingly, PEs were linked to increased neural stimulus (category) reactivation, dovetailing with evidence proposing that event boundaries, i.e., the PE event, induce rapid memory reinstatement to integrate prior information and support memory updating (Sinclair et al., 2021; Sols et al., 2017). Notably, this effect was dependent on the direction of the PE: Negative PEs enhanced stimulus reactivation after the outcome, most likely reflecting an adaptive updating of internal models in response to unexpected safety signals. This interpretation aligns with prior research showing that omissions of expected aversive outcomes engage mnemonic and learning-related processes, particularly those associated with model updating, highlighting

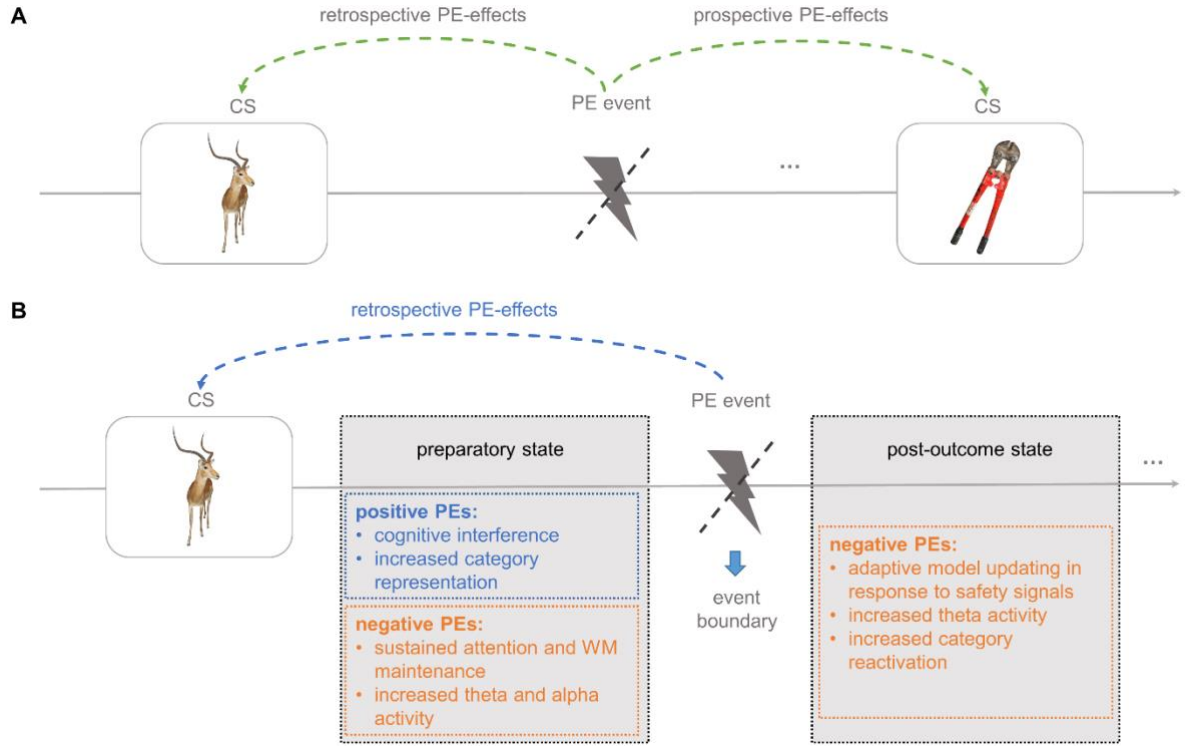
the salience of safety signals in guiding learning and memory adaptation (Iglesias et al., 2013). In contrast, positive PEs seemed to impair reactivation, presumably due to the attentional capture elicited by the unexpected shock, which may interrupt ongoing encoding processes (Biggs et al., 2012; Mather & Sutherland, 2011).

Crucially, however, the observed stimulus reactivation after the PE did not directly predict subsequent memory performance. This dissociation suggests that a reactivation may not serve a consolidation function per se, but may instead support online model updating, thereby enabling the integration of expectancy violations into an evolving internal representation of the environment (Gershman & Niv, 2010; Sinclair et al., 2021). This interpretation also aligns with findings from Study III, where PE effects emerged already during encoding, rather than reflecting offline consolidation. In other words, the PE-induced stimulus reactivation may serve as a computational mechanism for updating internal models rather than a driver of episodic memory enhancement, highlighting the complex and functionally distinct roles that reactivation can play in cognition.

In addition, oscillatory dynamics induced by PEs further support the notion of differential encoding pathways depending on the direction of the PE: Increased theta activity after negative PEs was associated with enhanced memory, suggesting a role of theta oscillations in the consolidation of memories following safety signals (see Figure 11). Theta activity has broadly been linked to error detection and correction (Kalfaoglu et al., 2018) and is sensitive to the degree of negative and positive PEs in adapting behavior in reinforcement learning paradigms (Cavanagh et al., 2010). In this context, increased theta activity following negative PEs may reflect increased encoding efficiency under surprise, speculatively via suppression of the default mode network, which might otherwise disrupt effective memory formation (Kalbe & Schwabe, 2022b; Klimesch, 1999; Kota et al., 2020; White et al., 2013). Furthermore, theta oscillations may facilitate the binding of item information to its spatiotemporal context, particularly through interactions within medial temporal structures and hippocampo-cortical feedback loops (Klimesch, 1999; Hanslmayr et al., 2011; Kota et al., 2020). Notably, this theta-linked enhancement was absent for positive PEs, emphasizing that distinct neural mechanisms guide memory modulation for expectancy violations depending on the emotional salience and attentional capture (Mather & Sutherland, 2011).

Figure 11

Dynamic states supporting PE-induced memory modulation



Note. Prediction errors (PEs) boost memory selectively for conditioned stimuli (CS) and act as event boundaries. (A) Unsigned PEs enhance memory retro- and prospectively (dashed green arrows), while signed PEs boost memory retrospectively depending on their direction (B). Overall, positive PEs (dashed blue arrow) enhance memory retrospectively, while negative PEs attenuate memory formation. The neural states surrounding the PE event, i.e., preparatory and post-outcome state, affect this PE-driven memory modulation separately. WM = working memory; three dots indicate the start of a new trial. Pictures, i.e., antelope and bold-cutter, taken from "The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research" by Brodeur et al. (2010) and from "Introducing the Open Affective Standardized Image Set (Oasis)" by Kurdi et al. (2016). CC BY 4.0.

6.4 Modulatory role of the rSPC in PE-induced memory enhancements

However, it still remains unclear whether the PE-driven memory enhancement necessitates these neural maintenance mechanisms. Interestingly, when applying cTBS over the rSPC as a causal probe, the underlying neural mechanisms of PE-induced memory modulations appeared to be altered. Specifically, the association between pre-PE alpha and theta oscillatory activity and subsequent memory was diminished following rSPC inhibition in Study V. This findings suggests that the rSPC may contribute to preparatory neural states that shape encoding around surprising events, i.e., a PE, presumably by guiding anticipatory attention and the maintenance of predictive representations (D'Esposito & Postle, 2015; Koenigs et al., 2009). At first glance, this disruption of preparatory neural states seems to contradict the observation that rSPC inhibition was linked to improved memory and enhanced stimulus reactivation after the PE. One possible interpretation is that the rSPC plays temporally distinct roles across the

PE window: it may support top-down modulation during the anticipatory phase, i.e., CS-outcome delay, while its engagement after outcome occurrence could interfere with the effective reactivation or integration of predictive stimuli, particularly in the face of highly salient events such as PEs. In this context, rSPC inhibition might have reduced such interference, thereby allowing for more effective memory tagging or consolidation processes after a PE (Moncada et al., 2015). However, this interpretation remains speculative and does not fully clarify how rSPC inhibition interacts with the PE effect on memory. While the PE-related memory benefit persisted under rSPC inhibition, its neural underpinnings may have shifted, perhaps relying on alternative pathways or compensatory mechanisms. Future work is needed to disentangle how parietal contributions to anticipatory control interact with PE-driven memory mechanisms, and whether rSPC involvement differentially supports predictive processing versus outcome-driven memory updating.

6.5 A dynamic framework of PE-induced memory modulation

This thesis proposes a dynamic, mechanistic framework by which PEs related to aversive events shape episodic memory formation through temporally and functionally distinct cognitive mechanisms (see Figure 12). Rather than a unitary effect, PE-induced memory modulation emerges from the interaction of predictive processing, attentional gating and post-encoding mnemonic updating.

PEs function as computational signals that demarcate event boundaries and facilitate memory segmentation (Gershman et al., 2014; Rouhani et al., 2020). By violating internal predictions, PEs reset ongoing encoding processes, thereby increasing the mnemonic distinctiveness of temporally adjacent events (Sols et al., 2017; Zacks et al., 2007). This segmentation does not enhance memory indiscriminately but reflects a prioritization mechanism wherein the relevance, salience and context of the PE determine which elements are strengthened or suppressed (Bein et al., 2023; Bein & Niv, 2025).

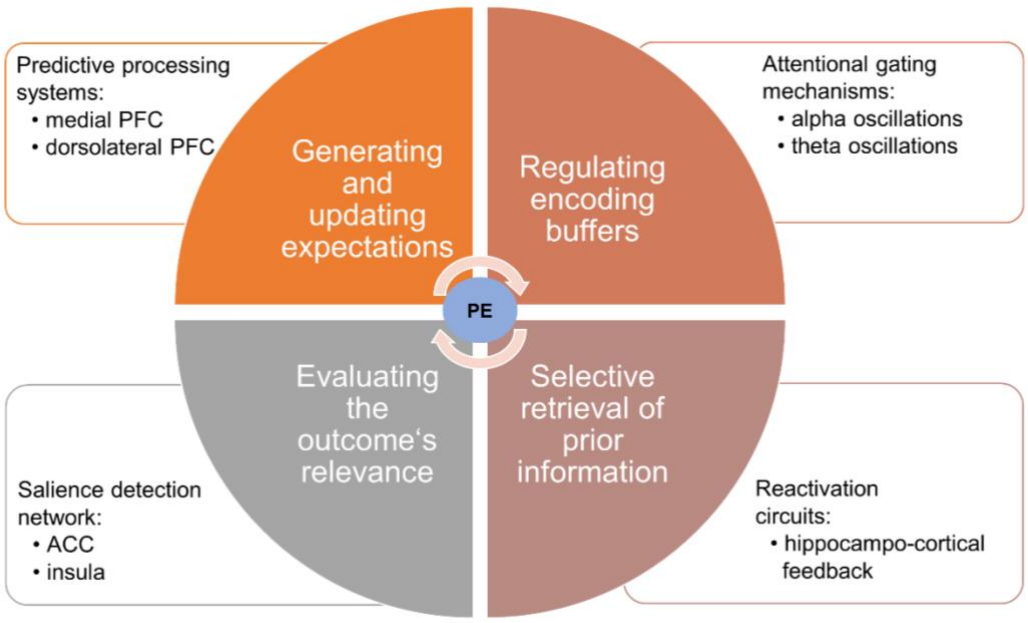
Neural states prior to a PE, particularly oscillatory dynamics in the alpha and theta bands, reflect the system's readiness to encode or update expectations. Elevated pre-PE alpha and theta activity may indicate increased internal attention and active maintenance of predictive representations. Mechanistically, these oscillations may gate incoming information by modulating sensory precision or predictive gain (Friston, 2009; Klimesch, 1999, 2012), thus determining whether a surprising event disrupts or reinforces current memory traces. Crucially, this gating effect seems valence-sensitive: Negative PEs, i.e., unexpected safety signals, are more likely to benefit from strong pre-PE engagement, while positive PEs, i.e., unexpected threats, may override attentional control, leading to interference and poorer encoding (Mather

& Sutherland, 2011). This dichotomy supports a model in which the emotional salience of the PE interacts with internal attentional states to determine encoding outcomes (see Figure 11).

PEs also induce rapid reactivation of preceding stimulus representations, consistent with a memory tagging account (Moncada et al., 2015; Frey & Morris, 1997). However, our findings suggest that such reactivation does not always correlate with improved memory performance, implying a functional dissociation. Rather than serving direct consolidation, PE-induced reactivation may primarily support internal model updating, i.e., retrieving relevant prior information to revise internal representations (Sinclair et al., 2021). If this process overlaps with episodic memory encoding, e.g., under conditions of emotional neutrality, it may speculatively also lead to memory strengthening. Contrarily, if there is a highly salient distraction, as in positive PEs, reactivation may fail to support consolidation. Theta oscillations observed after negative PEs further support this notion. They may serve as a mechanism for temporally binding reactivated stimuli to their new contextual relevance, facilitating adaptive memory formation (Hanslmayr et al., 2011; Roux et al., 2022). This aligns with reinforcement learning models where theta oscillations reflect the integration of surprise signals into behavioral updating (Cavanagh et al., 2010).

Figure 12

Framework of cognitive mechanisms underlying PE effects on memory



Note. Prediction error-related memory effects arise from a distributed, time-sensitive mechanism, relying on the coordination of attentional, predictive, and memory systems across distinct stages of event processing. PE = Prediction error, PFC = prefrontal cortex, ACC = anterior cingulate cortex.

Together, these findings support a model where memory modulation by PEs results from (i) coordinated interactions between predictive processing systems (e.g., mPFC, dlPFC) that generate and update expectations, (ii) attentional gating mechanisms, as indexed by alpha and theta oscillations regulating access to encoding buffers, (iii) salience detection networks (e.g., ACC, insula) that evaluate the relevance of unexpected outcomes and trigger mnemonic processing and (iv) reactivation circuits (e.g., hippocampo-cortical feedback) that selectively retrieve prior information to revise current models or support consolidation (see Figure 12). This integrative view proposes a distributed, temporally-resolved mechanism in which PE effects on memory depend on the alignment of attentional, predictive, and mnemonic systems at different stages of event processing. Although this framework is informed by converging behavioral and EEG findings, the involvement of specific neural circuits remains to be fully characterized and would benefit from future research using complementary methods (e.g., EEG-fMRI, intracranial EEG).

6.6 Methodological considerations and future perspectives

Despite offering a comprehensive framework for understanding the cognitive and neural dynamics of aversive PE-induced memory modulation, several limitations need to be considered when interpreting our findings.

Although cTBS allowed us to investigate the causal role of the rSPC in PE-induced memory modulations, the interpretation of these inhibitory effects remains complex. Specifically, applying cTBS does not necessarily imply that the region of interest is inhibited. Targeting the rSPC cannot rule out downstream or network-level effects, i.e., on the frontoparietal or salience networks, given that TMS has been shown to potentially also disrupt processing at distant sites that have not been targeted directly but could still interfere with task behavior (Siebner et al., 2009). Individual variability in functional anatomy may further limit stimulation efficacy, although using neuro-navigated TMS helps to mitigate this concern. To clarify the functional relevance of the rSPC, future studies could incorporate a group in the experimental design in which excitatory TMS of the rSPC is applied in the experimental design. Additionally, given that the rSPC seems to be rather involved in attentional mechanisms than in working memory processes (Study V), we might have forced an attention-biased view on PEs by stimulating the rSPC. To probe working memory contributions more directly, future studies might target the dlPFC that is implicated in active maintenance, error monitoring, and memory updating (D'Esposito & Postle, 2015; Kluen et al., 2019; Pine et al., 2018). However, stimulation of the dlPFC poses practical challenges, as it can lead to painful sensations on the forehead and unintended activation of facial nerves. In this context, transcranial direct current

stimulation may offer a less invasive and more tolerable, thus viable alternative (Woods et al., 2016). While well-controlled stimuli ensured reproducibility, it limits ecological validity (Shamay-Tsoory & Mendelsohn, 2019). PE-induced memory modulations may act in a more nuanced way in naturalistic or emotionally rich environments such as autobiographical memory or clinical populations that suffer from maladaptive prediction processes, e.g., post-traumatic stress disorders (PTSD) or anxiety disorders. Incorporating more realistic scenarios would enhance the translational relevance of expectancy violations in healthy and clinical populations.

Although separately investigating specific mechanisms in the five individual studies of this thesis helped to isolate PE effects, combining such manipulations in a single paradigm would provide greater explanatory power. For instance, integrating UI with variable CS-outcome delays could help to disentangle effects of attentional filtering versus working memory mechanisms and clarify their respective roles in PE-based memory modulation.

While the decoding approach in Study using MVPA, enabled us to examine to the dynamics of the neural patterns underlying PE-induced memory enhancements, its results are methodologically constrained to category-level effects (Naselaris & Kay, 2015; Treder, 2020), limiting our ability to detect trial-unique stimuli reactivations. Future studies that employ simultaneous EEG-fMRI could overcome this limitation by capturing both the temporal and spatial dynamics of PE-induced memory modulations. Crucially, such multimodal data can support stimulus-specific analyses on both EEG and fMRI data, enabling the investigation of memory replay of predictive stimuli and their role in consolidation processes (see Huang et al., 2024; Tambini & Davachi, 2019).

Building on the current findings, several promising avenues for future research could further elucidate the mechanisms of PE-induced memory modulations. Interestingly, recent research has shown that the updating of information in case of expectancy violations depends on the memory strength of the predictive cue, suggesting that stronger initial encoding can influence the likelihood and degree of subsequent reactivation (Yu & Davachi, 2025). In the case of the conducted studies, future approaches could manipulate the strength of the predictive stimulus to clarify how encoding strength interacts with PE processing and ultimately memory updating. This might be done by increasing the presentation time of the to-be-encoded pictures to enhance encoding depth (Craik & Tulving, 1975) or by varying the emotional valence, given that emotional items are assumed to be better remembered than neutral ones (Cahill & McGaugh, 1996). Although the PE-induced memory modulations seem to be distinct from MTL activation which is commonly assumed in memory consolidation and to rely on crosstalk between including the salience and frontoparietal networks (Alvarez & Squire, 1994; Hermans

et al., 2014; Kalbe & Schwabe, 2022b), our findings regarding theta oscillations (Study V) and their impact on PE-effects also point to an involvement of HC-based circuits. Future studies could test this directly leveraging hippocampal-targeted TMS (Tambini et al., 2018) to probe the causal role of the HC in memory formation and consolidation following PEs.

More broadly, the PE-induced memory enhancements are assumed to involve dopaminergic and noradrenergic neuromodulatory systems (Gershman et al., 2024). However, these mechanisms have not yet been investigated in the context of fear conditioning. Pharmacological studies, e.g., using noradrenaline or dopamine antagonists, could manipulate these systems to determine their roles in mediating the PE-related memory fromation.

6.7 Conclusion

Considering the pivotal role of expectancy violations in our daily experiences, this thesis contributes to an essential understanding of how PEs related to aversive events modulate episodic memory. When a surprising, aversive event occurs, our studies reveal that memory can be enhanced for events occurring both before and after the PE, although in a selective manner. The preparatory cognitive state preceding a PE, reflected in oscillatory alpha and theta activity, appears to set the stage for these mnemonic shifts, influencing whether perceived information is prioritized. While top-down attentional control mechanisms, potentially involving regions like the rSPC (D'Esposito & Postle, 2015), may contribute to this preparatory regulation, our findings suggest that such control is not strictly necessary for PE-related memory enhancements. In fact, if this control is inhibited, a counterintuitive effect emerges: Reactivation of information encoded before the PE is enhanced, while memory performance improves, indicating that releasing top-down constraints may facilitate memory integration of surprising events. Across studies, our findings indicate that PEs serve as temporal anchors that not only tag salient information but also initiate processes of mnemonic updating and speculatively reshape prior schemas by segmenting experiences (Bein & Niv, 2025). Oscillatory mechanisms induced by PEs, particularly in the theta band, and their broader interaction with attentional systems may thus provide the neural infrastructure for this adaptive flexibility. Collectively, our results advance a mechanistic account of how the brain forms episodic memory in response to unexpected events. It emphasizes the temporal interplay between salience detection and interference control, proposing that surprise-induced flexibility in information processing may be critical for adaptive memory formation. Moreover, the broader relevance of PEs extends beyond the confines of experimental research. In clinical settings, expectancy violations are central to the mechanisms underlying exposure-based therapies (Pittig et al., 2023), which are widely used in the treatment of fear-related disorders,

anxiety disorders and PTSD. These therapies aim to weaken maladaptive associations by repeatedly confronting individuals with feared stimuli in the absence of the expected negative outcome, thereby generating PEs that drive learning and emotional updating (Gagné et al., 2018; Kube et al., 2020; Putica et al., 2022; White et al., 2017; Winkler et al., 2025). By elucidating the neural and cognitive mechanisms through which PEs shape memory formation, our work may inform strategies to optimize therapeutic interventions, specifically by targeting maladaptive prediction processes.

References

Alexander, W. H., & Brown, J. W. (2019). The Role of the Anterior Cingulate Cortex in Prediction Error and Signaling Surprise. *Topics in Cognitive Science*, *11*(1), 119–135. https://doi.org/10.1111/tops.12307

Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110 Suppl 2*(Suppl 2), 10379–10386. https://doi.org/10.1073/pnas.1301199110

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(15), 7041–7045. https://doi.org/10.1073/pnas.91.15.7041

Anderson, C., & Colombo, M. (2019). Matching-to-Sample: Comparative Overview. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of Animal Cognition and Behavior*. Springer. https://doi.org/10.1007/978-3-319-47829-6_1708-1

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), Article 2. https://doi.org/10.1016/j.neuron.2020.10.029

Antony, J. W., Van Dam, J., 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

Ásgeirsson, Á. G., & Nieuwenhuis, S. (2019). Effects of arousal on biased competition in attention and short-term memory. *Attention, Perception, & Psychophysics*, *81*(6), 1901–1912. https://doi.org/10.3758/s13414-019-01756-x

Bach, D. R., Gerster, S., Tzovara, A., & Castegnetti, G. (2016). A linear model for event-related respiration responses. *Journal of Neuroscience Methods*, *270*, 147–155. https://doi.org/10.1016/j.jneumeth.2016.06.001

Baddeley, A. (1992). Working Memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Ballarini, F., Moncada, D., Martinez, M. C., Alen, N., & Viola, H. (2009). Behavioral tagging is a general mechanism of long-term memory formation. *Proceedings of the National Academy of Sciences*, *106*(34), 14599–14604. https://doi.org/10.1073/pnas.0907078106

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), Article 7. https://doi.org/10.1038/nrn3950

Bartlett, J. C., Till, R. E., & Levy, J. C. (1980). Retrieval characteristics of complex pictures: Effects of verbal encoding. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 430–449. https://doi.org/10.1016/S0022-5371(80)90303-5

Barto, A. G. (1995). Reinforcement Learning and Dynamic Programming. *IFAC Proceedings Volumes*, *28*(15), 407–412. https://doi.org/10.1016/S1474-6670(17)45266-9

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

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron*, *47*(1), Article 1. https://doi.org/10.1016/j.neuron.2005.05.020

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bein, O., Duncan, K., & Davachi, L. (2020). Mnemonic prediction errors bias hippocampal states. *Nature Communications*, *11*(1), Article 1. 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*, 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

Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience*, *27*(7), 1823–1835. https://doi.org/10.1111/j.1460-9568.2008.06138.x

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Bergt, A., Urai, A. E., Donner, T. H., & Schwabe, L. (2018). Reading memory formation from the eyes. *European Journal of Neuroscience*, *47*(12), 1525–1533. https://doi.org/10.1111/ejn.13984

Biggs, A. T., Kreager, R. D., Gibson, B. S., Villano, M., & Crowell, C. R. (2012). Semantic and affective salience: The role of meaning and preference in attentional capture and disengagement. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 531–541. https://doi.org/10.1037/a0027394

Bonnefond, M., & Jensen, O. (2012). Alpha Oscillations Serve to Protect Working Memory Maintenance against Anticipated Distracters. *Current Biology*, *22*(20), 1969–1974. https://doi.org/10.1016/j.cub.2012.08.029

Boucsein, W. (1992). *Electrodermal Activity*. Springer.

Braem, S., Coenen, E., Bombeke, K., van Bochove, M. E., & Notebaert, W. (2015). Open your eyes for prediction errors. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), Article 2. https://doi.org/10.3758/s13415-014-0333-4

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLoS ONE*, *9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. *Neuron*, *68*(5), 815–834. https://doi.org/10.1016/j.neuron.2010.11.022

Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychological Bulletin*, *133*(5), 761. https://doi.org/10.1037/0033-2909.133.5.761

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, *35*(4), 625–641. https://doi.org/10.1016/s0896-6273(02)00830-9

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index (PSQI): A new instrument for psychiatric research and practice. *Psychiatry Research*, *28*(2), 193–213.

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*(8), 613–625. https://doi.org/10.1038/nrn2459

Cahill, L., & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, *4*(4), 410–421. https://doi.org/10.1006/ccog.1995.1048

Cahill, L., & McGaugh, J. L. (1996). Modulation of memory storage. *Current Opinion in Neurobiology*, *6*(2), 237–242. https://doi.org/10.1016/S0959-4388(96)80078-X

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, *21*(7), 294–299. https://doi.org/10.1016/S0166-2236(97)01214-9

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. *Journal of Neuroscience*, *41*(8), 1716–1726. https://doi.org/10.1523/JNEUROSCI.1785-20.2020

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, *49*(4), 3198–3209. https://doi.org/10.1016/j.neuroimage.2009.11.080

Chadwick, M. J., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2010). Decoding Individual Episodic Memory Traces in the Human Hippocampus. *Current Biology*, *20*(6), 544–547. https://doi.org/10.1016/j.cub.2010.01.053

Charpentier, C. J., Faulkner, P., Pool, E. R., Ly, V., Tollenaar, M. S., Kluen, L. M., Fransen, A., Yamamori, Y., Lally, N., Mkrtchian, A., Valton, V., Huys, Q. J. M., Sarigiannidis, I., Morrow, K. A., Krenz, V., Kalbe, F., Cremer, A., Zerbes, G., Kausche, F. M., & O'Doherty, J. P. (2021). How representative are neuroimaging samples? Large-scale evidence for trait anxiety differences between fMRI and behaviour-only research participants. *Social Cognitive and Affective Neuroscience*, *16*(10), 1057–1070. https://doi.org/10.1093/scan/nsab057

Chen, D. Y., Ho, S. H., & Liang, K. C. (2000). Startle responses to electric shocks: Measurement of shock sensitivity and effects of morphine, buspirone and brain lesions. *The Chinese Journal of Physiology*, *43*(1), 35–47.

Christianson, S.-A. (2014). *The Handbook of Emotion and Memory: Research and Theory*. Psychology Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), Article 3. https://doi.org/10.1017/S0140525X12000477

Clewett, D., & McClay, M. (2024). Emotional arousal lingers in time to bind discrete episodes in memory. *Cognition and Emotion*, 1–20. https://doi.org/10.1080/02699931.2023.2295853

Clewett, D., & Murty, V. P. (2019). Echoes of Emotions Past: How Neuromodulators Determine What We Recollect. *ENeuro*, *6*(2), Article 2. https://doi.org/10.1523/ENEURO.0108-18.2019

Clewett, D., Schoeke, A., & Mather, M. (2014). Locus coeruleus neuromodulation of memories encoded during negative or unexpected action outcomes. *Neurobiology of Learning and Memory*, *111*, 65–70. https://doi.org/10.1016/j.nlm.2014.03.006

Clewett, D. V., Huang, R., Velasco, R., Lee, T.-H., & Mather, M. (2018a). Locus Coeruleus Activity Strengthens Prioritized Memories Under Arousal. *The Journal of Neuroscience*, *38*(6), 1558–1574. https://doi.org/10.1523/JNEUROSCI.2097-17.2017

Cohen, J. R., Sreenivasan, K. K., & D'Esposito, M. (2014). Correspondence Between Stimulus Encoding- and Maintenance-Related Neural Processes Underlies Successful Working Memory. *Cerebral Cortex*, *24*(3), Article 3. https://doi.org/10.1093/cercor/bhs339

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *The European Journal of Neuroscience*, *35*(7), 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x

Corbetta, M., Shulman, G. L., Miezin, F. M., & Petersen, S. E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science*, *270*(5237), 802–805. https://doi.org/10.1126/science.270.5237.802

Corlett, P. R., Mollick, J. A., & Kober, H. (2022). Meta-analysis of human prediction error for incentives, perception, cognition, and action. *Neuropsychopharmacology*, *47*(7), Article 7. https://doi.org/10.1038/s41386-021-01264-3

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), Article 7. https://doi.org/10.1016/j.tics.2006.05.004

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: Insights from relational and item-based learning. *Journal of Neurophysiology*, *88*(2), 982–990. https://doi.org/10.1152/jn.2002.88.2.982

Dayan, P. (1993). Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation*, *5*(4), Article 4. https://doi.org/10.1162/neco.1993.5.4.613

de Berker, A. O., Rutledge, R. B., Mathys, C., Marshall, L., Cross, G. F., Dolan, R. J., & Bestmann, S. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nature Communications*, *7*(1), 10996. https://doi.org/10.1038/ncomms10996

Delgado, M. R., Li, J., Schiller, D., & Phelps, E. A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1511), 3787–3800. https://doi.org/10.1098/rstb.2008.0161

Den Ouden, H. E., Kok, P., & De Lange, F. P. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00548

Den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A Dual Role for Prediction Error in Associative Learning. *Cerebral Cortex*, *19*(5), Article 5. https://doi.org/10.1093/cercor/bhn161

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 761–772. https://doi.org/10.1098/rstb.2007.2086

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59*(4), 447–456. https://doi.org/10.1016/j.jml.2007.11.004

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, *18*(5), Article 5. https://doi.org/10.1038/nn.3981

Dosher, B. A. (1999). Item interference and time delays in working memory: Immediate serial recall. *International Journal of Psychology*, *34*(5–6), 276–284. https://doi.org/10.1080/002075999399576

Dosher, B. A., & Ma, J.-J. (1998). Output loss or rehearsal loop? Output-time versus pronunciation-time limits in immediate recall for forgetting-matched materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 316. https://doi.org/10.1037/0278-7393.24.2.316

Dunsmoor, J. E., Murty, V. P., Clewett, D., Phelps, E. A., & Davachi, L. (2022). Tag and capture: How salient experiences target and rescue nearby events in memory. *Trends in Cognitive Sciences*, *26*(9), Article 9. https://doi.org/10.1016/j.tics.2022.06.009

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature*, *520*(7547), 345–348. https://doi.org/10.1038/nature14106

Düzel, E., Bunzeck, N., Guitart-Masip, M., Wittmann, B., Schott, B. H., & Tobler, P. N. (2009). Functional imaging of the human dopaminergic midbrain. *Trends in Neurosciences*, *32*(6), 321–328. https://doi.org/10.1016/j.tins.2009.02.005

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109–120. https://doi.org/10.1016/j.neuron.2004.08.028

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The Medial Temporal Lobe and Recognition Memory. *Annual Review of Neuroscience*, *30*(1), 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Ergo, K., De Loof, E., Janssens, C., & Verguts, T. (2019). Oscillatory signatures of reward prediction errors in declarative learning. *NeuroImage*, *186*, 137–145. https://doi.org/10.1016/j.neuroimage.2018.10.083

Ergo, K., De Loof, E., & 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

Eshel, N., Tian, J., Bukwich, M., & Uchida, N. (2016). Dopamine neurons share common response function for reward prediction error. *Nature Neuroscience*, *19*(3), Article 3. https://doi.org/10.1038/nn.4239

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Fanselow, M. S. (1998). Pavlovian Conditioning, Negative Feedback, and Blocking: Mechanisms that Regulate Association Formation. *Neuron*, *20*(4), Article 4. https://doi.org/10.1016/S0896-6273(00)81002-8

Fastenrath, M., Coynel, D., Spalek, K., Milnik, A., Gschwind, L., Roozendaal, B., Papassotiropoulos, A., & de Quervain, D. J. (2014). Dynamic modulation of amygdala–hippocampal connectivity by emotional arousal. *Journal of Neuroscience*, *34*(42), 13935–13947. https://doi.org/10.1523/JNEUROSCI.0786-14.2014

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

FeldmanHall, O., Montez, D. F., Phelps, E. A., Davachi, L., & Murty, V. P. (2021). Hippocampus guides adaptive learning during dynamic social interactions. *Journal of Neuroscience*, *41*(6), 1340–1348. https://doi.org/10.1523/JNEUROSCI.0873-20.2020

Ferreira-Santos, F. (2016). The role of arousal in predictive coding. *Behavioral and Brain Sciences*, *39*, e207. https://doi.org/10.1017/S0140525X15001788

Figner, B., & Murphy, R. O. (2011). *A handbook of process tracing methods for decision research*. Psychology Press.

Fiorillo, C. D. (2013). Two Dimensions of Value: Dopamine Neurons Represent Reward But Not Aversiveness. *Science*, *341*(6145), Article 6145. https://doi.org/10.1126/science.1238699

Fouragnan, E., Retzler, C., & Philiastides, M. G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. *Human Brain Mapping*, *39*(7), Article 7. https://doi.org/10.1002/hbm.24047

Frank, D., Kafkas, A., & Montaldi, D. (2022). Experiencing Surprise: The Temporal Dynamics of Its Impact on Memory. *Journal of Neuroscience*, *42*(33), 6435–6444. https://doi.org/10.1523/JNEUROSCI.1783-21.2022

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

Frey, S., & Frey, J. U. (2008). Chapter 7 ‘Synaptic tagging’ and ‘cross-tagging’ and related associative reinforcement processes of functional plasticity as the cellular basis for memory formation. In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *Progress in Brain Research* (Vol. 169, pp. 117–143). Elsevier. https://doi.org/10.1016/S0079-6123(07)00007-6

Frey, U., & Morris, R. G. (1997). Synaptic tagging and long-term potentiation. *Nature*, *385*(6616), 533–536. https://doi.org/10.1038/385533a0

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), Article 1456. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), Article 2. https://doi.org/10.1038/nrn2787

Gagné, J.-P., Kelly-Turner, K., & Radomsky, A. S. (2018). From the laboratory to the clinic (and back again): How experiments have informed cognitive–behavior therapy for obsessive–compulsive disorder. *Journal of Experimental Psychopathology*, *9*(4), 2043808718810030. https://doi.org/10.1177/2043808718810030

Gagnon, S. A., Waskom, M. L., Brown, T. I., & Wagner, A. D. (2019). Stress Impairs Episodic Retrieval by Disrupting Hippocampal and Cortical Mechanisms of Remembering. *Cerebral Cortex*, *29*(7), 2947–2964. https://doi.org/10.1093/cercor/bhy162

Ganesh, P., Donner, T., Cichy, R., Schuck, N., Finke, C., & Bruckner, R. (2024). *Pupil-linked arousal encodes uncertainty-weighted prediction errors*. OSF. https://doi.org/10.31234/osf.io/c6ujk

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *ELife*, *6*, e17086. https://doi.org/10.7554/eLife.17086

Gershman, S. J. (2017). Predicting the past, remembering the future. *Current Opinion in Behavioral Sciences*, *17*, 7–13. https://doi.org/10.1016/j.cobeha.2017.05.025

Gershman, S. J., Assad, J. A., Datta, S. R., Linderman, S. W., Sabatini, B. L., Uchida, N., & Wilbrecht, L. (2024). Explaining dopamine through prediction errors and beyond. *Nature Neuroscience, 27*(9), 1645–1655. https://doi.org/10.1038/s41593-024-01705-4

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*(1), 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256. https://doi.org/10.1016/j.conb.2010.02.008

Gershman, S. J., Radulescu, A., Norman, K. A., & Niv, Y. (2014). Statistical Computations Underlying the Dynamics of Memory Updating. *PLOS Computational Biology*, *10*(11), e1003939. https://doi.org/10.1371/journal.pcbi.1003939

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. *Neuron*, *66*(4), Article 4. https://doi.org/10.1016/j.neuron.2010.04.016

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, *108* (supplement_3), 15647-15654. https://doi.org/10.1073/pnas.1014269108

Gold, D. A., Zacks, J. M., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive Research: Principles and Implications*, *2*(1), 1. https://doi.org/10.1186/s41235-016-0043-2

Gordon, A. M., Rissman, J., Kiani, R., & Wagner, A. D. (2014). Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cerebral Cortex*, *24*(12), 3350–3364. https://doi.org/10.1093/cercor/bht194

Grandchamp, R., & Delorme, A. (2011). Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. *Frontiers in Psychology*, *2*, 236. https://doi.org/10.3389/fpsyg.2011.00236

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Green, S. A., Hernandez, L., Bookheimer, S. Y., & Dapretto, M. (2016). Salience Network Connectivity in Autism Is Related to Brain and Behavioral Markers of Sensory Overresponsivity. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*(7), 618-626.e1. https://doi.org/10.1016/j.jaac.2016.04.013

Gresch, D., Boettcher, S. E. P., Van Ede, F., & Nobre, A. C. (2021). Shielding working-memory representations from temporally predictable external interference. *Cognition*, *217*, 104915. https://doi.org/10.1016/j.cognition.2021.104915

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

Grob, A. M., Heinbockel, H., Milivojevic, B., Doeller, C. F., & Schwabe, L. (2024). Causal role of the angular gyrus in insight-driven memory reconfiguration. *Elife*, *12*, RP91033. https://doi.org/10.7554/eLife.91033.3

Gurunandan, K., Greve, A., & Henson, R. (2025). *Does Signed Prediction Error drive Declarative Memory? Evidence from Variable Choice Paradigms*. OSF. https://doi.org/10.31234/osf.io/3sbfk_v1

Hanslmayr, S., Volberg, G., Wimber, M., Raabe, M., Greenlee, M. W., & Bäuml, K. H. T. (2011). The relationship between brain oscillations and BOLD signal during memory formation: A combined EEG–fMRI study. *Journal of Neuroscience*, *31*(44), 15674–15680. https://doi.org/10.1523/JNEUROSCI.3140-11.2011

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 1. https://doi.org/10.1038/s41467-020-19828-0

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Platt, M. L. (2011). Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. *Journal of Neuroscience*, *31*(11), Article 11. https://doi.org/10.1523/JNEUROSCI.4652-10.2011

Hein, T. P., & Herrojo Ruiz, M. (2022). State anxiety alters the neural oscillatory correlates of predictions and prediction errors during reward-based learning. *NeuroImage*, *249*, 118895. https://doi.org/10.1016/j.neuroimage.2022.118895

Heinbockel, H., Wagner, A. D., & Schwabe, L. (2024). Post-retrieval stress impairs subsequent memory depending on hippocampal memory trace reinstatement during reactivation. *Science Advances*, *10*(18), eadm7504. https://doi.org/10.1126/sciadv.adm7504

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

Hermans, E. J., Battaglia, F. P., Atsak, P., de Voogd, L. D., Fernández, G., & Roozendaal, B. (2014). How the amygdala affects emotional memory by altering brain network properties. *Neurobiology of Learning and Memory*, *112*, 2–16. https://doi.org/10.1016/j.nlm.2014.02.005

Herry, C., Bach, D. R., Esposito, F., Salle, F. D., Perrig, W. J., Scheffler, K., Lüthi, A., & Seifritz, E. (2007). Processing of Temporal Unpredictability in Human and Animal Amygdala. *Journal of Neuroscience*, *27*(22), Article 22. https://doi.org/10.1523/JNEUROSCI.5218-06.2007

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1075. https://doi.org/10.1037/xlm0000503

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679. https://doi.org/10.1037/0033-295X.109.4.679

Huang, Q., Xiao, Z., Yu, Q., Luo, Y., Xu, J., Qu, Y., Dolan, R., Behrens, T., & Liu, Y. (2024). Replay-triggered brain-wide activation in humans. *Nature Communications*, *15*(1), 7185. https://doi.org/10.1038/s41467-024-51582-5

Huang, R., & Clewett, D. (2024). The Locus Coeruleus: Where Cognitive and Emotional Processing Meet the Eye. In M. H. Papesh & S. D. Goldinger (Eds.), *Modern Pupillometry: Cognition, Neuroscience, and Practical Applications* (pp. 3–75). Springer International Publishing. https://doi.org/10.1007/978-3-031-54896-3_1

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., Den Ouden, H. E. M., & Stephan, K. E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*, *80*(2), 519–530. https://doi.org/10.1016/j.neuron.2013.09.009

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

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), Article 7. https://doi.org/10.1038/s41562-019-0597-3

Jannati, A., Oberman, L. M., Rotenberg, A., & Pascual-Leone, A. (2023). Assessing the mechanisms of brain plasticity by transcranial magnetic stimulation. *Neuropsychopharmacology*, *48*(1), 191–208. https://doi.org/10.1038/s41386-022-01453-8

Jordan, R., & Keller, G. B. (2023). The locus coeruleus broadcasts prediction errors across the cortex to promote sensorimotor plasticity. *ELife*, *12*, RP85111. https://doi.org/10.7554/eLife.85111

Kalbe, F., Bange, S., Lutz, A., & Schwabe, L. (2020). Expectancy Violation Drives Memory Boost for Stressful Events. *Psychological Science*, *31*(11), 1409–1421. https://doi.org/10.1177/0956797620958650

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. (2022a). On the search for a selective and retroactive strengthening of memory: Is there evidence for category-specific behavioral tagging? *Journal of Experimental Psychology: General*, *151*(1), 263–284. https://doi.org/10.1037/xge0001075

Kalbe, F., & Schwabe, L. (2022b). 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

Kalfaoğlu, Ç., Stafford, T., & Milne, E. (2018). Frontal theta band oscillations predict error correction and posterror slowing in typing. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 69–88. https://doi.org/10.1037/xhp0000417

Kamin, L. J. (1968). Attention-like processes in classical conditioning. In *Symposium on aversive motivation Miami: Bd. No. TR-5* (pp. 9–31). University of Miami Press.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In *Punishment and aversive behavior* (pp. 279–296). Appleton-Century-Crofts.

Keiflin, R., & Janak, P. H. (2015). Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. *Neuron*, *88*(2), Article 2. https://doi.org/10.1016/j.neuron.2015.08.037

Kerzel, D., & Schönhammer, J. (2013). Salient stimuli capture attention and action. *Attention, Perception, & Psychophysics*, *75*(8), 1633–1643. https://doi.org/10.3758/s13414-013-0512-3

Khader, P. H., Jost, K., Ranganath, C., & Rösler, F. (2010). Theta and alpha oscillations during working-memory maintenance predict successful long-term memory encoding. *Neuroscience Letters*, *468*(3), 339–343. https://doi.org/10.1016/j.neulet.2009.11.028

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195. https://doi.org/10.1016/s0165-0173(98)00056-3

Klimesch, W. (2012). α-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Kluen, L. M., Dandolo, L. C., Jocham, G., & Schwabe, L. (2019). Dorsolateral Prefrontal Cortex Enables Updating of Established Memories. *Cerebral Cortex (New York, N.Y.: 1991)*, *29*(10), 4154–4168. https://doi.org/10.1093/cercor/bhy298

Koenigs, M., Barbey, A. K., Postle, B. R., & Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *Journal of Neuroscience*, *29*(47), 14980–14986. https://doi.org/10.1523/JNEUROSCI.3706-09.2009

Kolada, E., Bielski, K., Wilk, M., Rymarczyk, K., Bogorodzki, P., Kazulo, P., Kossowski, B., Wypych, M., Marchewka, A., Kaczmarek, L., Knapska, E., & Szatkowska, I. (2023). The Human Centromedial Amygdala Contributes to Negative Prediction Error Signaling during Appetitive and Aversive Pavlovian Gustatory Learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *43*(17), 3176–3185. https://doi.org/10.1523/JNEUROSCI.0926-22.2023

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165

Kota, S., Rugg, M. D., & Lega, B. C. (2020). Hippocampal theta oscillations support successful associative memory formation. *Journal of Neuroscience*, *40*(49), 9507–9518. https://doi.org/10.1523/JNEUROSCI.0767-20.2020

Krebs, R. M., Schott, B. H., Schütze, H., & Düzel, E. (2009). The novelty exploration bonus and its attentional modulation. *Neuropsychologia*, *47*(11), Article 11. https://doi.org/10.1016/j.neuropsychologia.2009.01.015

Kroes, M. C., & Fernández, G. (2012). Dynamic neural systems enable adaptive, flexible memories. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1646–1666. https://doi.org/10.1016/j.neubiorev.2012.02.014

Krugers, H. J., Hoogenraad, C. C., & Groc, L. (2010). Stress hormones and AMPA receptor trafficking in synaptic plasticity and memory. *Nature Reviews. Neuroscience*, *11*(10), 675–681. https://doi.org/10.1038/nrn2913

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted Cognitive Processes in Major Depression: A Predictive Processing Perspective. *Biological Psychiatry,* *87*(5), 388–398. https://doi.org/10.1016/j.biopsych.2019.07.017

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, *49*(2), 457–470. https://doi.org/10.3758/s13428-016-0715-3

LaBar, K. S. (2003). Emotional memory functions of the human amygdala. *Current Neurology and Neuroscience Reports*, *3*(5), 363–364. https://doi.org/10.1007/s11910-003-0015-z

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*(1), 54–64. https://doi.org/10.1038/nrn1825

Laing, P. A. F., & Dunsmoor, J. E. (2025). Event Segmentation Promotes the Reorganization of Emotional Memory. *Journal of Cognitive Neuroscience*, *37*(1), 110–134. https://doi.org/10.1162/jocn_a_02244

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*, *14*(10), 1250–1252. https://doi.org/10.1038/nn.2904

Lisman, J., Grace, A. A., & Duzel, E. (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends in Neurosciences*, *34*(10), Article 10. https://doi.org/10.1016/j.tins.2011.07.006

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., & others. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, *77*, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026

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

Loof, E. D., Ergo, K., Naert, L., Janssens, C., Talsma, D., Opstal, F. V., & Verguts, T. (2018). Signed reward prediction errors drive declarative learning. *PLOS ONE*, *13*(1), Article 1. https://doi.org/10.1371/journal.pone.0189212

Maes, E. J. P., Sharpe, M. J., Usypchuk, A. A., Lozzi, M., Chang, C. Y., Gardner, M. P. H., Schoenbaum, G., & Iordanova, M. D. (2020). Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. *Nature Neuroscience*, *23*(2), Article 2. https://doi.org/10.1038/s41593-019-0574-1

Maier, M. E., Ernst, B., & Steinhauser, M. (2019). Error-related pupil dilation is sensitive to the evaluation of different error types. *Biological Psychology*, *141*, 25–34. https://doi.org/10.1016/j.biopsycho.2018.12.013

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, *39*, e200. https://doi.org/10.1017/S0140525X15000667

Mather, M., & Sutherland, M. R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspectives on Psychological Science*, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234

Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience*, *10*(5), Article 5. https://doi.org/10.1038/nn1890

McCutcheon, R. A., Abi-Dargham, A., & Howes, O. D. (2019). Schizophrenia, Dopamine and the Striatum: From Biology to Symptoms. *Trends in Neurosciences*, *42*(3), 205–220. https://doi.org/10.1016/j.tins.2018.12.004

McGaugh, J. L. (2000). Memory - A Century of Consolidation. *Science*, *287*(5451), 248–251. https://doi.org/10.1126/science.287.5451.248

McGaugh, J. L. (2018). Emotional arousal regulation of memory consolidation. *Current Opinion in Behavioral Sciences*, *19*, 55–60. https://doi.org/10.1016/j.cobeha.2017.10.003

McGaugh, J. L., & Roozendaal, B. (2002). Role of adrenal stress hormones in forming lasting memories in the brain. *Current opinion in neurobiology*, *12*(2), 205–210. https://doi.org/10.1016/s0959-4388(02)00306-9

McHugh, S. B., Barkus, C., Huber, A., Capitão, L., Lima, J., Lowry, J. P., & Bannerman, D. M. (2014). Aversive Prediction Error Signals in the Amygdala. *Journal of Neuroscience*, *34*(27), Article 27. https://doi.org/10.1523/JNEUROSCI.4465-13.2014

McNamara, C. G., & Dupret, D. (2017). Two sources of dopamine for the hippocampus. *Trends in Neurosciences*, *40*(7), 383–384. https://doi.org/10.1016/j.tins.2017.05.005

McReynolds, J. R., & McIntyre, C. K. (2012). Emotional modulation of the synapse. *Reviews in the Neurosciences*, *23*(5–6). https://doi.org/10.1515/revneuro-2012-0047

Meier, J. K., Staresina, B. P., & Schwabe, L. (2022). Stress diminishes outcome but enhances response representations during instrumental learning. *ELife*, *11*, e67517. https://doi.org/10.7554/eLife.67517

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure & Function*, *214*(5–6), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Metcalfe, J. (2017). Learning from Errors. *Annual Review of Psychology*, *68*(Volume 68, 2017), Article Volume 68, 2017. https://doi.org/10.1146/annurev-psych-010416-044022

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363–386. https://doi.org/10.1037/0033-2909.117.3.363

Möhring, L., & Gläscher, J. (2023). Prediction errors drive dynamic changes in neural patterns that guide behavior. *Cell Reports*, *42*(8), 112931. https://doi.org/10.1016/j.celrep.2023.112931

Moncada, D., Ballarini, F., & Viola, H. (2015). Behavioral tagging: A translation of the synaptic tagging and capture hypothesis. *Neural Plasticity*, *2015*, 650780. https://doi.org/10.1155/2015/650780

Moncada, D., & Viola, H. (2007). Induction of long-term memory by exposure to novelty requires protein synthesis: Evidence for a behavioral tagging. *Journal of Neuroscience*, *27*(28), 7476–7481.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *16*(5), 1936–1947. https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996

Moscovitch, M. (1995). Recovered consciousness: A hypothesis concerning modularity and episodic memory. *Journal of Clinical and Experimental Neuropsychology*, *17*(2), 276–290. https://doi.org/10.1080/01688639508405123

Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., & Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, *117*, 12–22. https://doi.org/10.1016/j.bandl.2010.09.013

Murty, V. P., & Adcock, R. A. (2014). Enriched Encoding: Reward Motivation Organizes Cortical Networks for Hippocampal Detection of Unexpected Events. *Cerebral Cortex*, *24*(8), Article 8. https://doi.org/10.1093/cercor/bht063

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Is survival processing special? *Journal of Memory and Language*, *59*(3), 377–385. https://doi.org/10.1016/j.jml.2008.06.001

Naselaris, T., & Kay, K. N. (2015). Resolving Ambiguities of MVPA Using Explicit Models of Representation. *Trends in Cognitive Sciences*, *19*(10), 551–554. https://doi.org/10.1016/j.tics.2015.07.005

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/nn.3130

Nieh, E. H., Schottdorf, M., Freeman, N. W., Low, R. J., Lewallen, S., Koay, S. A., Pinto, L., Gauthier, J. L., Brody, C. D., & Tank, D. W. (2021). Geometry of abstract learned knowledge in the hippocampus. *Nature*, *595*(7865), 80–84. https://doi.org/10.1038/s41586-021-03652-7

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), Article 3. https://doi.org/10.1016/j.jmp.2008.12.005

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*, *12*(7), 265–272. https://doi.org/10.1016/j.tics.2008.03.006

Noh, E., Herzmann, G., Curran, T., & de Sa, V. R. (2014). Using single-trial EEG to predict and analyze subsequent memory. *NeuroImage*, *84*, 712–723. https://doi.org/10.1016/j.neuroimage.2013.09.028

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1023–1035. https://doi.org/10.1016/j.neubiorev.2009.12.014

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411. https://doi.org/10.1037/0278-7393.28.3.411

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory. *Intelligence*, *31*(2), 167–193. https://doi.org/10.1016/S0160-2896(02)00115-0

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron*, *38*(2), Article 2. https://doi.org/10.1016/S0896-6273(03)00169-7

Olmos, A., & Kingdom, F. A. A. (2004). A Biologically Inspired Algorithm for the Recovery of Shading and Reflectance Images. *Perception*, *33*(12), 1463–1473. https://doi.org/10.1068/p5321

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

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

Papalini, S., Beckers, T., & Vervliet, B. (2020). Dopamine: From prediction error to psychotherapy. *Translational Psychiatry*, *10*(1), 1–13. https://doi.org/10.1038/s41398-020-0814-x

Pape, H.-C., & Pare, D. (2010). Plastic Synaptic Networks of the Amygdala for the Acquisition, Expression, and Extinction of Conditioned Fear. *Physiological Reviews*, *90*(2), 419–463. https://doi.org/10.1152/physrev.00037.2009

Payne, L., & Sekuler, R. (2014). The Importance of Ignoring: Alpha Oscillations Protect Selectivity. *Current Directions in Psychological Science*, *23*(3), 171–177. https://doi.org/10.1177/0963721414529145

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*(2), 198–202. https://doi.org/10.1016/j.conb.2004.03.015

Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., & Mendelsohn, A. (2018). Knowledge acquisition is governed by striatal prediction errors. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-03992-5

Pittig, A., Heinig, I., Goerigk, S., Richter, J., Hollandt, M., Lueken, U., Pauli, P., Deckert, J., Kircher, T., Straube, B., & others. (2023). Change of Threat Expectancy as Mechanism of Exposure-Based Psychotherapy for Anxiety Disorders: Evidence From 8,484 Exposure Exercises of 605 Patients. *Clinical Psychological Science*, *11*(2), 199–217. https://doi.org/10.1177/21677026221101379

Pupillo, F., & Bruckner, R. (2023). Signed and unsigned effects of prediction error on memory: Is it a matter of choice? *Neuroscience and Biobehavioral Reviews*, *153*, 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*(1), Article 1. https://doi.org/10.1038/s41539-023-00166-x

Putica, A., Felmingham, K. L., Garrido, M. I., O'Donnell, M. L., & Van Dam, N. T. (2022). A predictive coding account of value-based learning in PTSD: Implications for precision treatments. *Neuroscience & Biobehavioral Reviews*, 138, 104704. https://doi.org/10.1016/j.neubiorev.2022.104704

Queirazza, F., Fouragnan, E., Steele, J. D., Cavanagh, J., & Philiastides, M. G. (2019). Neural correlates of weighted reward prediction error during reinforcement learning classify response to cognitive behavioral therapy in depression. *Science Advances*, *5*(7), eaav4962. https://doi.org/10.1126/sciadv.aav4962

Reisberg, D., & Hertel, P. (Eds.). (2004). *Memory and Emotion*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195158564.001.0001

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory: Vol. 2*.

Ridderinkhof, K. R., Ramautar, J. R., & Wijnen, J. G. (2009). To PE or not to PE: A P3-like ERP component reflecting the processing of response errors. *Psychophysiology*, *46*(3), 531–538. https://doi.org/10.1111/j.1469-8986.2009.00790.x

Ripollés, P., Ferreri, L., Mas-Herrero, E., Alicart, H., Gómez-Andrés, A., Marco-Pallares, J., Antonijoan, R. M., Noesselt, T., Valle, M., Riba, J., & Rodriguez-Fornells, A. (2018). Intrinsically regulated learning is modulated by synaptic dopamine signaling. *ELife*, *7*, e38113. https://doi.org/10.7554/eLife.38113

Roozendaal, B., & Hermans, E. J. (2017). Norepinephrine effects on the encoding and consolidation of emotional memory: Improving synergy between animal and human studies. *Current Opinion in Behavioral Sciences, 14,* 115–122. https://doi.org/10.1016/j.cobeha.2017.02.001

Roozendaal, B., Okuda, S., Van der Zee, E. A., & McGaugh, J. L. (2006). Glucocorticoid enhancement of memory requires arousal-induced noradrenergic activation in the basolateral amygdala. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(17), 6741–6746. https://doi.org/10.1073/pnas.0601874103

Rosenbaum, G. M., Grassie, H. L., & Hartley, C. A. (2022). Valence biases in reinforcement learning shift across adolescence and modulate subsequent memory. *ELife*, *11*, e64620. https://doi.org/10.7554/eLife.64620

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*, 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*, 104269. https://doi.org/10.1016/j.cognition.2020.104269

Roux, F., Parish, G., Chelvarajah, R., Rollings, D. T., Sawlani, V., Hamer, H., Gollwitzer, S., Kreiselmeyer, G., ter Wal, M. J., Kolibius, L., Staresina, B. P., Wimber, M., Self, M. W., & Hanslmayr, S. (2022). Oscillations support short latency co-firing of neurons during human episodic memory formation. *ELife*, *11*, e78109. https://doi.org/10.7554/eLife.78109

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience*, *17*(11), Article 11. https://doi.org/10.1038/nn.3832

Rubínová, E., & Kontogianni, F. (2023). Sources and destinations of misattributions in recall of instances of repeated events. *Memory & Cognition*, *51*(1), 188–202. https://doi.org/10.3758/s13421-022-01300-7

Sambrook, T. D., & Goslin, J. (2014). Mediofrontal event-related potentials in response to positive, negative and unsigned prediction errors. *Neuropsychologia*, *61*, 1–10. https://doi.org/10.1016/j.neuropsychologia.2014.06.004

Schacter, D. L., & Addis, D. R. (2007). 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

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492.

Schimmelpfennig, J., Topczewski, J., Zajkowski, W., & Jankowiak-Siuda, K. (2023). The role of the salience network in cognitive and affective deficits. *Frontiers in Human Neuroscience*, *17*, 1133367.

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

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, *57*, 87–115. https://doi.org/10.1146/annurev.psych.56.091103.070229

Schultz, W. (2016). Dopamine reward prediction-error signalling: A two-component response. *Nature Reviews Neuroscience*, *17*(3), Article 3. https://doi.org/10.1038/nrn.2015.26

Schultz, W. (2017). Reward prediction error. *Current Biology*, *27*(10), Article 10. https://doi.org/10.1016/j.cub.2017.02.064

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schulz, P., Schlotz, W., & Becker, P. (2004). *Trier inventory for chronic stress*. Hogrefe.

Schwiedrzik, C. M., & Freiwald, W. A. (2017). High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. *Neuron*, *96*(1), 89-97.e4. https://doi.org/10.1016/j.neuron.2017.09.007

Seligman, M. E. P., Maier, S. F., & Solomon, R. L. (1971). CHAPTER 6—Unpredictable and Uncontrollable Aversive Events. In F. R. Brush (Ed.), *Aversive Conditioning and Learning* (pp. 347–400). Academic Press. https://doi.org/10.1016/B978-0-12-137950-6.50011-0

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., & Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*(6992), Article 6992. https://doi.org/10.1038/nature02581

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: An ecological approach to brain and behavior research. *Perspectives on Psychological Science*, *14*(5), 841–859. https://doi.org/10.1177/1745691619856350

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., Niv, Y., & Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience*, *20*(5), Article 5. https://doi.org/10.1038/nn.4538

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences*, *14*(10), 464–472. https://doi.org/10.1016/j.tics.2010.08.002

Siebner, H. R., Hartwigsen, G., Kassuba, T., & Rothwell, J. C. (2009). How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *45*(9), 1035–1042. https://doi.org/10.1016/j.cortex.2009.02.007

Silvetti, M., Castellar, E. N., Roger, C., & Verguts, T. (2014). Reward expectation and prediction error in human medial frontal cortex: An EEG study. *NeuroImage*, *84*, 376–382. https://doi.org/10.1016/j.neuroimage.2013.08.058

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 of the United States of America*, *118*(51), e2117625118. https://doi.org/10.1073/pnas.2117625118

Smulders, F. T. Y., Ten Oever, S., Donkers, F. C. L., Quaedflieg, C. W. E. M., & van de Ven, V. (2018). Single-trial log transformation is optimal in frequency analysis of resting EEG alpha. *The European Journal of Neuroscience*, *48*(7), 2585–2598. https://doi.org/10.1111/ejn.13854

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event Boundaries Trigger Rapid Memory Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. *Current Biology: CB*, *27*(22), 3499-3504.e4. https://doi.org/10.1016/j.cub.2017.09.057

Spielberger, C. D., Gorsuch, R. L., & Luchene, R. E. (1970). *The State-Trait Anxiety Inventory*. Consulting Psychology Press.

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89. https://doi.org/10.1016/j.tics.2013.12.001

Stanek, J. K., Dickerson, K. C., Chiew, K. S., Clement, N. J., & Adcock, R. A. (2019). Expected Reward Value and Reward Uncertainty Have Temporally Dissociable Effects on Memory Formation. *Journal of Cognitive Neuroscience*, *31*(10), Article 10. https://doi.org/10.1162/jocn_a_01411

Staresina, B. P., Alink, A., Kriegeskorte, N., & Henson, R. N. (2013). Awake reactivation predicts memory in humans. *Proceedings of the National Academy of Sciences*, *110*(52), 21159–21164. https://doi.org/10.1073/pnas.1311989110

Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Current Biology*, *23*(12), 1101–1106. https://doi.org/10.1016/j.cub.2013.04.074

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*(7), Article 7. https://doi.org/10.1038/nn.3413

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), Article 1. https://doi.org/10.1007/BF00115009

Takeuchi, T., Duszkiewicz, A. J., & Morris, R. G. M. (2014). The synaptic plasticity and memory hypothesis: Encoding, storage and persistence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1633), Article 1633. https://doi.org/10.1098/rstb.2013.0288

Takeuchi, T., Duszkiewicz, A. J., Sonneborn, A., Spooner, P. A., Yamasaki, M., Watanabe, M., Smith, C. C., Fernández, G., Deisseroth, K., & Greene, R. W. (2016). Locus coeruleus and dopaminergic consolidation of everyday memory. *Nature*, *537*(7620), 357–362.

Talmi, D. (2013). Enhanced Emotional Memory: Cognitive and Neural Mechanisms. *Current Directions in Psychological Science*, *22*(6), 430–436. https://doi.org/10.1177/0963721413498893

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience*, *33*(19), 8264–8269. https://doi.org/10.1523/JNEUROSCI.5695-12.2013

Tambini, A., & Davachi, L. (2019). Awake Reactivation of Prior Experiences Consolidates Memories and Biases Cognition. *Trends in Cognitive Sciences*, *23*(10), 876–890. https://doi.org/10.1016/j.tics.2019.07.008

Tambini, A., & D'Esposito, M. (2020). Causal contribution of awake post-encoding processes to episodic memory consolidation. *Current Biology*, *30*(18), 3533–3543. https://doi.org/10.1016/j.cub.2020.06.063

Tambini, A., Nee, D. E., & D'Esposito, M. (2018). Hippocampal-targeted theta-burst stimulation enhances associative memory formation. *Journal of Cognitive Neuroscience*, *30*(10), 1452–1472. https://doi.org/10.1162/jocn_a_01300

Teyler, T. J., & Rudy, J. W. (2007). The hippocampal indexing theory and episodic memory: Updating the index. *Hippocampus*, *17*(12), 1158–1169. https://doi.org/10.1002/hipo.20350

Trapp, S., O'Doherty, J. P., & Schwabe, L. (2018). Stressful events as teaching signals for the brain. *Trends in Cognitive Sciences*, *22*(6), 475–478. https://doi.org/10.1016/j.tics.2018.03.007

Treder, M. S. (2020). MVPA-light: A classification and regression toolbox for multi-dimensional data. *Frontiers in Neuroscience*, *14*, 289. https://doi.org/10.3389/fnins.2020.00289

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*, *1*(381–403), 1. Cambridge, MA: Academic Press.

Turan, G., Spiertz, V., Bein, O., Shing, Y. L., & Nolden, S. (2025). Unexpected Twists: Electrophysiological Correlates of Encoding and Retrieval of Events Eliciting Prediction Error. *Psychophysiology*, *62*(1), e14752. https://doi.org/10.1111/psyp.14752

Tzovara, A., Korn, C. W., & Bach, D. R. (2018). Human Pavlovian fear conditioning conforms to probabilistic learning. *PLOS Computational Biology*, *14*(8), Article 8. https://doi.org/10.1371/journal.pcbi.1006243

Van Kesteren, M. T., 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

Vogel, S., Kluen, L. M., Fernández, G., & Schwabe, L. (2018). Stress leads to aberrant hippocampal involvement when processing schema-related information. *Learning & Memory*, *25*(1), 21–30. https://doi.org/10.1101/lm.046003.117

Vogt, B. A., Finch, D. M., & Olson, C. R. (1992). Functional heterogeneity in cingulate cortex: The anterior executive and posterior evaluative regions. *Cerebral Cortex (New York, N.Y.: 1991)*, *2*(6), 435–443. https://doi.org/10.1093/cercor/2.6.435-a

Wagatsuma, A., Okuyama, T., Sun, C., Smith, L. M., Abe, K., & Tonegawa, S. (2018). Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proceedings of the National Academy of Sciences*, *115*(2). https://doi.org/10.1073/pnas.1714082115

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, *3*(4), 255–274. https://doi.org/10.3758/cabn.3.4.255

Wang, W.-C., Wing, E. A., Murphy, D. L. K., Luber, B. M., Lisanby, S. H., Cabeza, R., & Davis, S. W. (2018). Excitatory TMS Modulates Memory Representations. *Cognitive Neuroscience*, *9*(3–4), 151–166. https://doi.org/10.1080/17588928.2018.1512482

Watabe-Uchida, M., Eshel, N., & Uchida, N. (2017). Neural Circuitry of Reward Prediction Error. *Annual Review of Neuroscience*, *40*(Volume 40, 2017), Article Volume 40, 2017. https://doi.org/10.1146/annurev-neuro-072116-031109

Weissman, D. H., Gopalakrishnan, A., Hazlett, C. J., & Woldorff, M. G. (2005). Dorsal Anterior Cingulate Cortex Resolves Conflict from Distracting Stimuli by Boosting Attention toward Relevant Events. *Cerebral Cortex*, *15*(2), 229–237. https://doi.org/10.1093/cercor/bhh125

White, S. F., Geraci, M., Lewis, E., Leshin, J., Teng, C., Averbeck, B., Meffert, H., Ernst, M., Blair, J. R., Grillon, C., & Blair, K. S. (2017). Prediction Error Representation in Individuals With Generalized Anxiety Disorder During Passive Avoidance. *American Journal of Psychiatry,* *174*(2), 110–117. https://doi.org/10.1176/appi.ajp.2016.15111410

White, T. P., Jansen, M., Doege, K., Mullinger, K. J., Park, S. B., Liddle, E. B., Gowland, P. A., Francis, S. T., Bowtell, R., & Liddle, Peter. F. (2013). Theta power during encoding predicts subsequent-memory performance and default mode network deactivation. *Human Brain Mapping*, *34*(11), 2929–2943. https://doi.org/10.1002/hbm.22114

Wianda, E., & Ross, B. (2019). The roles of alpha oscillation in working memory retention. *Brain and Behavior*, *9*(4), e01263. https://doi.org/10.1002/brb3.1263

Winkler, C. D., Pittig, A., Phillips, L. J., & Felmingham, K. L. (2025). Associations among threat prediction error, prediction change, and anxiety during an exposure therapy analogue in adults with healthy to clinical social anxiety. *Behaviour Research and Therapy,* 187, 104709. https://doi.org/10.1016/j.brat.2025.104709

Woods, A. J., Antal, A., Bikson, M., Boggio, P. S., Brunoni, A. R., Celnik, P., Cohen, L. G., Fregni, F., Herrmann, C. S., Kappenman, E. S., Knotkova, H., Liebetanz, D., Miniussi, C., Miranda, P. C., Paulus, W., Priori, A., Reato, D., Stagg, C., Wenderoth, N., & Nitsche, M. A. (2016). A technical guide to tDCS, and related non-invasive brain stimulation tools. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *127*(2), 1031–1048. https://doi.org/10.1016/j.clinph.2015.11.012

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970

Yau, J. O.-Y., & McNally, G. P. (2023). The Rescorla-Wagner model, prediction error, and fear learning. *Neurobiology of Learning and Memory*, *203*, 107799. https://doi.org/10.1016/j.nlm.2023.107799

Yu, W., & Davachi, L. (2025). *Weak and strong memories are equally reactivated during counterfactual learning, but only weak memories are modified*. OSF. https://doi.org/10.31234/osf.io/js6qd

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

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

Zerbes, G., Kausche, F. M., Müller, J. C., Wiedemann, K., & Schwabe, L. (2019). Glucocorticoids, Noradrenergic Arousal, and the Control of Memory Retrieval. *Journal of Cognitive Neuroscience*, *31*(2), 288–298. https://doi.org/10.1162/jocn_a_01355

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Erklärung gemäß *(bitte Zutreffendes ankreuzen)*

☐ **§ 4 (1c) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**

☒ **§ 5 (4d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich,

Kaja Loock (Vorname, Nachname),

dass ich mich an einer anderen Universität oder Fakultät noch keiner Doktorprüfung unterzogen oder mich um Zulassung zu einer Doktorprüfung bemüht habe.

Hamburg, 04.06.25
Ort, Datum


Unterschrift

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Eidesstattliche Erklärung nach *(bitte Zutreffendes ankreuzen)*

☐ **§ 7 (4) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**

☒ **§ 9 (1c und 1d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich an Eides statt,

1. dass die von mir vorgelegte Dissertation nicht Gegenstand eines anderen Prüfungsverfahrens gewesen oder in einem solchen Verfahren als ungenügend beurteilt worden ist.
2. dass ich die von mir vorgelegte Dissertation selbst verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die wörtlich oder inhaltlich übernommenen Stellen habe ich als solche kenntlich gemacht.

Hamburg, 04.06.25
Ort, Datum


Unterschrift