Contextual Prediction Errors Reorganize Episodic Memories in Time

Fahd Yazin[1], Moumita Das[1], Arpan Banerjee[1], Dipanjan Roy[1]

[1]Cognitive Brain Dynamics Lab, National Brain Research Centre, NH 8, Manesar, Gurgaon 122052, India

Correspondence: dipanjan.nbrc@gov.in

Abstract

Episodic memories are contextual experiences ordered in time. Despite knowing its adaptive nature, whether our memories for events can be altered in time is not adequately addressed. The role of prediction errors (PEs) in strengthening declarative memory is well established but has not been investigated in the time dimension of episodic memory. We tested this dual possibility out using a 3-day movie viewing paradigm, wherein memorized events are subjected to expectation violations and later tested for sequence memory. Our results revealed that contextual PEs fundamentally weaken the older memorized sequence of mispredicted events. Interestingly, the newer events having the PEs in them, had comparable memory strength as control events with repeated viewings but without having any PEs. A hierarchical drift diffusion model indicated that a change in decision criterion rather than drift rate accounted for the behavioural measures of memory strength. In summary, these results suggest that the temporal arrangement of naturalistic episodic memories are reorganized by contextual prediction errors.

Introduction

Imagine while entering the office, you see your favorite actor sitting in your chair, much to your surprise. The memory of such an event would be harder to forget than others.

One reason for strong memory consolidation of this event is the low expectation of such events to occur within the given context. This displays the dependency of our day-to-day memories on the underlying contextual (Howard et al 2002, Yonelinas et al 2019, Polyn et al 2009, Hupbach et al 2011, Davachi et al 2006) and predictive processes (Smith et al 2013, Henson et al 2010, Mullaly et al 2014, Bar et al 2009 , Lisman et al 2009). Episodes or events being the canonical components of episodic memory (Tulving 1983), are not only marked by a clear beginning and an end, but also relatable to each other temporally (Tulving 1983). As such, our memories are organized sequentially in contexts that evolves in time. But whether and how unpredicted events can affect this temporal sequencing of our experienced memories is something that surprisingly remains largely unexplored.

In real life situations, time represents the sequential arrangement of the memories as and how they occurred in their respective contexts. A context in its simplest form is described as any aspect of the episode that binds its constituent elements together, be it spatial, temporal or conceptual (Yonelinas et al 2019, Polyn et al 2009, Hupbach et al 2011, Davachi et al 2006, Tulving 1983). A recent perspective (Yonelinas et al 2019) places contextual binding account at the forefront of cognitive theories of episodic memory with the hippocampus playing a central role in item and context binding. One may ask what could possibly happen when the same context is shared by two different memories? The Temporal Context Model (TCM) (Howard et al 2002, Sederberg et al 2011) of memory posits the two memories would be sharing the same temporal context, hence creating source confusion during memory recall. This does not necessarily result in memory trace weakening however recent evidence has shown mispredicted items in

the same context are in fact weakened (Kim et al 2014). Previous studies have also pointed out that competing memories in the same context interfere with each other, and cause intrusions. This important line of work termed as reconsolidation (Hupbach et al 2011, Alberini et al 2013, Hupbach et al 2007, Hupbach et al 2009, Lee et al 2017), thought to be mediated by prediction error (Exton-McGuinness et al 2015 , Fernandez et al 2016, Sinclair et al 2018) and is demonstrated by an asymmetric intrusion of memories in the same context.

Predictions are a hallmark of episodic memory recall (Smith et al 2013, Henson et al 2010, Lisman et al 2009). Therefore, whenever a context is re-experienced, the sequence of episodes is automatically predicted (Smith et al 2013, Bar et al 2009, Lisman et al 2009, Sinclair et al 2018). Prediction errors resulting from the violation of these predictions have been shown to influence declarative memory by strengthening incidental encoding (Jang et al 2019,), semantic memory acquisition (Pine et al 2018), paired association learning (De Loof et al 2018, Greve et al 2017) and by playing a role in reconsolidation (Sinclair et al 2018). How they interact with declarative memories in the sequential aspect has not been investigated. This sets up an intriguing question in how they affect episodic memory sequences ordered in time.

Our overarching hypothesis is that these contextual prediction errors would fundamentally alter the memorized sequence of events, such that the newer updated memory sequences are prioritized during recall over older outdated ones. From the perspective of predictive coding framework (Friston et al 2005, Clark et al 2013), new surprising information drives learning and the ensuing sequential order would be

strengthened over older encoded sequential information, so as to minimize the PEs in the future.

To systematically validate our hypothesis, we employed a 3-day paradigm with naturalistic movies strategically edited into contextually different events containing multiple segments. There are substantial gaps about important behavioral questions like how the sequences themselves gets strengthened or weakened, ease of access, subjective confidence during episodic sequence recall. We also wanted to understand the crucial relationship between prediction error and the evidence accumulation process during sequence recall. After watching the movies on *Day1*, participants saw an incongruent version of the movies with different but contextually fitting segments added onto the original event during *Day2*. A sequence memory test was conducted on *Day3* which allowed us to examine the temporal association strength between adjacent memory segments. Our key finding is that contextual prediction errors strengthens the newer memory sequences in time while weakening the order of previously encoded sequences. This enhanced performance results from increased evidence quality in memories rather than a faster evidence accumulation process while remembering. Additionally, this effect is not due to novel associations between segments. Our findings potentially reveal how prediction errors are key in determining how memories are organized in time.

Materials and Methods

Participants

Participants were 24 healthy young right-handed adults (16 males, 8 females, ages 22-35, *Mean*: 27, *SD*: 3.3). The study was approved by the Institutional Human Ethics Committee of National Brain Research Centre, India (IHEC, NBRC). All participants signed informed consent in consonance with the Declaration at Helsinki, declared normal or normal to corrected vision and no history of hearing problems, neurological and neuropsychiatric disorders.

Materials

Participants saw 2 short films on *Day 1*. The selected films had engaging plots with multiple different contexts, spatially and conceptually with slightly surprising storylines. Secondly, being short films, they had no famous or otherwise identifiable actors whom the participants could easily recognize. This control on *prior* memory formation was a necessary step in our study. We wanted the natural scenes to appear as a 'first impression' in which aspects like characters and storyline remain unknown beforehand. Third, being an episodic sequence encoding task, it was imperative that the performance should be measured objectively. In conventional study designs with naturalistic stimuli, authors typically measure memory performance based on questionnaires, interviews etc. In order to have the same methodological rigor of controlled memory tasks we needed to edit the movie and scenes appropriately to suit our goal (see supplementary materials).

We divided the movie into different events (Kurby et al 2008), each event defined by a distinct underlying context conceptually/spatially. Each event had multiple segments in them. For example, a series of segments occurring in a kitchen at night, in a bar, by the car park all constitute separate events. Each segment is defined by the

actions/interactions happening between entities (people) concerning the underlying event. Thus, we divided the whole movie into different contextual events with different segments making up an event (**Figure 1**). (See supplementary materials for more details).

Both movies had 15 events where prediction errors occurred. Each event had 4 segments. Each segment was 7s long, with a 1s interstimulus interval, where a black blank screen is presented between every segment. This also allowed us to reorient attention between the scenes.

Both movies were taken from YouTube. The copyright for the scientific usage and publishing of both these movies were obtained from the creators. First movie was titled *Betrayed*, about a wife finding out her husband's affair with her best friend, with the latter trying to cover it up. The second one, titled, *The Man and the Thief* was about a young man helping out a girl at a railway station only to be deceived by her in the end when she steals his wallet.

Design and Sequence Encoding

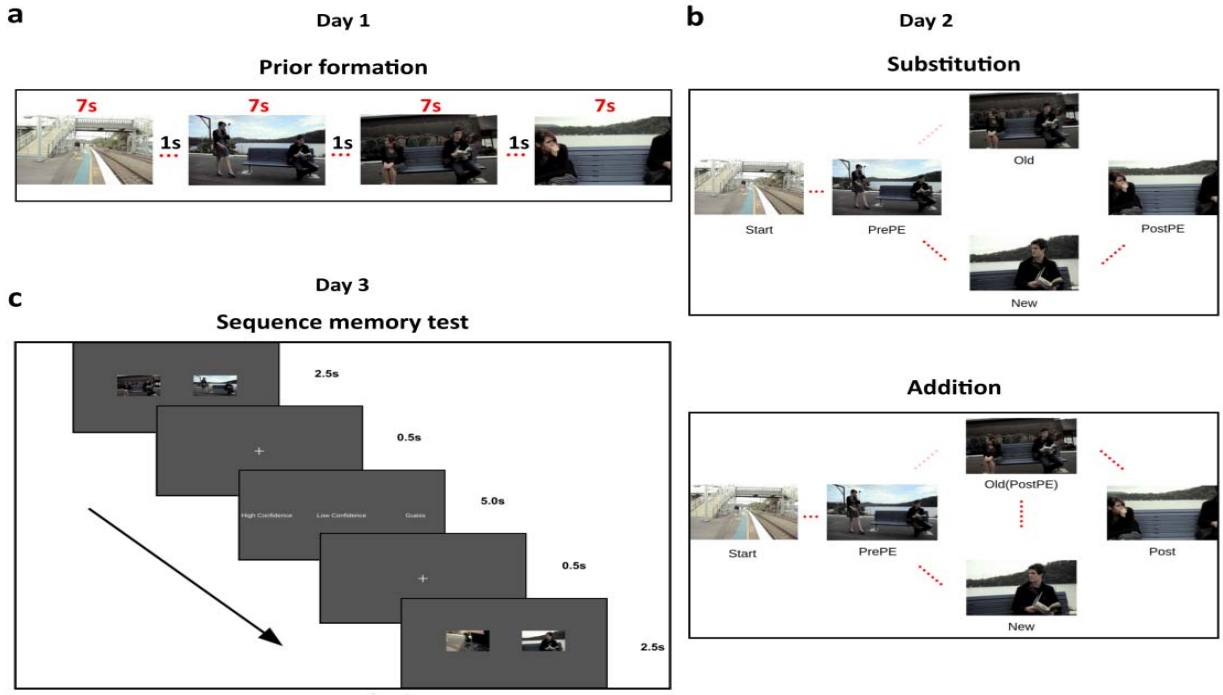

Figure 1: Experiment Design. (**a**) Example of an event seen on *Day1*. Each segment is 7s with 1s blank screen in between (not shown). (**b**) *Day2* conditions. Faded red dots represent predicted sequences while bold red dots represent actual sequences viewed. Representative event of this Day in Substitution(top), Addition (bottom). c) Schematic of *Day3* Sequence memory test block.

Day 1: Formation of Priors

All participants saw both movies during *Day1* prior formation. The order of the movies seen was counterbalanced across participants. Each event had 4 segments on *Day1,* for both movies. They were named (in order of their appearance in the event) *Start*, *PrePE*, *Old* and *PostPE*. Each event began with the *Start* segment.

Day 2: Inducing Prediction Errors

To induce PEs, the original versions of the movie were edited into 3 versions. First version being the *Prior* was shown to all the participants on *Day1*. The second version was the *Substitution* version with a segment substituting another within an event and the third version was the *Addition* version with an additional segment inserted in an event. These scenes termed *New* segments were contextually fitting so as to avoid making abrupt transitions while watching. Importantly, because of this, we could reliably introduce contextual PEs in both movies. Furthermore, on *Day2*, participants saw the two movies in either of these two versions.

First segment of every event was termed as *Start* since it was always the first event to be shown in a segment and was unchanged across both days' viewings (for this reason, it acted as Control). This also enabled the subjects to predict the upcoming segment in that specific event on *Day2*. Second segment was termed the Pre-Prediction Error segment (*PrePE* henceforth) since the prediction violation segment always happened after this. This segment was also similar to a Control in that it did not deviate from prior viewing as well. The segment that followed *PrePE* was termed as *Old* because this segment was shown on *Day1* but was replaced on *Day2* in the *Substitution* condition. The segment replacing it on *Day2*, the prediction error segment, was labelled as *New*. The segment that occurred in the event right after the *New* segment was termed Post-Prediction Error (*PostPE* henceforth) because it always followed the Prediction error segment.

In the *Addition* condition, *Start* and *PrePE* remained the same, similar to *Substitution*. The Prediction error segment, *New*, occurs after *PrePE*. The *Old* segment as seen on

Day1 which was expected to happen after *PrePE*, was seen after the *New* event hence termed aptly as *PostPE* in this condition.

Thus, in the *Substitution* condition, on *Day2* each event had a *Start* segment, followed by the *PrePE* segment, then the *New* segment inducing prediction error (since subjects were predicting *Old*), and afterwards *PostPE* segment. In the *Addition* condition, the *Old* (which was seen after the *New*), functioned as the *PostPE* segment. We purposely put the PE segments in the middle of events, rather than beginning or end to control for primacy and recency effects, respectively.

Betrayed was 5mins 42s in both *Prior* formation and *Substitution* condition, and 6mins 55s in *Addition* condition. It had 11 different events of which 9 were altered during second days viewing. *The Man and the Thief* had 6 contexts instead with a duration of 3mins 14s in Prior and *Substitution* condition, and 4mins 2s in *Addition*. All the 9 events had 4 segments (termed from start to end: Start, *PrePE*, *Old*, *PostPE*) during Prior viewing. *Old* segment gets replaced by *New* in *Substitution* while in *Addition New* is inserted in between *PrePE* and *Old* incurring a total 5 episodes in one contextual event in this condition. Hence in *Addition*, we deliberately took out a segment from each event in the Prior version only to be added back in *Day2* so as to induce the contextual prediction errors. Participants were instructed to pay full attention to both Days' movies and not to use any specific strategies of encoding. They were asked on *Day2* if they noticed changes in scenes compared to the previous day to confirm for attention for both days' viewings but were not asked to recall them in any way.

Day 3: Sequence Memory Test

On *Day3* participants underwent a sequence memory test. The Sequence Memory test was a two-alternative forced task choice (2AFC) task. A screenshot from each segment was shown besides a screenshot of an adjacent segment within an event and subjects were asked to choose the first one that came in the movie. The stimulus presentation on *Day3* was using PsychoPy software (Pierce et al 2019).

It was designed in such a way that segments exclusive to one day (such as *New* and *Old* in *Substitution* condition) were not seen together to prevent a conflict of decision. Since participants were instructed to encode both Days' movies equally without giving preference to one, there were no discrepancies between choices.

Screenshots of adjacent segments within an event were shown side by side (in normal and reversed order randomly) and participants were asked to indicate which one came first in the sequence. This enabled us to probe into the sequential association strength between those two segments, within an event. We had categorized the association strength between the *Start* segment and *Pre-PE* segment within an event as Control Sequence. *PrePE* segment and *Old* segment (*PrePE-Old*) which reflected the strength of the Older (*Prior)* sequence was termed Old Sequence. *PrePE* segment to *New* segment (*PrePE-New*) was called New Sequence and finally the sequence of *New* segment to *PostPE* (*New-PostPE*) segment as Novel association sequence. Trials were binned across participants into these categories by movies and conditions for analysis.

Trials were shown for 2500ms with 500ms ITI between them. Subjects were asked to answer as accurately and fast as possible. After the response subjects had to choose their confidence on a 3-point scale - High Confidence, Low Confidence and Guess.

There were 180 trials in total. We calculated the accuracy by including only the High Confidence and Low Confidence hits; Guesses were completely excluded from analysis. Confidence performance of High Confidence responses were derived by counting the percentage total of High Confidence hits per total hits (High Confidence + Low Confidence).

Hierarchical Drift-Diffusion Model

We used a Hierarchical version of the DDM to fit the behavioral data (Weicki et al 2013). The DDM (Ratcliff et al 2008) is one of a general class of sequential sampling models aptly suited for reaction time data from two-alternative forced choice paradigms (Ratcliff et al 1998). It operates under the assumption that decisions are made by accumulated evidence from a noisy sensory signal. The decision is made when the evidence crosses a threshold. The main parameters in DDM is a drift rate, a rate of accumulation of evidence and the thresholds of the boundaries for these to cross or evidence required for the decision to be made. Additional components include a response bias, which determines whether the responses towards either boundaries are biased or not depending on the starting position. The total reaction time is assumed to be a combination of processing time required to make that decision, as well as encoding time of the stimulus and the time required for the motor execution response. The latter two are assumed not to vary and are combined as another component called non-decision time. Thus, the DDM gives the response choice depending on which boundary threshold (upper or lower) is crossed and gives the response time as a combination of the total time required to cross these boundaries and the non-decision time involved (Ratcliff et al 2008, Ratcliff et al 1998).

The HDDM toolbox (Weicki et al 2013) in python was used to model the data. Bayesian inference which naturally fits a hierarchical DDM, can be used to not only recover the parameters but more importantly to estimate the uncertainty involved in the model parameters (Weicki et al 2013). They also provide solutions for parameter estimations of individual participants which are assumed to be drawn from a group-level prior distribution and are constrained by it. Furthermore, Bayesian methods are more powerful when the trial numbers are low which is desirable since the usual DDM requires larger datasets. The joint posteriors of all the model parameters are estimated by standard Markov Chain Monte Carlo (MCMC) methods (Gamerman et al 2006). A direct Bayesian inference was performed on the posteriors of different conditions by computing the overlap between their distributions.

In our main model, we allowed both drift rate and boundary parameters to vary with Stimulus, that is *Old* and *New* memory sequences. This was to compare which of the two parameters demonstrated the empirical effect of PEs, which is the reaction times. Drift rate also varied across Confidence. Confidence being High Confidence and Low Confidence as Guesses were discarded from main analysis. Non-decision time parameter was set to vary across *Substitution* and *Addition*. We also generated other models with different combinations of drift rate and boundary varying with Stimulus and Confidence. Models were adjudicated using Deviance Information Criterion (DIC) which penalizes *Addition*al model complexity (Spiegelhalter et al 2002). DIC values above 10 are generally considered significant, with the lowest having best goodness-of-fit. (Suppl **Figure 2**)

For each model, we used MCMC methods to generate 100,000 samples from the posterior distribution and discarded the first 20,000 samples as burn-in. After visually inspecting the chains and autocorrelation for proper convergence, the Gelman Rubin R-hat statistic was confirmed to be between 98-1.02(Gelman et al 1992). The latter was computed by running the (main) model 3 times and checking within-chain and between-chain variance. All procedures were done as described in the HDDM toolbox in python (Weicki et al 2013).

Results

Behavioral Analysis

In order to determine the sequence memory performance, we categorized the trials into *New* sequences (temporal order judgements between *PrePE* and *New* segments), and *Old* sequences (temporal order judgements between *PrePE* and *Old* segments). All trials were binned into either *Addition* or *Substitution* by participants depending on which version of the movie they watched on *Day2*. Furthermore, we also categorized trials into Novel sequences (temporal order judgements between *New* and *PostPE* segments) in *Substitution* and *Addition* as well.

Subjective confidence in memories is usually teased out via either numerical scales (1-10) or nominal scales (sure, not sure). Here we deployed a version of nominal scale having High confidence, Low confidence and Guess as labels in a 3-point scale to test the strength of subjective opinion. We calculated and compared the percentage number of High Confidence responses given by participants out of the total (High Confidence + Low Confidence) responses in Old and New sequences.

We also did a Bayes factor (BF) analysis on the paired *t-tests* (Rouder et al 2009) using a default Jeffreys Zellner Siow (JZS) prior using the *BayesFactor* package in R.

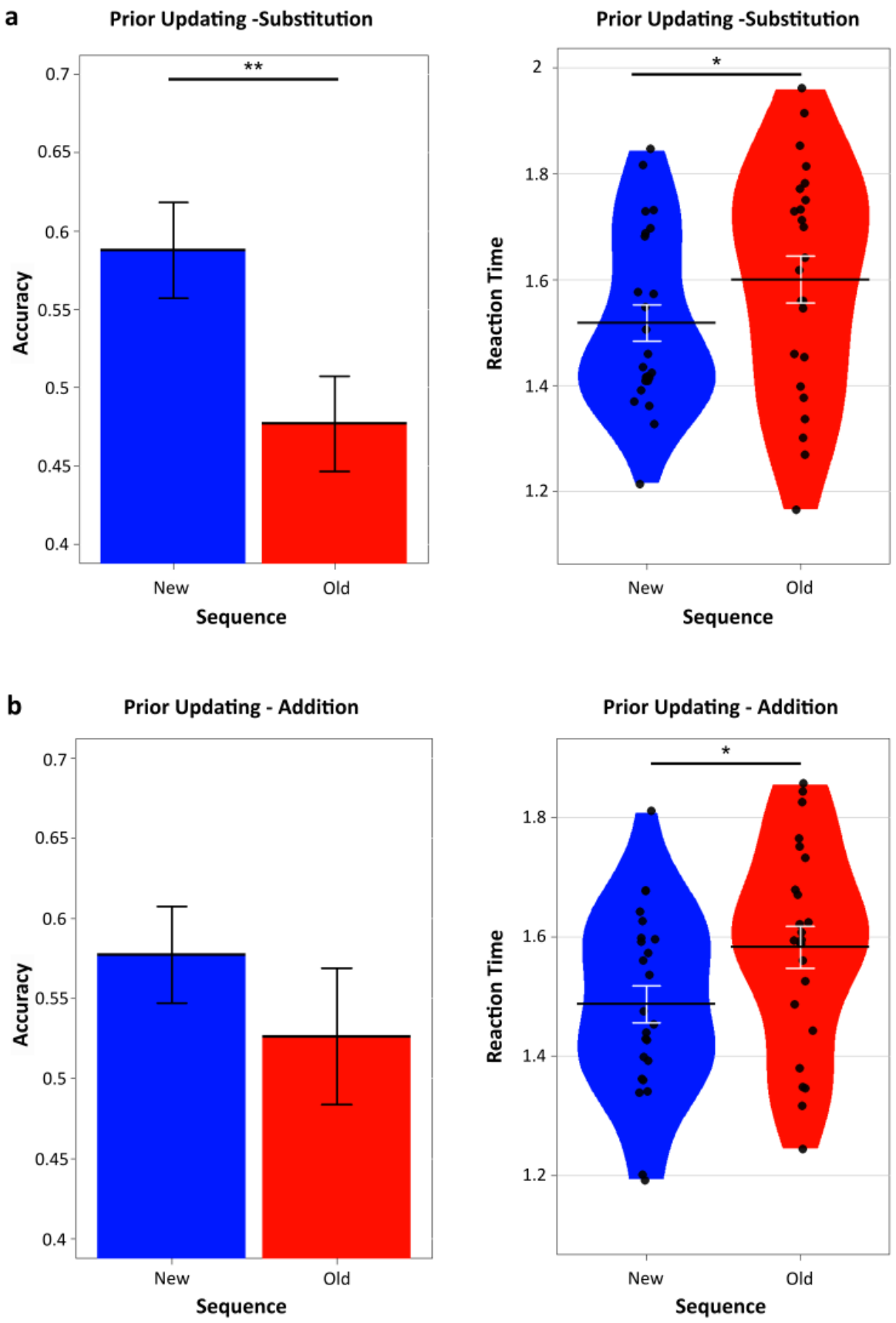
a
Prior Updating -Substitution
**
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Accuracy
New
Old
Sequence
Prior Updating -Substitution
*
2
1.8
1.6
1.4
1.2
Reaction Time
New
Old
Sequence
b
Prior Updating - Addition
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Accuracy
New
Old
Sequence
Prior Updating - Addition
*
1.8
1.6
1.4
1.2
Reaction Time
New
Old
Sequence

Figure 2: Prediction errors affect temporal order of the memory. Accuracy and Reaction times for (**a**) Substitution condition (**b**) Accuracy and Reaction times for Addition condition. Dots represent individual participant performance. Error bars represent standard error of the mean.

Prediction errors drive episodic sequence learning

We hypothesized that the association strength of the mispredicted older sequence would be weaker compared to the newer sequence because of the prediction error. Subsequently, we tested the sequential order memory between *PrePE* segment and *Old* segment with the *PrePE* segment and the *New* segment. Indeed, there was a significant difference of in percentage accuracy between the Old(Old Sequence Mean = 47.69, SE = 3.04) and New sequence (New Sequence Mean = 58.76, SE = 3.05) in *Substitution* condition ($t_{(23)}$ = 3.416, p = .002, 95% CI [4.36 17.77], BF = 16.60, d = .74) (**Figure 2**). There was a significant decrease in reaction time for the New sequence (New Sequence Mean RT = 1517ms, SE = 34ms) compared to Older sequence (Old Sequence Mean RT = 1600ms, SE = 44ms) $t_{(23)}$ = -2.42, p = .02, 95% CI [-151.8 -12.13], BF = 2.39 , d = .42.

Next, we hypothesized that the segments omitted in *Substitution* if re-experienced again, would result in stronger memory activation of those segments. That is, if the mispredicted segment episode is re-experienced, its subsequent reactivation would strengthen its sequential memory. The *Addition* condition was used to test out this specific hypothesis. This allowed us to tease out the specific effect of PE by comparing New and Old Sequences by its interaction with memory reactivation. We found that the group differences in mean accuracy of New (New Sequence Mean = 57.71, SE = 3.02)

and Old (Old Sequence Mean = 52.62, SE = 4.25) sequences did not differ significantly $t_{(23)}$ = 0.906, p = .37, 95% CI [-6.5 16.69], BF = 0.31, d = *.28*. This is not entirely surprising given that in *Addition* condition re-encountering of the segment after the prediction error, in contrast to *Substitution* condition where the segment was omitted. Reaction times however, showed a significant difference (New Sequence Mean = 1486ms, SE = 31ms, Old Sequence Mean = 1582ms, SE = 35ms), similar to the *Substitution* condition $t_{(23)}$ = -2.43, p = .02, 95% CI [-177 -14.32], BF = 2.41, d = *.58.* The effects of prediction errors on the accessibility of the memory sequences is reflected by this result. In other words, the new sequences are always recalled faster compared to the older ones.

Subjective confidence responses did not increase in the newer memory sequences

In *Substitution*, there were no differences observed between the two sequences with respect to their High Confidence responses (New Sequence Mean HC percentage = 85.34, SE =3.6, Old Sequence Mean HC percentage = 81.03, SE = 4.6) $t_{(23)}$ = 1.26, p = 0.218, 95% CI [-0.027 0.113], BF = 0.43, d = .21 (See supplementary material). The result confirms that despite changes in temporal order memory, participants felt similar subjective confidence their responses to both new and old sequences.

Similarly in the *Addition* condition, no increased subjective confidence percentage responses were observed in the Newer sequences over Older sequences (New Sequence Mean HC hits = 85.60, SE = 4.18, Old Sequence Mean HC hits= 78.78, SE = 4.92), $t_{(23)}$ = 1.15, p = 0.26, 95% CI [2.42 15.36], BF = 0.38, d = .30.

Hierarchical drift diffusion modelling

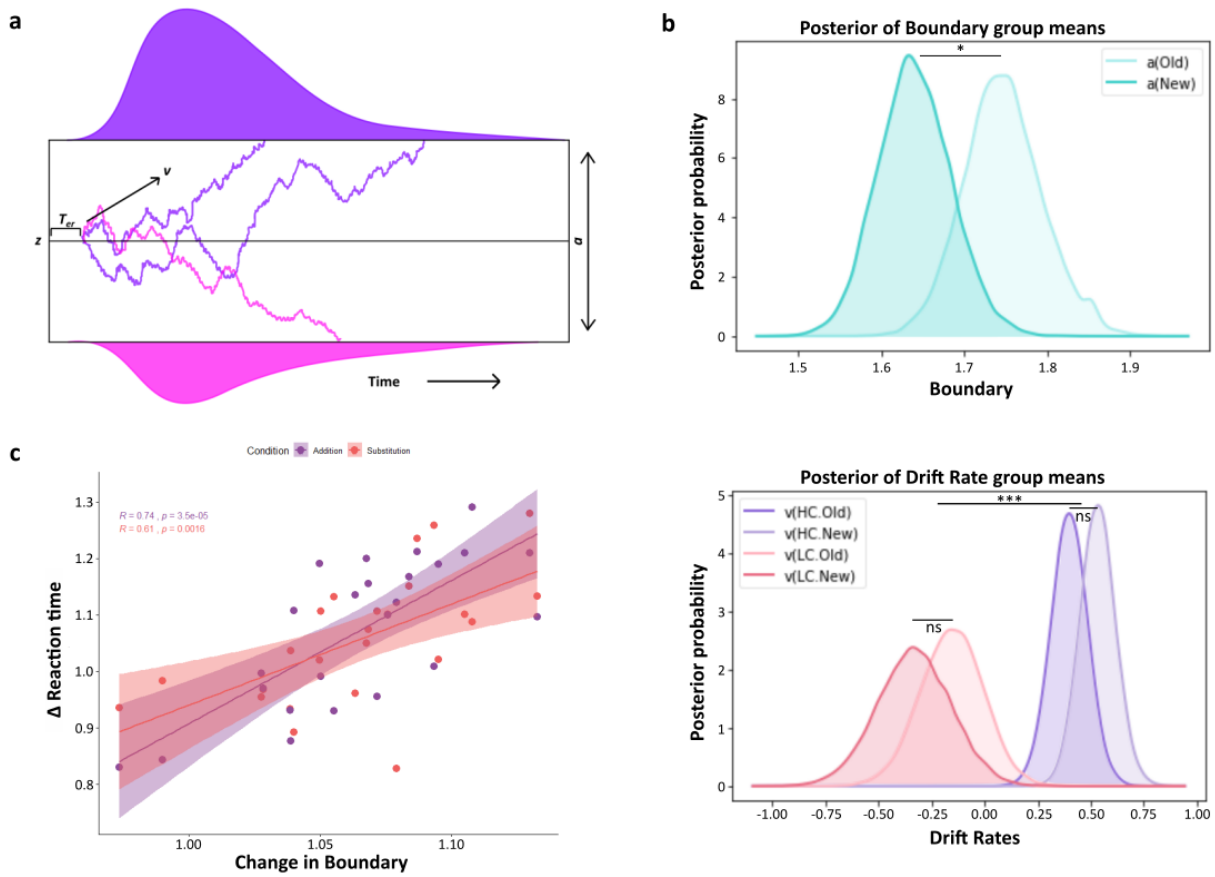

Figure 3: Hierarchical DDM results. (**a**) An illustration of the DDM. The drift rate v reflects the rate of noisy accumulation of evidence till it reaches either of the two boundaries which are separated by a parameter a. The process starts at point z which may or may not have a response bias (not included in the main model) towards either boundaries, which results in the model making the response choices. The response times are a combination of diffusion process and the non-decision time T_{er}, which includes no accumulation. (**b**) Posterior density plots of the group means boundary

parameter showing differences between the Old and New sequences (top). Posterior density plots of the group means drift-rate parameter showing no difference between the Old and New sequences but only between High and Low confidence responses (bottom). (**c**) Change in Boundary condition correlation with change in reaction times in Addition (Purple) and Substitution (Pink). Dots represent individual participants.

Prediction errors cause increased evidence quality, but not increased evidence accumulation rate during recall

Our main hypothesis was to distinguish whether the faster recall of PE memories was due to lower decision threshold or faster drift rates. The best fitting model had both drift rates and boundary allowed to vary with the Stimulus, that is Old and New sequences. Drift rate was also allowed to vary with Confidence measures for each trial response. The HDDM model reliably accounted for reaction times. Moreover, a posterior predictive check was conducted to asses model performance (by generating 100 datasets from the model) which could predict the observed RTs (**Figure 4**).

Significant differences between the Boundary parameter *a* of Old and New sequences explained the corresponding reaction time differences shown in the empirical data ($Boundary_{Old}$ group means = 1.747, 95% credible interval = 1.657-1.846, $Boundary_{New}$ group means = 1.642, 95% credible interval = 1.558-1.730, p = 0.026) (**Figure 3**). Specifically, New sequences with PEs required lower threshold to arrive at a decision compared to Old sequences. In other words, a better quality of evidence allows PE based memories to be recalled faster during making decisions on their temporal order.

Next, we wanted to quantify how much the within-subject reaction time differences between Old and New sequences were driven by the changes in decision boundaries. To this end, we used the ratio of Boundary parameter $a_{(Old)}$/ Boundary parameter $a_{(New)}$ to quantify the change in evidence requirement and correlated this quantity with the ratio of $RT_{(Old)}/RT_{(New)}$ as a measure of change in reaction times for each participant. Significant correlations were found in both *Substitution* (r = 0.61, p = .0016) and *Addition* (r = 0.74, p < .001) revealing how the changes in evidence requirement drives corresponding changes in the reaction times.

We estimated the drift-rate parameter v, to quantify the differences between RTs in Old and New sequences. The drift-rate parameter v, crucially did not show any significant differences between the Old and New sequences in both Low (Drift-rate Low Confidence_{Old} group means = -0.153, 95% credible interval = -0.445 - 0.129, Drift-rate Low Confidence_{New} group means = -0.326, 95% credible interval = -0.666-0.009, p = 0.217) and High Confidence responses (Drift-rate High Confidence_{Old} group means = 0.389, 95% credible interval = 0.221-0.556, Drift-rate High Confidence_{New} group means = 0.52, 95% credible interval = 0.358-0.685, p = 0.867). This is also in line with the behavioral result of participants' subjective confidence judgements which did not significantly differ between Old and New sequences, in both *Substitution* and *Addition*. However, drift-rates did show significant differences between Low Confidence and High Confidence responses, across the Stimulus (Drift-rate High $\text{Confidence}_{New+Old}$ group means = 0.455, Drift-rate Low $\text{Confidence}_{New+Old}$ group means = -0.24, p < .001). This means that rate of evidence accumulation while recalling the temporal order of memory determines the subjective confidence accompanying those responses.

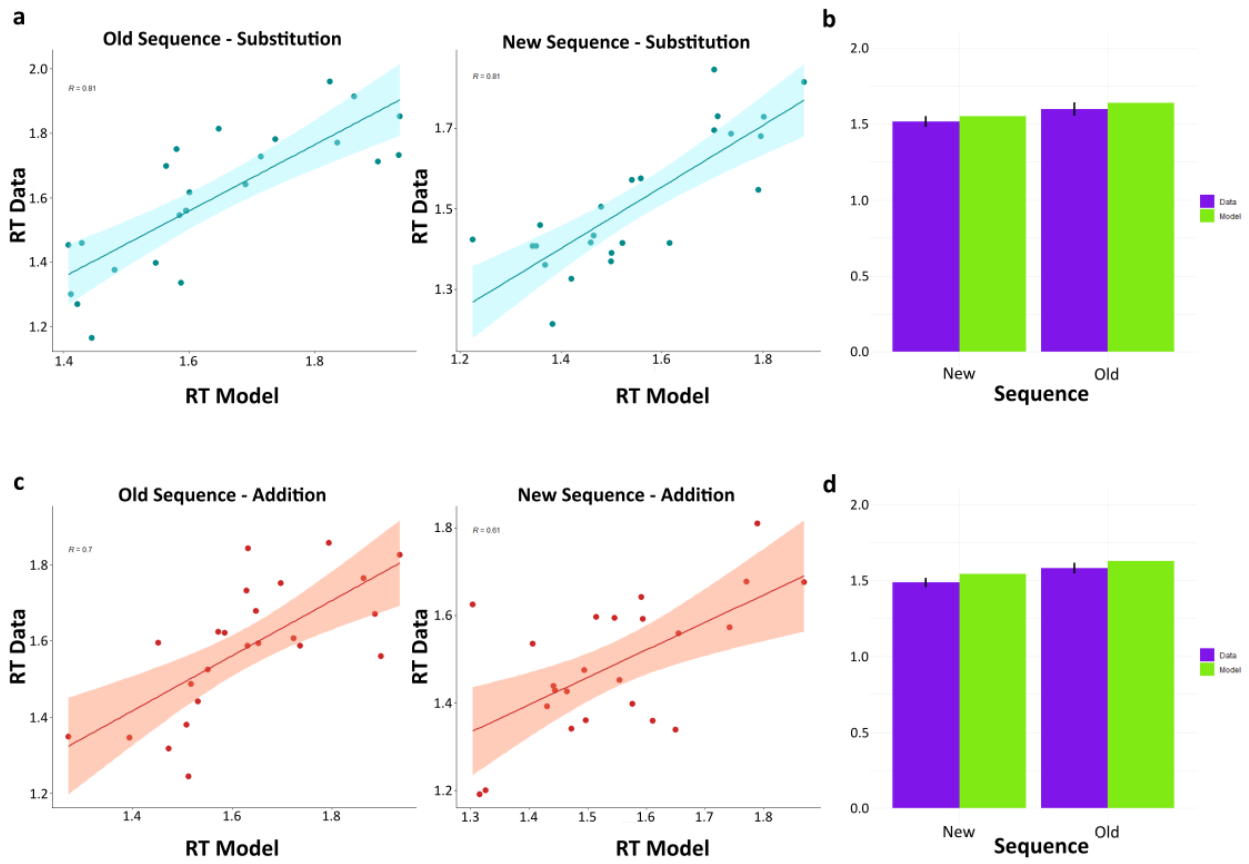

Figure 4: Model Fits (**a**) Behavioral RT data correlates with model generated RT data in Substitution for Old sequence(left), $r = 0.81$, $p < .00001$ and New sequence (right), $r = 0.81$, $p < .00001$. (**b**) Reaction time fits of model with empirical RT data in Substitution for New sequence(left) and Old sequence(right). (**c**) Behavioral RT data correlates with model generated RT data in Addition for Old sequence(left), $r = 0.61$, $p = .0014$ and New sequence (right), $r = 0.70$, $p = .00012$. (**d**) Reaction time fits of model with empirical RT data in Addition for New sequence(left) and Old sequence(right). Data (Purple), Model (Green).

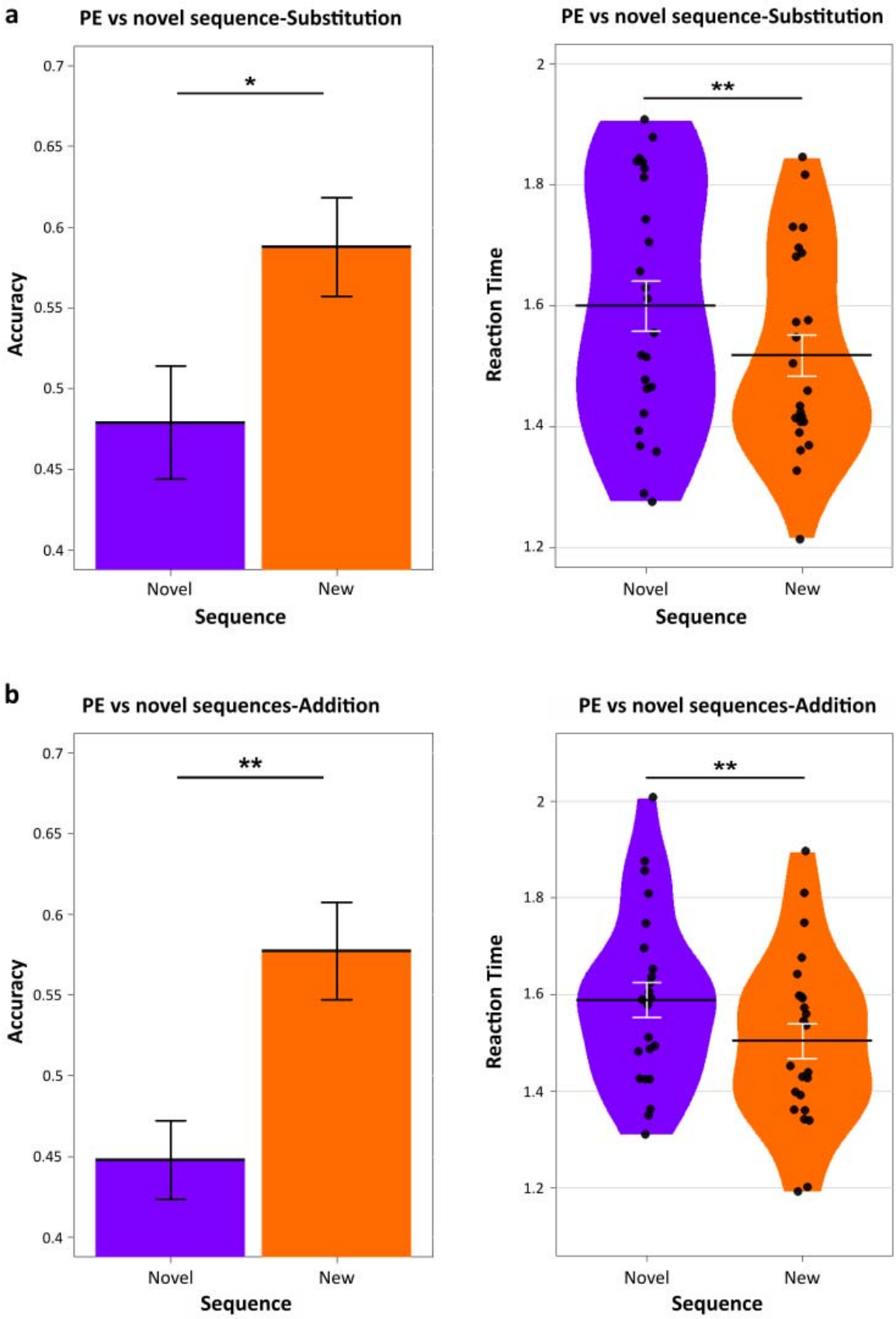
a
PE vs novel sequence-Substitution
*
Accuracy
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Novel
New
Sequence
PE vs novel sequence-Substitution
**
Reaction Time
2
1.8
1.6
1.4
1.2
Novel
New
Sequence
b
PE vs novel sequences-Addition
**
Accuracy
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Novel
New
Sequence
PE vs novel sequences-Addition
**
Reaction Time
2
1.8
1.6
1.4
1.2
Novel
New
Sequence

Figure 5: Effects are due to Prediction Errors and not novel associations (**a**) Accuracy and Reaction times in sequence memory of New (PE) sequence (Orange) and Novel sequence (Purple) in Substitution condition. Dots represent individual participant performance. Error bars represent standard error of the mean. (**b**) Accuracy and Reaction times in sequence memory of New (PE) sequence (Orange) and Novel sequence (Purple) in Addition condition. Dots represent individual participant performance. Error bars represent standard error of the mean

The effects on memory are specifically because of PEs and not due to novel associations

In order to truly verify that the memory effects are due to PEs but no other factors like novel associations, we compared the New sequence (*PrePE-New*) with the subsequent sequence, *PostPE-New*, which was termed as novel sequence. The reasoning being, since both these memories are encoded on *Day2*, unless there was a specific effect of PEs, both would justifiably have similar memory encoding. Thus, we hypothesized since the memory strengthening can only be due to PEs, the New sequences would have significantly better memories over novel associations.

In *Substitution*, there was a significant difference ($t_{(23)}$ = 2.12, p < .05, 95% CI [0.002 0.214], BF = 1.41 , d = .67) in accuracy between *New*er sequence, (*New* Sequence Mean = 58.76, SE = .03) compared to the Novel sequence (Novel Sequence Mean = 47.88, SE = .03). This difference was significant in response times as well ($t_{(23)}$ = -3.07, p = 0.0054, 95% CI [-0.136 -0.026], BF = 8.18, d = .44) between *New* Sequence Mean RT (1517ms, SE = 34ms) and the Novel Sequence Mean RT (1600ms, SE = 41ms) (**Figure 5**). This shows that in *Substitution*, even though both the sequences were seen

on the same day, there is a stark difference in memory performance specifically due to PEs and not seen in novel sequences. Next, we sought to replicate this in the *Addition* condition.

We hypothesized that the differences in accuracy would be present in *Addition* as well. Indeed a significant difference ($t_{(23)}$ = 3.29, p = 0.0031, 95% CI [0.048 0.210], BF = 12.8, d = .96) was observed in accuracy between PE sequence (*New* Sequence Mean = 57.71, SE = .03) compared to the Novel sequence here as well (Novel Sequence Mean = 44.77, SE = .02). Response times demonstrated significant differences as well ($t_{(23)}$ = -2.82, p = 0.0096, 95% CI [-0.175 -0.027], BF = 5.00, d = .61) (*New* Sequence Mean RT = 1486ms, SE = 31ms, Novel Sequence Mean RT= 1588ms, SE = 36ms). This conclusively shows the memory strengthening effects are exclusive when strong prior expectations are violated, but not when novel associations are formed.

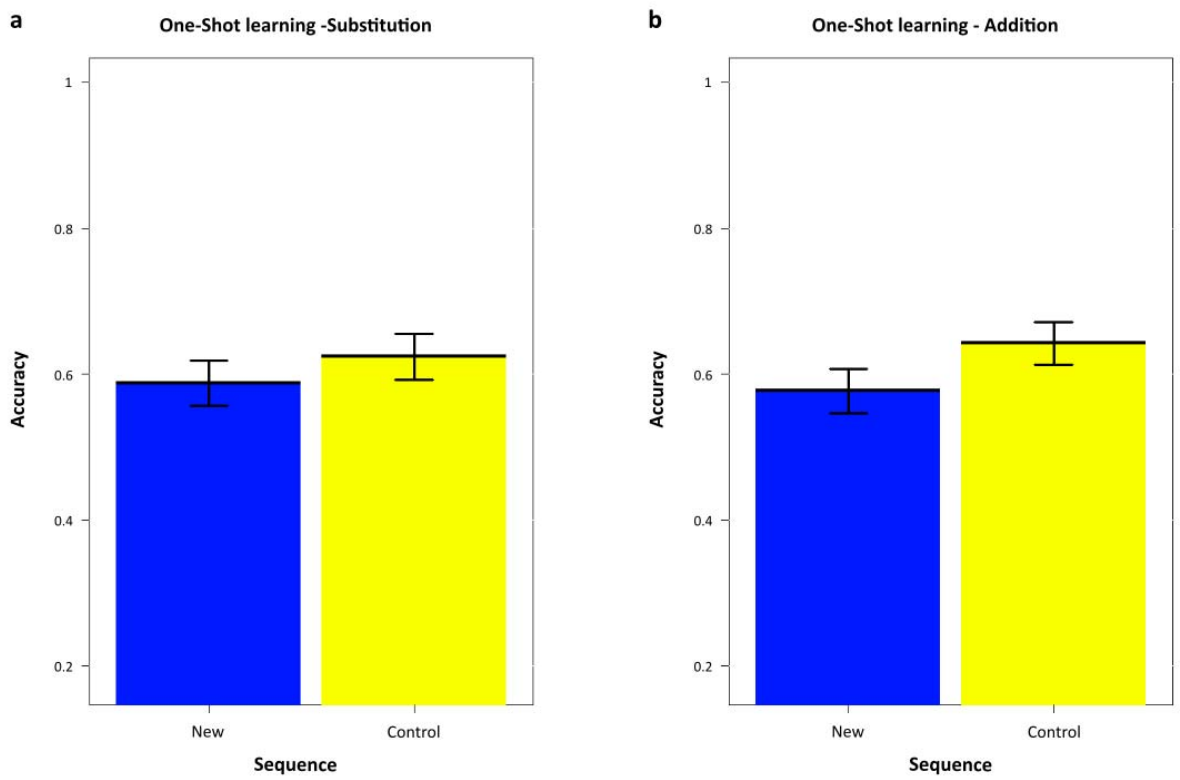

Figure 6: One-shot learning exhibited by PE sequences. Similar accuracy performance in sequence memory between New(blue) and Control(yellow) sequences in (**a**) Substitution and (**b**) Addition suggesting one-shot learning. Dots in represent individual participant performance and error bars represent standard error of the mean.

PEs promote one shot learning in the Newer sequences

Recent studies have put forth a one-shot encoding property of PEs (30). We wondered whether this holds true in extended naturalistic memories as well. We compared the Control sequences (*Start-PrePE*) so named since they are seen repeatedly on *Day1* and *Day2* without any changes or violations. This enabled us to contrast how much of

learning did PEs enhance in the memory accuracy, compared with a sequence experienced twice.

Remarkably, in *Substitution* there was no significant difference ($t_{(23)}$ = 0.83, p = .41, 95% CI [-0.054 0.126], d = 0.23, BF = 0.29) between them in terms of accuracy (*New* Sequence Mean = 58.76, SE = .03, Control Sequence Mean = 62.38, SE = .03) (**Figure 6**)

This effect of PEs on the memories was noted in *Addition* condition as well with similar memory accuracies ($t_{(23)}$ = 1.54, p = .13, 95% CI [-0.022 0.152], Cohen's d = 0.44, BF = 0.60) between them (*New* Sequence Mean = 57.71, SE = .03, Control Sequence Mean = 64.22, SE = .029). Firstly, this shows the effect is present independent of conditions and secondly it solidifies the importance of prediction errors in driving one-shot episodic sequence learning.

Discussion

Episodic memory enables us to recall sequences of past experiences in time. Can unpredicted events disrupt such encoded order of memories is the question we set out to address in this work. To this end, we designed a paradigm where participants had to recall the temporal order of a naturalistic sequence of events with contextual prediction errors. Our results reveal that encoded order of mispredicted memories were significantly weakened due to PEs. Secondly, the newer sequences with PEs are encoded with comparable strength as control sequences experienced with repetition, suggesting a one-shot encoding. Finally, by conceptualizing the sequence memory

retrieval as an evidence accumulation process over time, we offer insights into how PEs increased evidence quality for the new sequences during recall.

We observed that introducing prediction errors in a temporal context affects the sequential order of the encoded memories. Specifically, the temporal order in the Older sequences of memories that were mispredicted in the contexts, were significantly weaker. Crucially, this is paralleled by the increased memory strength noticed in the Newer memory sequences. These results support the fact that PEs can selectively disrupt episodic memories in the time domain. An expanding literature shows how PEs can affect declarative memory (Jang et al 2019, Pine et al 2018, De Loof et al 2018, Greve et al 2017), and even destabilize it (Exton-McGuinness et al 2015, Sinclair et al 2018). But an emphasis on how temporal components of episodic memories are affected by PEs is largely unexplored. Our paradigm ensured that a prior temporal context of memories is encoded and violated at specific segments, thereby allowing us to tease out the specific effect of PEs on sequential arrangement of memories.

Furthermore, in a second condition, we made the participants re-experience the mispredicted segments, after inducing the PE to reactivate those sequences strongly. Strongly reactivating memories have been shown to strengthen them (Kim et al 2017, Ritvo et al 2019). Despite these sequences having similar accuracies compared to the Newer sequences, there was a stark increase in response times. In other words, even though the segments were seen again, participants were slower in recalling the temporal order. In doing so, we uncovered an interaction between memory reactivation and PEs. Taken together, the key emerging finding from the above two results are a

specific effect of PEs in slowing the recall of mispredicted memory sequences. We sought to model empirical observations to gain insight as to why this could possibly happen.

Deploying a hierarchical drift diffusion model allowed us to disentangle the mechanisms underlying these reaction time differences. That is, whether the slower reaction times were due to slower evidence accumulation rate or a higher decision threshold. Model output showed the effects of PE are due to increased decision threshold for the Older sequences, and subsequent decrease in threshold for the Newer, PE sequences. Importantly, this relative change in decision threshold correlated significantly with the relative change in response times within participants. In contrast, no such support was found in favor of evidence accumulation rate determining reaction time differences. Although, the speed of evidence buildup did determine subjective confidence in the memory sequences. This confirms our behavioral finding where confidence measures did not change significantly due to the PEs, in line with similar studies (20). Together with the behavioral results, this can explain some outstanding findings in the literature (Hupbach et al 2009, Sinclair et al 2018, Klingmuller et al 2017).

The distinctive feature of reconsolidation studies is the intrusions of one set of memories when recalling another. For example, a list of words into a second list, when both are learned separately in the same context (Hupbach et al 2007, Hupbach et al 2009). This finding was displayed strongly in naturalistic stimuli as well (Sinclair et al 2018). Such intrusions are by nature, asymmetric, in the sense that only the second list can intrude onto the first and never the other way around. We interpret this longstanding

finding in the light of our result as follows. During remembering the temporal order of memories sharing the same context, a source confusion ensues (Howard et al 2002, Sederberg et al 2011). The PE sequences owing to their lower decision threshold are recalled faster. While freely recalling memories, due to the decreased evidence required PE memories intrudes into an older memory sharing the same context, reflecting as errors in remembering. The opposite is impossible since older memories require more evidence. Most tasks that (Jang et al 2019, De Loof et al 2018) induce prediction errors have a reward task structure learned over many trials. However, in our paradigm each segment created a PE only once, which was sufficient to destabilize the prior encoded sequence of events. It might be as if accessibility of episodic memories depends on contextual PEs while recalling and not only about reward or value. The increased evidence quality in newer sequences also adds credence to predictive coding frameworks of episodic memory, where newer unexpected information is inherently prioritized to update past knowledge.

Alternatively, the results of our experiment can be explained with a temporal advantage effect. Particularly, since participants saw the Newer segments during *Day2*, they will have a recency effect during *Day3* compared to segments encoded during *Day1*. However, we tried to minimize this recency effect by specifically instructing the participants to encode both days' scenes with equal priority. Furthermore, our hypothesized effect of PEs on temporal order was reported in the *Addition* condition as well, which had the memory segments of *Day1* re-experienced, possibly ruling out this confound. Comparison of Newer sequence memories, having PEs to novel association sequences showed significant differences in memory strength and accessibility. Thus,

we validated that the reported effects are due to violation of predictions only, and not due to novel associations. If the effects were due to the effects of just association between events, then there should not have been any differences between the two sequences. But our results prove convincingly otherwise. Generally, in encoding temporal order of sequences, the association strength between two episodes depends on the informational relation on which they differ (Rescorla 1990). Thus, the more surprise involved, the more information gets equated with the association (Shannon 1948). While we did not quantify the surprise elicited per segment mathematically, one can hypothesize Newer PE sequences carry more surprises than the novel sequences following it.

One compelling question to keep in mind when understanding the effect of PEs on temporal arrangement of memories, is how it measures up with events of repeated experience. In our experiment, the control sequences were repeated unchanged across both days' viewings. Intuitively speaking, these should have better accuracies than PE sequences that were seen on just once, owing to multiple exposures. Yet strikingly, the memory accuracies were similar, implying a one-shot learning for the *New*er sequences. This is because it took the *New*er PE sequences only one exposure to gain a performance similar to the repetitive sequences. A recent study (Greve et al 2017) has shown that one-shot learning is seen with PEs in paired association learning. We extend this finding to the time domain as memory sequences with PEs showed similar accuracies as control sequences.

A promising future endeavor is investigating whether such sequential reorganization of prior encoded memories can be observed in language processing as well as in

navigation, since episodic memory and navigation are underpinned by similar neural structures**.** Recalling, ordering and integrating memories in time is widely considered to be a function of MTL structures (Lisman et al 2009, Davachi et al 2015, Tubridy et al 2011, Fortin et al 2002, Montchal et al 2019, Tsao et al 2018). The hippocampus has also been implicated in detecting mismatched sequences (Kumaran et al 2006). Not surprisingly, it has been hypothesized to be a generator of sequences recently (Buzsaki et al 2018) and draws attention to this being its singular computation. A central theme of a recent theoretical work (Stachenfeld et al 2017) also proposed the hippocampus as a predictive map as opposed to its legacy in the literature as a spatial map (Breathnach et al 2012). Future studies can uncover how effects of contextual prediction errors in episodic memories is mediated by the hippocampus.

In closing, we summarize the effects of contextual prediction errors on temporal sequence of memories. While most laboratory experiments are highly controlled it usually doesn't have the ecological validity of real life, which we tried to incorporate in our design. While mimicking real life scenarios have certain distinct advantage such paradigms often involves many parameters that are uncontrollable. To ensure that we have carefully divided the two movies into specific events and segments which allowed for introducing contextual prediction errors during episodic sequence encoding. Our main contribution is that we provide conclusive evidence to how PEs mediate the temporal order of episodic memories which is observed as contextual changes (DuBrow et al 2013). This can be extended to clinical settings where such PE based naturalistic

viewings could not only reveal contextual memory encoding/updating deficits early on, but also provide a cognitive tool to weaken undesirable memories in PTSD and anxiety.

Acknowledgements

This study was supported by NBRC Core funds, Ramalingaswami Fellowships (Department of Biotechnology, Government of India) to DR (BT/RLF/Re-entry/07/2014). AB also acknowledges the support of the Centre of Excellence in Epilepsy and MEG. DR was also supported by SR/CSRI/21/2016 extramural grant from the Department of Science and Technology (DST) Ministry of Science and Technology, Government of India. AB and DR are also supported by BT/MED-III/NBRC/Flagship/Program/2019. We would also like to thank Gargi Majumdar and Dipanjan Ray for their helpful suggestions and comments on the manuscript.

References

1. Howard, M. W. & Kahana, M. J. A distributed representation of temporal context. J. Math. Psychol. 46, 269–299 (2002).
2. Yonelinas, A. P., Ranganath, C., Ekstrom, A. D. & Wiltgen, B. J. A contextual binding theory of episodic memory: systems consolidation reconsidered. Nat. Rev. Neurosci. 20, 364–375 (2019).
3. Polyn, S. M., Norman, K. A. & Kahana, M. J. A Context Maintenance and Retrieval Model of Organizational Processes in Free Recall. Psychol. Rev. 116, 129–156 (2009).
4. Hupbach, A., Gomez, R. & Nadel, L. Episodic memory updating: The role of context familiarity. Psychon. Bull. Rev. 18, 787–797 (2011).

5. Davachi, L. Item, context and relational episodic encoding in humans. Current Opinion in Neurobiology 16, 693–700 (2006).
6. Smith, T. A., Hasinski, A. E. & Sederberg, P. B. The context repetition effect: Predicted events are remembered better, even when they don't happen. J. Exp. Psychol. Gen. 142, 1298–1308 (2013).
7. Henson RN, Gagnepain P. Predictive, interactive multiple memory systems. Hippocampus. 2010
8. Mullally, Sinéad L, and Eleanor A Maguire. Memory, Imagination, and Predicting the Future: A Common Brain Mechanism? The Neuroscientist (2014)
9. Bar, M. The proactive brain: Memory for predictions. Philos. Trans. R. Soc. B Biol. Sci. 364, 1235–1243 (2009).
10. Lisman, J. & Redish, A. D. Prediction, sequences and the hippocampus. Philos. Trans. R. Soc. B Biol. Sci. 364, 1193–1201 (2009).
11. Tulving, E. Elements of episodic memory. (Clarendon Press, 1983).
12. Sederberg, P. B., Gershman, S. J., Polyn, S. M. & Norman, K. A. Human memory reconsolidation can be explained using the temporal context model. Psychon. Bull. Rev. 18, 455–468 (2011)
13. Kim, G., Lewis-Peacock, J. A., Norman, K. A. & Turk-Browne, N. B. Pruning of memories by context-based prediction error. Proc. Natl. Acad. Sci. 111, 8997–9002 (2014).
14. Alberini, C. M. & Ledoux, J. E. Memory reconsolidation. Current Biology 23, R746–R750 (2013).
15. Hupbach, A., Gomez, R., Hardt, O. & Nadel, L. Reconsolidation of episodic memories: A subtle reminder triggers integration of *New* information. Learn. Mem. 14, 47–53 (2007).
16. Hupbach, A., Gomez, R. & Nadel, L. Episodic memory reconsolidation: Updating or source confusion? Memory 17, 502–510 (2009).
17. Lee, J. L. C., Nader, K. & Schiller, D. An Update on Memory Reconsolidation Updating. Trends in Cognitive Sciences 21, 531–545 (2017)
18. Exton-McGuinness, M. T. J., Lee, J. L. C. & Reichelt, A. C. Updating memories-The role of prediction errors in memory reconsolidation. Behavioural Brain Research 278, 375–384 (2015).
19. Fernández, R. S., Boccia, M. M. & Pedreira, M. E. The fate of memory: Reconsolidation and the case of Prediction Error. Neuroscience and Biobehavioral Reviews 68, 423–441 (2016).
20. Sinclair, A. H. & Barense, M. D. Surprise and destabilize: Prediction error influences episodic memory reconsolidation. Learn. Mem. 25, 369–381 (2018).
21. Jang, A. I., Nassar, M. R., Dillon, D. G. & Frank, M. J. Positive reward prediction errors during decision-making strengthen memory encoding. Nat. Hum. Behav. 3, 719–732 (2019).

22. Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y. & Mendelsohn, A. Knowledge acquisition is governed by striatal prediction errors. Nat. Commun. 9, 1–14 (2018).
23. De Loof, E. et al. Signed reward prediction errors drive declarative learning. PLoS One 13, (2018).
24. Greve, A., Cooper, E., Kaula, A., Anderson, M. C. & Henson, R. Does prediction error drive one-shot declarative learning? J. Mem. Lang. 94, 149–165 (2017).
25. Friston, K. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 815–836 (2005).
26. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36, 181–204 (2013).
27. Kurby, C. A. & Zacks, J. M. Segmentation in the perception and memory of events. Trends Cogn. Sci. 12, 72–79 (2008).
28. Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. PsychoPy2: experiments in behavior made easy. Behavior Research Methods , 195-203 (2019).
29. Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. Front. Neuroinform. 7, 14 (2013).
30. Ratcliff, R. & McKoon, G. The diffusion decision model: Theory and data for two-choice decision tasks. Neural Computation 20, 873–922 (2008).
31. Ratcliff, R. & Rouder, J. N. Modeling Response Times for Two-Choice Decisions. Psychol. Sci. 9, 347–356 (1998).
32. Gamerman, D. & Lopes, H. F. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd edition Taylor and Francis, (2006).
33. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B Stat. Methodol. 64, 583–616 (2002).
34. Gelman, A. & Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. Stat. Sci. 7, 457–472 (1992).
35. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. & Iverson, G. Bayesian t tests for accepting and rejecting the null hypothesis. Psychonomic Bulletin and Review 16, 225–237 (2009).
36. Kim, G., Norman, K. A. & Turk-Browne, N. B. Neural differentiation of incorrectly predicted memories. J. Neurosci. 37, 2022–2031 (2017).
37. Ritvo, V. J. H., Turk-Browne, N. B. & Norman, K. A. Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends Cogn. Sci.* **23**, 726–742 (2019).
38. Klingmüller, A., Caplan, J. B. & Sommer, T. Intrusions in episodic memory: Reconsolidation or interference? Learn. Mem. 24, 216–224 (2017).
39. Rescorla, R. A. The Role of Information About the Response-Outcome Relation in Instrumental Discrimination Learning. J. Exp. Psychol. Anim. Behav. Process. 16, 262–270 (1990).

40. Shannon, C. E. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423 (1948).
41. Davachi, L. & DuBrow, S. How the hippocampus preserves order: The role of prediction and context. Trends in Cognitive Sciences 19, 92–99 (2015).
42. Tubridy, S. & Davachi, L. Medial temporal lobe contributions to episodic sequence encoding. Cereb. Cortex 21, 272–280 (2011).
43. Fortin, N. J., Agster, K. L. & Eichenbaum, H. B. Critical role of the hippocampus in memory for sequences of events. Nat. Neurosci. 5, 458–462 (2002).
44. Montchal, M. E., Reagh, Z. M. & Yassa, M. A. Precise temporal memories are supported by the lateral entorhinal cortex in humans. *Nat. Neurosci.* **22**, 284–288 (2019).
45. Tsao, A. *et al.* Integrating time from experience in the lateral entorhinal cortex. *Nature* **561**, 57–62 (2018).
46. Kumaran, D. & Maguire, E. A. An unexpected sequence of events: Mismatch detection in the human hippocampus. PLoS Biol. 4, 2372–2382 (2006).
47. Buzsáki, G. & Tingley, D. Space and Time: The Hippocampus as a Sequence Generator. Trends Cogn. Sci. 22, 853–869 (2018).
48. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. Nat. Neurosci. 20, 1643–1653 (2017).
49. Breathnach, C. S. The Hippocampus as a Cognitive Map. Philosophical Studies 27, (2012).
50. DuBrow, S. & Davachi, L. The influence of context boundaries on memory for the sequential order of events. J. Exp. Psychol. Gen. 142, 1277–1286 (2013).