Prediction Errors Disrupt Hippocampal Representations and Update Episodic Memories

*Alyssa H. Sinclair[1,2] Grace M. Manalili[2], Iva K. Brunec[3,4],

R. Alison Adcock[1], & Morgan D. Barense[2,5]

[1] Duke University, *Department of Psychology & Neuroscience*, Durham, NC, 27705, USA

[2] University of Toronto, *Department of Psychology*, Toronto, ON, M5S 3G3, Canada

[3] Temple University, *Department of Psychology*, Philadelphia, PA, 19122, USA

[4] University of Pennsylvania, *Department of Psychology*, Philadelphia, PA, 19104, USA

[5] Baycrest Hospital, *Rotman Research Institute*, Toronto, ON, M64 1W1, Canada

* Corresponding Author: allie.sinclair@duke.edu

Keywords: prediction error, memory, learning, hippocampus, basal forebrain, acetylcholine, false memory, cognitive neuroscience, fMRI

Significance

1   Our brains draw on memories to predict the future; when our predictions are incorrect,

2   we must update our memories to improve future predictions. Past studies have demonstrated that

3   the hippocampus signals *prediction error*, or surprise, but have not linked this neural signal to

4   memory updating. Here, we uncover this missing connection: We show that mnemonic

5   prediction errors change the role of the hippocampus, reversing the relationship between

6   hippocampal activation and memory outcomes. We examine the mechanisms of this shift in

7   neural processing, showing that prediction errors disrupt the temporal continuity of hippocampal

8   patterns. We propose that prediction errors disrupt sustained representations and enable memory

9   updating. Our findings bear implications for improving education, understanding eyewitness

10   memory distortion, and treating pathological memories.

1   **Abstract**

2   The brain supports adaptive behavior by generating predictions, learning from errors, and

3   updating memories to incorporate new information. *Prediction error,* or surprise, triggers

4   learning when reality contradicts expectations. Prior studies have shown that the hippocampus

5   signals prediction errors, but the hypothesized link to memory updating has not been

6   demonstrated. In a human fMRI study, we elicited mnemonic prediction errors by interrupting

7   familiar narrative videos immediately before the expected endings. We found that prediction

8   error reversed the relationship between univariate hippocampal activation and memory: greater

9   hippocampal activation predicted memory preservation after expected endings, but memory

10   updating after surprising endings. In contrast to previous studies, we showed that univariate

11   activation was insufficient for understanding hippocampal prediction error signals. We explained

12   this surprising finding by tracking both the evolution of hippocampal activation patterns and

13   connectivity between the hippocampus and neuromodulatory regions. We found that

14   hippocampal activation patterns stabilized as each narrative episode unfolded, suggesting

15   sustained episodic representations. Prediction errors disrupted these sustained representations,

16   and the degree of disruption predicted memory updating. The relationship between hippocampal

17   activation and subsequent memory depended on concurrent basal forebrain activation, supporting

18   the idea that cholinergic modulation regulates attention and memory. We conclude that

19   prediction errors create conditions that favor memory updating, prompting the hippocampus to

20   abandon ongoing predictions and make memories malleable.

Introduction

In daily life, we continuously draw on past experiences to predict the future. Expectation and surprise shape learning across many situations, such as when we discover misinformation in the news, receive feedback on an exam, or make decisions based on past outcomes. When our predictions are incorrect, we must update our mnemonic models of the world to support adaptive behavior. *Prediction error* is a measure of the discrepancy between expectation and reality; this surprise signal is both evident in brain activity and related to learning (1–6). The brain dynamically reconstructs memories during recall, recreating and revising past experiences based on current information (7). The intuitive idea that surprise governs learning has long shaped our understanding of memory, reward learning, perception, action, and social behavior (2, 8–14). Yet, the neural mechanisms that allow prediction error to update memories remain unknown.

Past research has implicated the hippocampus in each of the mnemonic functions required for learning from prediction errors: retrieving memories to make predictions, identifying discrepancies between past and present, and encoding new information (2, 15–20). Functional MRI (fMRI) studies have shown that hippocampal activation increases after predictions are violated; this surprise response has been termed *mismatch detection* (18, 19, 21–23), or *mnemonic prediction error* (20). These past studies have shown that the hippocampus *detects* mnemonic prediction errors. Several theoretical frameworks have hypothesized that this hippocampal prediction error signal could update memories (17, 20, 24–27), but this crucial link for understanding how we learn from error has not yet been demonstrated.

What mechanisms could link hippocampal prediction errors to memory updating? A leading hypothesis is that prediction errors shift the focus of attention and adjust cognitive processing (20, 28–32). After episodes that align with expectations, we should continue

1   generating predictions and shift attention *internally*, sustaining and reinforcing existing

2   memories. However, after mnemonic prediction errors, we should reset our expectations and

3   shift attention *externally*, preparing to encode new information and update memories. Consistent

4   with this idea, mnemonic prediction errors have been shown to enhance the hippocampal *input*

5   pathway that supports encoding, but suppress the *output* pathway that supports retrieval (20). We

6   propose that surprising events may also change intrinsic hippocampal processing, changing the

7   effect of hippocampal activation on memory outcomes.

8   Neuromodulation may be a critical factor that regulates hippocampal processing and

9   enables memory updating. Currently, there is mixed evidence supporting two hypotheses:

10   acetylcholine and/or dopamine could act upon the hippocampus to regulate processing after

11   surprising events (24–27, 29, 31, 33, 34). Several models have proposed that acetylcholine from

12   the medial septum (within the basal forebrain) regulates the balance between input and output

13   pathways in the hippocampus (27–29, 35–38), thus allowing stored memories to be compared

14   with perceptual input (31, 38, 39). After prediction errors, acetylcholine release could change

15   hippocampal processing and enhance encoding or memory updating (26, 29, 33, 37, 39). On the

16   other hand, dopamine released from the ventral tegmental area (VTA), if transmitted to the

17   hippocampus, could also modulate hippocampal plasticity after prediction errors. Past studies

18   have shown that the hippocampus and VTA are co-activated after surprising events (40, 41).

19   Other work has shown that co-activation of the hippocampus and VTA predicts memory

20   encoding and integration (42–45). Overall, basal forebrain and VTA neuromodulation are both

21   candidate mechanisms for regulating hippocampal processing and memory updating.

22   In the present study, we used an fMRI task with human participants to examine trial-wise

23   hippocampal responses to prediction errors during narrative videos. During the **Encoding phase**,

1   participants viewed 70 full-length videos that featured narrative episodes with salient endings

2   (e.g., a baseball batter hitting a home run) (Figure 1A). During the **Reactivation phase** the

3   following day, participants watched the videos again (Figure 1B). We elicited mnemonic

4   prediction errors by interrupting half of the videos immediately before the expected narrative

5   ending (e.g., the video ends while the baseball batter is mid-swing). These surprising

6   interruptions were comparable to the prediction errors employed in prior studies of memory

7   updating (1). Half of the videos were presented in Full-length form (as previously seen during

8   the Encoding phase), and half were presented in Interrupted form (eliciting prediction error).
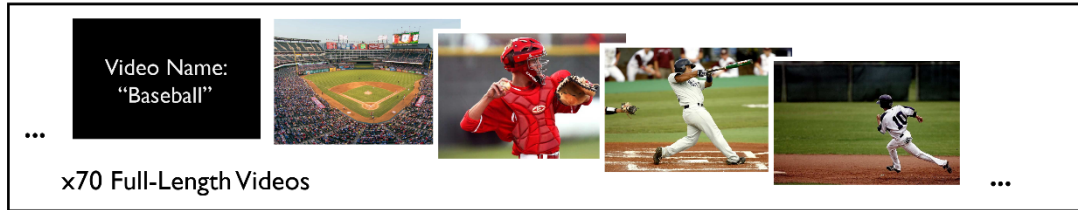
9   During the **Test phase**, participants completed a memory test in the form of a structured

10   interview (Figure 1C). On each trial, participants were cued with the name of the video and

11   recalled the narrative. The experimenter then probed for further details with pre-determined

12   questions (e.g., "Can you describe the baseball batter's ethnicity, age range, or clothing?"). Our

13   critical measure of memory updating was *false memories*, because the presence of a false

14   memory indicates that the original memory was changed in some way. Although it can be

15   adaptive to update real-world memories by incorporating relevant new information, we expected

16   that our laboratory paradigm would induce false memories because participants would integrate

17   interfering details across similar episodes (1, 7). Because we were interested in false memories as

18   a measure of memory updating, we instructed participants not to guess and permitted them to

19   skip details they could not recall.

20   Prior research in human and animals has shown that some memory updating effects only

21   emerge after delays that allow protein synthesis to occur during consolidation and

22   reconsolidation (1, 46–48). Therefore, to test our primary question about the neural correlates of

23   memory updating, fMRI participants completed the Encoding, Reactivation, and Test phases

1   over three days, with 24-hours between each session (*Delayed* group, n = 24). In addition, we

2   tested the behavioral prediction that memory updating would *require* a delay (i.e., because

3   transforming a memory trace requires protein synthesis) by recruiting a separate group of

4   participants who completed the Test phase immediately after the Reactivation phase on Day 2

5   (*Immediate* group, n = 24) (Figure 1D). Delayed group participants completed the Reactivation

6   phase while undergoing an fMRI scan, whereas Immediate group participants (n = 24) were not

7   scanned. Our primary fMRI analyses examined the fixation period immediately following the

8   offset of Full and Interrupted videos (Post-Video period; Figure 1B, right) during the

9   Reactivation phase in the Delayed group. Importantly, this design compares neural responses to

10   surprising and expected video endings while controlling for visual and auditory input.

11   Our approach allowed us to test several questions set up by the prior literature. First, we

12   used naturalistic video stimuli to examine the effect of mnemonic prediction error on

13   hippocampal activation and episodic memories. Second, to investigate hippocampal processing,

14   we used multivariate analyses to track how episodic representations were sustained or disrupted

15   over time. Third, to test hypotheses about neuromodulatory mechanisms, we related hippocampal

16   activation and memory updating to activation in the basal forebrain and VTA.

A   Encoding Phase: Example Stimulus Video

Video Name: "Baseball"

... x70 Full-Length Videos ...

B   Reactivation Phase: Example Trial

Video Name: "Baseball"

Full or **Interrupted**

Cue (3s)   Video Playback (M=30s, SD=7s)   Post-Video Period (7s) (Focus of fMRI analyses)
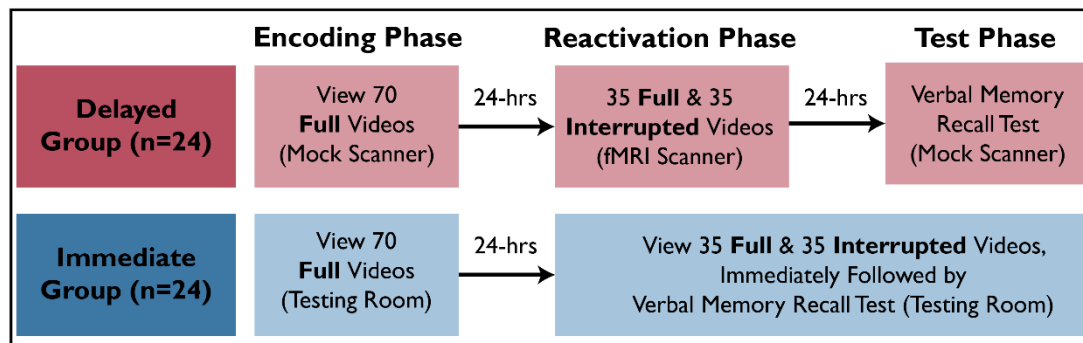
C   Test Phase: Example Memory Test

Experimenter: The next video is "Baseball." Can you describe the main event of the video?
Participant: Okay, so they're **in a stadium**, and there are **lots of people watching.** The **pitcher throws the ball** and the batter **hits it hard** and **gets a home run.**
Experimenter: Can you describe the baseball batter? Age range, hair color, ethnicity, or clothing?
Participant: He was **Caucasian**, in his **mid-40s.** He was **wearing a green uniform**.
Experimenter: Do you remember hair color?
Participant: No, I don't remember.

Legend: **Correct Details**   **False Memories**

D   Overview of Paradigm

	Encoding Phase		Reactivation Phase		Test Phase
Delayed Group (n=24)	View 70 **Full** Videos (Mock Scanner)	24-hrs →	35 **Full** & 35 **Interrupted** Videos (fMRI Scanner)	24-hrs →	Verbal Memory Recall Test (Mock Scanner)
Immediate Group (n=24)	View 70 **Full** Videos (Testing Room)	24-hrs →	View 35 **Full** & 35 **Interrupted** Videos, Immediately Followed by Verbal Memory Recall Test (Testing Room)		

2   *Figure 1*. Overview of experimental paradigm. A) During the Encoding phase, all videos were
3   presented in Full-length form. Here we show example frames depicting a stimulus video. B)
4   During the Reactivation phase, participants viewed the 70 videos again, but half (35) were
5   interrupted to elicit mnemonic prediction error. Participants were cued with the video name,
6   watched the video (Full or Interrupted), and then viewed a fixation screen. The "Baseball" video
7   was interrupted when the batter was mid-swing. fMRI analyses focused on the Post-Video
8   fixation periods after each video (red highlighted box). Thus, visual and auditory stimulation
9   were matched across Full and Interrupted conditions, allowing us to compare Post-Video neural

1 activation while controlling for perceptual input. C) During the Test phase, participants answered
2 structured interview questions about all 70 videos, and were instructed to answer based on their
3 memory of the Full video originally shown during the Encoding phase. Here we show example
4 text illustrating the memory test format and scoring of correct details (our measure of memory
5 preservation) and false memories (our measure of memory updating, because false memories
6 indicate that the memory has been modified). The void response ("I don't remember") is not
7 counted as a false memory. D) Overview of the experiment. All participants completed
8 Encoding, Reactivation, and Test phases of the study. The Delayed group (fMRI participants)
9 completed the Test phase 24 hours after Reactivation, because prior studies have shown that
10 memory updating becomes evident only after a delay (e.g., to permit protein synthesis). The
11 Immediate group completed the Test phase immediately after Reactivation and was not scanned.
12 The purpose of the Immediate group was to test the behavioral prediction that memory updating
13 required a delay.

14

15   **Results**

16 **Behavioral Results**

17   We transcribed and scored memory tests for two key measures: number of unique *correct*

18 *details* (Figure 2A) and *false memories* (Figure 2B) – reflecting memory preservation and

19 updating, respectively. We also collected *confidence ratings* and scored the number of *forgotten*

20 *videos* (Supplementary Information, *Confidence and Forgetting*, Figure S1). We defined false

21 memories as distorted details that the participant recalled from the episode (e.g., "The pitcher

22 wore a green uniform"). Void responses (e.g., "I don't remember") were not counted as false

23 memories, but were missed opportunities to earn points for correct details. Importantly, our

24 measures for correct details and false memories were independent; there was no limit to the

25 number of details a participant could recall about a video, and each detail was scored as correct

26 or false. We conducted linear mixed-effects regression to predict memory outcomes (either

27 *correct details* or *false memories*) from the factors *group* (Delayed vs. Immediate) and

28 *reactivation type* (Full vs. Interrupted). In all models, we included random effects to account for

1   by-subject and by-video variability (Supplementary Information, *Supplementary Methods,*

2   *Linear Mixed Effects Regression*).

3   ***Prediction Error Increased Correct Details***

4   We found that prediction errors during memory reactivation enhanced recall of correct

5   details (main effect of reactivation type: β = -0.07, 95% CI [-0.12, -0.02], t = -2.75, p = .008)

6   (Figure 2A), such that participants in both groups reported more correct details for Interrupted

7   videos than Full videos (Delayed Group: β = -0.05, t = -2.07, p = .042; Immediate Group: β = -

8   0.08, t = -2.50, p = .015) (Table S1A). Even though the video endings were omitted, prediction

9   errors strengthened and preserved existing memories. Participants in the Delayed group recalled

10   fewer correct details than participants in the Immediate group (main effect of group: β = 0.16,

11   95% CI [0.01, 0.31], t = 2.16, p < .036), likely because the Delayed group completed the

12   memory test after 24 hours. There was no interaction between group and reactivation type (β = -

13   0.01, 95% CI [-0.04, 0.02], t = -0.68, p = .495), indicating that the effect of prediction error

14   enhancing correct details did not require a delay.

15   ***Prediction Error Increased False Memories***

16   We found that prediction errors selectively increased false memories in the Delayed

17   group, replicating our past behavioral results (48) (significant interaction between reactivation

18   type and group, β = 0.04, 95% CI [0.01, 0.07], t = 2.61, p = .010) (Figure 2B, Table S1B). In

19   other words, Interrupted videos increased false memories in the Delayed group (β = -0.09, t = -

20   3.50, p < .001), but not the Immediate group (β = -0.01, t = -0.78, p = .437). We also found a

21   main effect of group indicating more false memories in the Delayed group (β = -0.36, 95% CI [-

22   0.43, -0.29], t = -9.68, p < .001). Lastly, there was a main effect of reactivation type (β = -0.05,

23   95% CI [-0.08, -0.02], t = -3.31, p = .001), driven by the effect of prediction error increasing

1   false memories in the Delayed group. To ensure that the effect of prediction error on memory

2   updating was not driven by the first few trials (which are presumably most surprising), we also

3   conducted a control analysis that found no effect of trial number on false memories

4   (Supplementary Information, *Trial Number Control*, Table S2).

5   In sum, our behavioral results showed a novel dissociation between reinforcing and

6   updating memories: Prediction errors during memory reactivation strengthened memories,

7   evident both immediately and after a delay (Figure 2A). However, memory updating (as revealed

8   by false memories) was not evident until after a delay (Figure 2B), consistent with prior studies

9   that have been interpreted in terms of reconsolidation theory (1, 46–48).

10   ***Surprise Ratings and Semantic Similarity Predicted False Memories***

11   Expanding on the results reported above, we recruited an independent sample to watch

12   the videos and rate (on a 5-point Likert scale) the degree of surprise elicited by the narrative

13   interruptions (Supplementary Information, *Supplementary Methods, Online Ratings of Stimulus*

14   *Videos*). We then calculated the average surprise rating for each video and then related these

15   surprise ratings to memory outcomes in our laboratory sample. Using linear mixed effects

16   regression, we predicted subsequent *false memories* from the variables *reactivation type* (Full vs.

17   Interrupted), *group* (Delayed vs. Immediate), *surprise ratings* (continuous), and all relevant

18   interactions. All model parameters are reported in Table S3A. There was a significant interaction

19   between surprise ratings and group (β = -0.03, 95% CI [-0.06, -0.01], t = -2.07, p = .039), such

20   that more surprising videos were associated with more false memories selectively in the Delayed

21   group (Delayed: β = 0.10, z = 2.49, p = .013; Immediate: β = 0.03, z = 0.95, p = .344). However,

22   the three-way interaction among surprise ratings, group, and reactivation type was not significant

23   (β = 0.01, 95% CI [-0.02, 0.04], t = 0.75, p = .451). Because our surprise ratings were collected

1   from a separate online sample, this measure may not be sensitive enough to detect the expected

2   three-way interaction when applied to the laboratory sample. In a separate model, we found that

3   surprise ratings were not associated with correct details (Table S3B).

In the current study, we indexed memory updating in terms of false memories; however,

incorporating relevant information into memory can be an adaptive function. We hypothesized

that our paradigm would induce *false* memories because information would be integrated across

semantically-related episodes. To test this hypothesis, we quantified semantic similarity among

the 70 videos with a text-based analysis (Supplementary Information, *Supplementary Methods*,

Scoring of Memory Tests). Using linear mixed effects regression, we predicted subsequent *false*

memories from the variables *reactivation type* (Full vs. Interrupted), *group* (Delayed vs.

Immediate), *semantic similarity* (continuous), and all relevant interactions. All model parameters

are reported in Table S4. We found that videos that were more semantically-similar to other

videos in the stimulus set produced more false memories ($\beta = 0.11$, 95% CI [0.04, 0.19], $t = 3.09$,

$p = .003$) (Table S4). An interaction between semantic similarity and group predicted false

memories ($\beta = -0.04$, 95% CI [-0.07, -0.01], $t = -2.77$, $p = .006$), such that the effect of semantic

similarity was stronger in the Delayed group ($\beta = 0.16$, $z = 3.90$, $p < .001$) than in the Immediate

group ($\beta = 0.07$, $z = 1.80$, $p = .073$). There was also an interaction between semantic similarity

and reactivation type that predicted false memories ($\beta = -0.03$, 95% CI [-0.06, -0.01], $t = -2.08$, p

$= .038$), such that the effect of semantic similarity was stronger for Interrupted videos ($\beta = 0.15$,

$z = 3.44$, $p < .001$) than for Full videos ($\beta = 0.08$, $z = 2.18$, $p = .030$). However, the three-way

interaction among reactivation type, group, and semantic similarity was not significant ($\beta = 0.02$,

95% CI [-0.02, 0.05], $t = 0.99$, $p = .323$). Overall, these results suggest that interfering details

from semantically-related videos distorted memories; this interference effect was strongest for

Interrupted videos and for participants in the Delayed group. Consistent with an adaptive

updating process, memories may have been updated with relevant information from other videos.
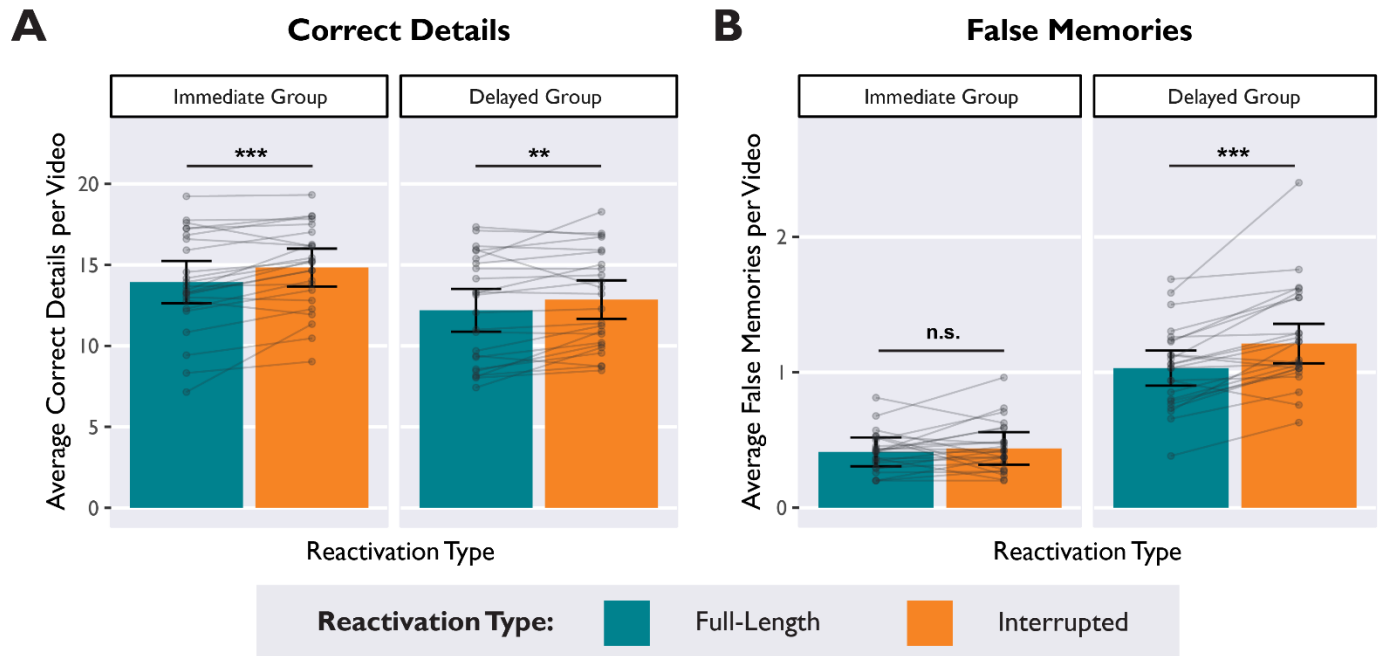

Figure 2. Prediction errors strengthened and updated memories over distinct time-courses. A) In both the Delayed and Immediate groups, average Correct Details were higher for videos that were Interrupted during memory reactivation, demonstrating that prediction error can strengthen memory recall both immediately and after a delay. B) In the Delayed group (but not the Immediate group), average False Memories were higher for videos that were Interrupted during memory reactivation. This interaction demonstrates that prediction error enabled memory updating, but only after a delay. Bars depict estimated marginal means from a linear mixed effects model. Subject averages are overlaid on top to display the distribution: Dots indicate average scores by-participant, and lines connect within-subjects measures. Error bars depict 95% confidence intervals. * $p < .05$, ** $p < .01$, *** $p < .001$.

Univariate fMRI Results

The primary aim of our univariate fMRI analyses was to test the following questions: (1) Is hippocampal activation related to *reactivation type* (Full vs. Interrupted) and memory updating as indexed by subsequent *false memories?* (2) If so, does activation in the *basal forebrain* or the *VTA* moderate the relationship between hippocampal activation and memory updating?

1   We analyzed the blood oxygen level-dependent (BOLD) signal from the 24 subjects in

2   the Delayed group (the Immediate group was not scanned). Our analyses focused on the fixation

3   screen presented during the Post-Video period immediately after each video offset. This fixation

4   screen was preceded by the narrative ending of each video, which was either as-expected (Full)

5   or a surprising prediction error (Interrupted). Importantly, visual and auditory input was identical

6   across conditions during this Post-Video fixation screen, and thus, by analyzing neural activation

7   during the Post-Video period we controlled for differences in visual and auditory input (Figure

8   2B). Another possibility is that condition differences in the duration of visual stimulation (during

9   video playback) could affect the magnitude and duration of the ongoing BOLD response during

10   the Post-Video period. We conducted control analyses to confirm that our results were not

11   confounded by video duration (Supplementary Information, *Video Duration Control*, Table S18).

12   ***The Effect of Hippocampal Activation on Memory Depended on Prediction Error***

13   Whole-brain mass univariate results are provided in the Supplementary Information

14   (*Whole-Brain Analysis*, Table S5, Figure S2). To investigate our primary research questions, we

15   used single-trial modelling to relate post-video hippocampal activation to subsequent false

16   memories. For our univariate analyses, we modelled a 2s impulse during the Post-Video period

17   (fixation screen), convolved with the canonical double-gamma hemodynamic response function

18   and phase-shifted 2s after video offset. This 2s shift targeted the peak Post-Video hippocampal

19   response identified in previous studies (49, 50). We isolated BOLD activation during the Post-

20   Video period on each trial and averaged parameter estimates across all voxels within each

21   hippocampal ROI (Methods, *Univariate fMRI Analyses*).

22   Some past studies have shown that prediction error signals are stronger in left

23   hippocampus and anterior hippocampus (18, 20, 21, 51), whereas posterior hippocampus is more

1   sensitive to video offsets (52). Other studies have shown that anterior and posterior hippocampus

2   parse continuous experience at different timescales (53, 54). On the basis of these findings, we

3   tested separate (non-overlapping) ROIs for left anterior, right anterior, left posterior, and right

4   posterior hippocampus (Supplementary Information, *Supplementary Methods, ROI Masks*).

5   Activation estimates from these four ROIs are included within each model to test for left/right

6   and anterior/posterior differences.

7   Using linear mixed-effects regression, we predicted trial-wise *hippocampal activation*

8   from the following variables: *reactivation type* (Full vs. Interrupted), *false memories* (continuous

9   measure), *hemisphere* (left vs. right), *axis* (anterior vs. posterior), and all relevant interactions.

10   We found a significant interaction between reactivation type and subsequent false memories

11   predicting hippocampal activation (β = -0.06, 95% CI [-0.09, -0.03], t = -4.33, p < .001) (Figure

12   3A) (Table S6). As predicted, this interaction demonstrated that the relationship between

13   hippocampal activation and subsequent memory differed between conditions. After Full videos,

14   greater hippocampal activation was associated with fewer subsequent false memories (Figure

15   3A, blue; β = -0.07, z = -2.81, p = .005), consistent with the idea that the hippocampus reinforces

16   memory for episodes that just concluded (49, 50, 55). However, we observed the opposite effect

17   when events were surprising. After Interrupted videos, greater hippocampal activation was

18   associated *more* subsequent false memories (Figure 3A, orange; β = 0.05, z = 2.23, p = .026),

19   consistent with the idea that surprise drives memory updating. Overall, this interaction

20   demonstrated that the *same amount* of hippocampal activation predicted different memory

21   outcomes depending on whether the video was Full (fewer false memories) or Interrupted (more

22   false memories).

1   Neither the main effect of reactivation type on hippocampal activation ($\beta = 0.04$, 95% CI

2   [-0.01, 0.09], $t = 1.85$, $p = .069$), nor the main effect of false memories on hippocampal

3   activation ($\beta = -0.01$, 95% CI [-0.05, 0.03], $t = -0.47$, $p = .640$) were statistically significant.

4   These null results demonstrate the value of examining the effect of prediction error on both

5   hippocampal activation and memory outcomes. There was a main effect of axis indicating that

6   average activation was greater in anterior hippocampus than posterior hippocampus ($\beta = 0.04$,

7   95% CI [0.01, 0.06, $t = 3.12$, $p = .002$). There was no main effect of hemisphere, and the axis and

8   hemisphere variables did not interact with reactivation type or false memories. Parameter

9   estimates for all variables are provided in Table S6.

10   Lastly, although our primary research questions pertained to false memories, we also

11   tested a model that included both false memories and correct details (Table S7). We found that

12   correct details also interacted with reactivation type to predict hippocampal activation ($\beta = 0.03$,

13   95% CI [0.01, 0.06], $t = 2.10$, $p = .036$), but numerically contrasted with our results for false

14   memories. After Full videos, hippocampal activation was positively associated with correct

15   details ($\beta = 0.01$, $z = 0.33$, $p = .742$), but was negatively associated with false memories ($\beta = $ -

16   0.06, $z = -2.44$, $p = .015$). After Interrupted videos, these associations reversed direction:

17   hippocampal activation was negatively associated with correct details ($\beta = -0.05$, $z = -1.55$, $p = $

18   .120), but was positively associated with false memories ($\beta = 0.05$, $z = 2.07$, $p = .039$). Notably,

19   although the overall interaction term was significant, the associations between hippocampal

20   activation and correct details did not significantly differ from zero. Importantly, the effect of

21   hippocampal activation on false memories remained unchanged after controlling for correct

22   details (Full: $\beta = -0.06$, $z = -2.44$, $p = .015$; Interrupted: $\beta = 0.05$, $z = 2.07$, $p = .039$),

1   demonstrating that the effect of memory updating was distinct from recall success for correct

2   details (Table S7).

Hippocampal-Basal Forebrain Connectivity Predicted Memory Outcomes

4   Next, we tested hypotheses about neuromodulatory mechanisms by examining activation

5   in the basal forebrain (which contains the medial septal nucleus, the primary source of

6   acetylcholine in the hippocampus) (29, 31, 37) and the VTA (which contains dopaminergic

7   neurons that project to the hippocampus) (16, 24, 56). First, we extracted average basal forebrain

8   and VTA activation during the Post-Video period (i.e., the 2-second segment of the Post-Video

9   fixation period, consistent with modelling of hippocampal activation).

10   Using linear mixed effects regression, we predicted *hippocampal activation* from the

11   variables *reactivation type* (Full vs. Interrupted), *false memories* (continuous), *basal forebrain*

12   *activation* (continuous), *hemisphere* (left vs. right), *axis* (anterior vs. posterior), and all relevant

13   interactions. There was a significant three-way interaction among basal forebrain activation,

14   reactivation type, and false memories predicting hippocampal activation (β = -0.04, 95% CI [-

15   0.06, -0.01], t = -2.66, p = .008) (Figure 3B). This interaction demonstrated that the relationship

16   between hippocampal activation and subsequent memory (Figure 3A) was evident only when the

17   basal forebrain was also strongly activated (Figure 3B, right) (Full: β = -0.12, z = -4.01, p < .001;

18   Interrupted: β = 0.07, z = 2.37, p = .018). When basal forebrain activation was weak,

19   hippocampal activation was unrelated to memory (Figure 3B, left) (Full: β = 0.01, z = -0.33, p =

20   .739; Interrupted: β = 0.04, z = 1.39, p = .166). There was also a main effect of basal forebrain

21   activation predicting hippocampal activation during the Post-Video period (β = 0.10, 95% CI

22   [0.02, 0.19], t = 2.34, p = .029). Other results from this extended model were consistent with the

1   base model (without basal forebrain variables) described above. There were no significant

2   interactions with hemisphere or axis. All parameter estimates are provided in Table S8.

3   Next, we examined the role of VTA activation. We modified the model described above

4   by replacing the basal forebrain activation variable (and interactions) with VTA activation

5   parameters. VTA activation was positively related to hippocampal activation during the Post-

6   Video period ($\beta = 0.15$, 95% CI [0.05, 0.25], $t = 3.03$, $p = .006$). However, there was no

7   interaction among VTA activation, reactivation type, and false memories ($\beta = -0.004$, 95% CI [-

8   0.03, 0.02], $t = -0.32$, $p = .747$). Thus, there was no evidence that VTA moderated the effect of

9   hippocampal activation on memory. All parameter estimates are provided in Table S9. In

10   separate model, we also tested for a main effect of prediction error on basal forebrain activation

11   (Table S10A) or VTA activation (Table S10B), and found no significant effects.
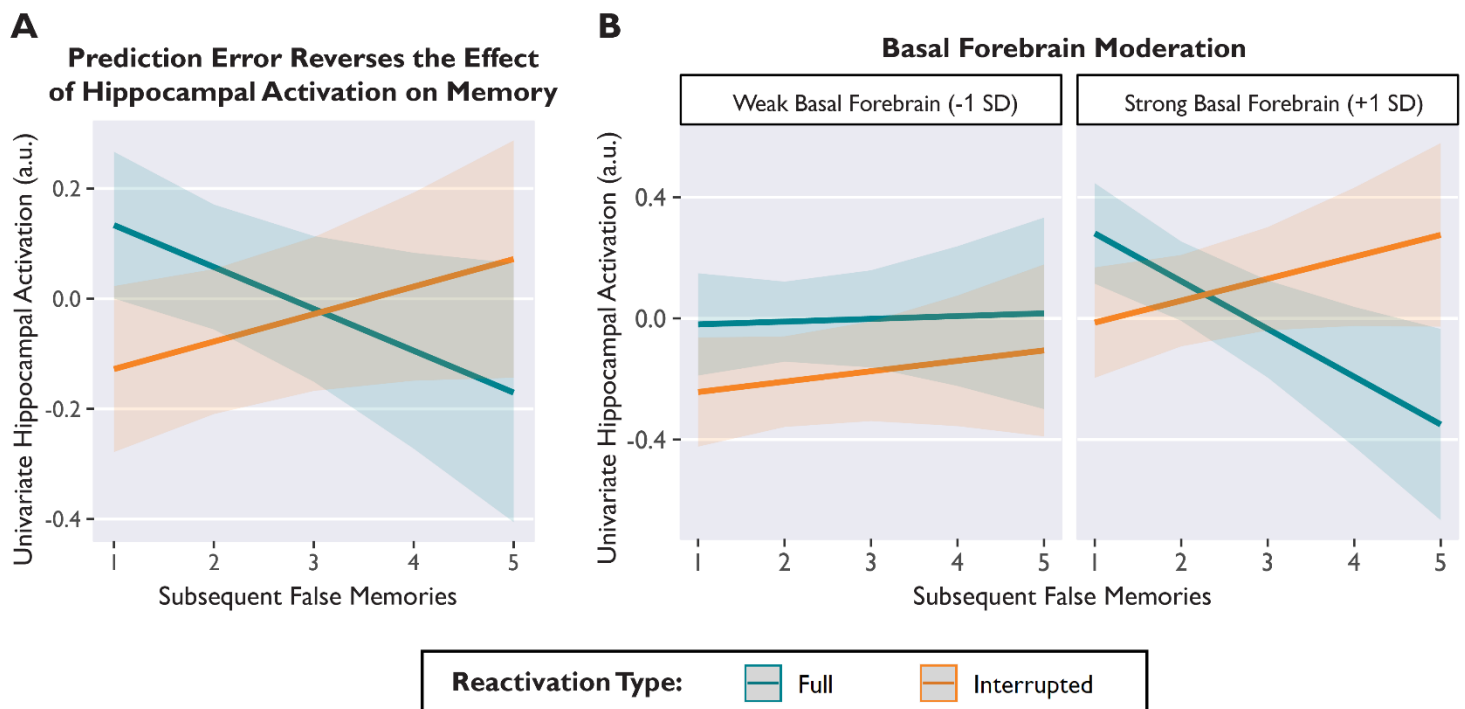

Figure 3. Prediction error reversed the relationship between average hippocampal activation (arbitrary units, a.u.) and subsequent memory, and this effect depended on concurrent basal forebrain activation. A) After Full-length videos, hippocampal activation was associated with memory *preservation*, predicting fewer false memories (blue). After Interrupted videos,

hippocampal activation was associated with memory *updating*, predicting more false memories (orange). B) The effect of prediction error on hippocampal activation and memory was observed only when basal forebrain activation was strong (right). When basal forebrain activation was weak, hippocampal activation was unrelated to memory (left). Basal forebrain activation is binned (weak vs. strong) for visualization, but statistical models used a continuous variable. Lines depict model-predicted estimates, and shaded bands depict the 95% confidence interval. Model-derived estimates are shown instead of individual data points in order to show within-subjects effects, while controlling for subject and stimulus variance.

Multivariate fMRI Results

1   Overall, our univariate results suggest that prediction error changed the role of the

2   hippocampus: the same magnitude of hippocampal activation predicted opposing effects on

3   memory depending on whether events were expected or surprising. Moreover, this effect of

4   hippocampal activation on memory depended on concurrent basal forebrain activation, consistent

5   with the idea that acetylcholine regulates hippocampal processing (27, 28). On the basis of our

6   univariate findings, we proposed that during video playback, the hippocampus continually

7   generates predictions and sustains episodic representations (57–59). If no prediction error is

8   detected, these representations should be sustained, and the hippocampus should preserve and

9   reinforce the memory (i.e., decreasing false memories). If a prediction error is detected, then the

10   hippocampus should abandon ongoing predictions and prepare to update a memory (i.e.,

11   increasing false memories) (28, 29). Therefore, we hypothesized that prediction errors would

12   disrupt sustained representations in the hippocampus, and that disrupting hippocampal

13   representations would lead to memory updating. Furthermore, we predicted that activation in the

14   basal forebrain and/or VTA would link hippocampal representations to memory outcomes, via

15   neuromodulation of hippocampal processing (24–29, 34, 60, 61).

16

1   ***Prediction Errors Disrupted Sustained Representations in the Hippocampus***

2   Past studies in rodents and humans have used *autocorrelation* measures, which quantify

3   similarity across neural patterns, to investigate hippocampal representations during naturalistic

4   tasks (53, 54). *Temporal autocorrelation* is an index of multivariate information that is preserved

5   over time; this measures moment-to-moment overlap of activation patterns (53, 58, 62).

6   Intracranial recordings in humans have shown that temporal autocorrelation in the hippocampus

7   ramps up over the course of familiar episodes (58). Ramping autocorrelation reflects sustained

8   neural representations, consistent with the hippocampus generating predictions and anticipating

9   upcoming stimuli (57, 58). To test whether hippocampal representations were sustained or

10   disrupted over time, we calculated temporal autocorrelation by correlating the activation of all

11   voxels within the hippocampus at timepoint *T* with the activation pattern at timepoint *T+1 sec*

12   (Methods, *Multivariate fMRI Analyses*). Importantly, we also investigated autocorrelation in two

13   control regions (inferior lateral occipital cortex and white matter) to demonstrate that these

14   representational changes were not a spurious whole-brain phenomenon (Supplementary

15   Information, *Autocorrelation Control*).

16   First, we tested whether autocorrelation increased during video playback (binned into 5-

17   second *video segments* over time). We used linear mixed effects regression to predict

18   *hippocampal autocorrelation* (averaged over 5s bins) from the variables *video segment*

19   (continuous), *hemisphere, axis,* and all interaction terms. We found that hippocampal

20   autocorrelation increased linearly as videos progressed ($\beta = 0.025$, 95% CI [0.01, 0.04], $t = 3.41$,

21   $p = .002$), suggesting that episodic representations were sustained and stabilized during video

22   playback (Figure 4A) (58). There were no significant effects of hemisphere or axis. All

23   parameter estimates are provided in Table S11A.

1   Next, we tested whether prediction error disrupted this ramping autocorrelation. To

2   analyze Post-Video *change in autocorrelation*, we calculated a difference score for each trial by

3   subtracting the average autocorrelation value from the 5s bin immediately *before* video offset

4   from the average autocorrelation value from the 5s video immediately *after* video offset. We

5   then used linear mixed effects regression to predict average *change in autocorrelation* from the

6   variables *reactivation type*, *hemisphere*, *axis*, and interaction terms. There was a significant main

7   effect of reactivation type ($\beta = 0.04$, 95% CI [0.01, 0.07], $t = 2.21$, $p = .038$), such that

8   autocorrelation increased after the offset of Full videos but not Interrupted videos (Figure 4C). In

9   other words, prediction errors disrupted the continuity of hippocampal representations. This

10   Post-Video divergence is visualized in Figure 4B. There were no significant interactions with

11   hemisphere or axis. All parameter estimates are provided in Table S11B.
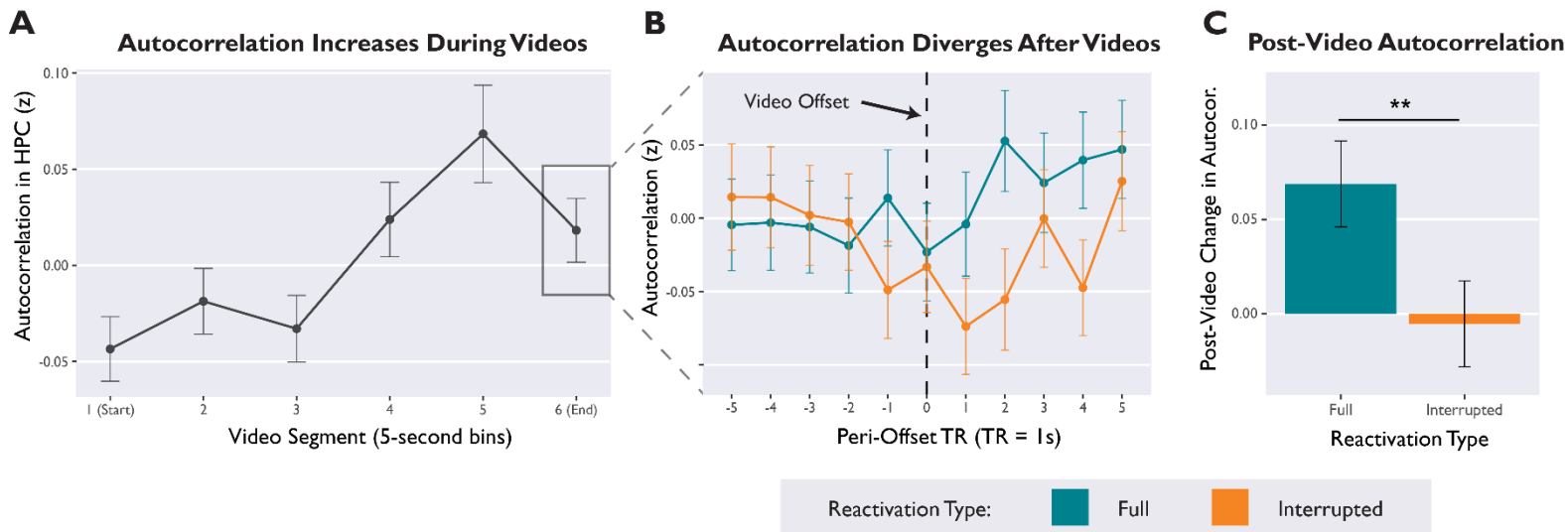
12

Figure 4. Hippocampal representations are sustained or disrupted over time, depending on whether or not episodes align with expectations. A) Temporal autocorrelation in the hippocampus gradually increased over the course of a video, suggesting that episodic representations were sustained over time. Autocorrelation values were averaged over 5-second bins of video playback. B) Autocorrelation trajectories for Full and Interrupted videos diverged during the Post-Video period. Plot visualizes second-by-second autocorrelation values in the hippocampus, time-locked to the moment of video offset (black dotted line). C) Average Post-

Video change in autocorrelation (average autocorrelation scores for the 5-sec bin immediately *after* video offset, minus average autocorrelation for the bin immediately *before* offset). Hippocampal representations were sustained after Full videos, but disrupted after Interrupted videos. Error bars depict SEM.

1   ***Disruption of Hippocampal Autocorrelation Predicted False Memories***

2   Next, we tested whether disruption of hippocampal representations predicted memory

3   updating. Using linear mixed effects regression, we predicted subsequent *false memories* from

4   the variables *reactivation type*, Post-Video *change in autocorrelation* (continuous), *hemisphere*,

5   *axis*, and all relevant interactions. We also included a continuous covariate for univariate

6   hippocampal activation (thus controlling for any autocorrelation effects that may be a

7   consequence of univariate activation). There was a significant interaction between reactivation

8   type and change in autocorrelation predicting false memories ($\beta = 0.04$, 95% CI [0.02, 0.07], $t =$

9   3.34, $p < .001$) (Figure 5A). After Interrupted videos, disrupting hippocampal representations led

10   to memory updating ($\beta = -0.06$, $z = -2.89$, $p = .004$). Conversely, after Full videos, hippocampal

11   autocorrelation was not related to false memories ($\beta = 0.02$, $z = 1.03$, $p = .303$). There were no

12   significant interactions with hemisphere or axis. All other parameters are reported in Table

13   S12A.

14   ***Basal Forebrain Activation Links Hippocampal Autocorrelation to Memory***

15   What determines whether hippocampal representations are sustained or disrupted? To

16   investigate candidate neuromodulatory mechanisms, we extended the model described above by

17   adding parameters for average Post-Video basal forebrain activation and interaction terms. The

18   model included all relevant interaction terms, reported in full in Table S12B. Paralleling our

19   univariate findings, we found a significant three-way interaction among basal forebrain

20   activation, reactivation type, and hippocampal autocorrelation that predicted subsequent false

21   memories ($\beta = 0.03$, 95% CI [0.01, 0.06], $t = 2.54$, $p = .011$) (Figure 5B). In other words,

1   prediction errors disrupted hippocampal representations and led to memory updating, but only

2   when the basal forebrain was also strongly activated (at +1SD basal forebrain activation,

3   Interrupted: $\beta = -0.07$, $z = -2.46$, $p = .014$; Full: $\beta = 0.08$, $z = 2.92$, $p = .004$) (Figure 5B, right).

4   Next, we tested whether VTA activation was related to hippocampal autocorrelation and

5   memory. We modified the model described above (predicting subsequent false memories) by

6   replacing the basal forebrain activation variable (and interaction terms) with the VTA activation

7   variable. In contrast to our basal forebrain findings, there was no three-way interaction among

8   VTA activation, reactivation type, and hippocampal autocorrelation ($\beta = -0.01$, $t = -0.54$, $p =$

9    $.587$, 95% CI [-0.03, 0.02]). All parameter estimates are reported in Table S12C. Overall, we

10   found that our autocorrelation results paralleled our univariate findings: basal forebrain

11   activation, but not VTA activation, was crucial for connecting hippocampal representations to
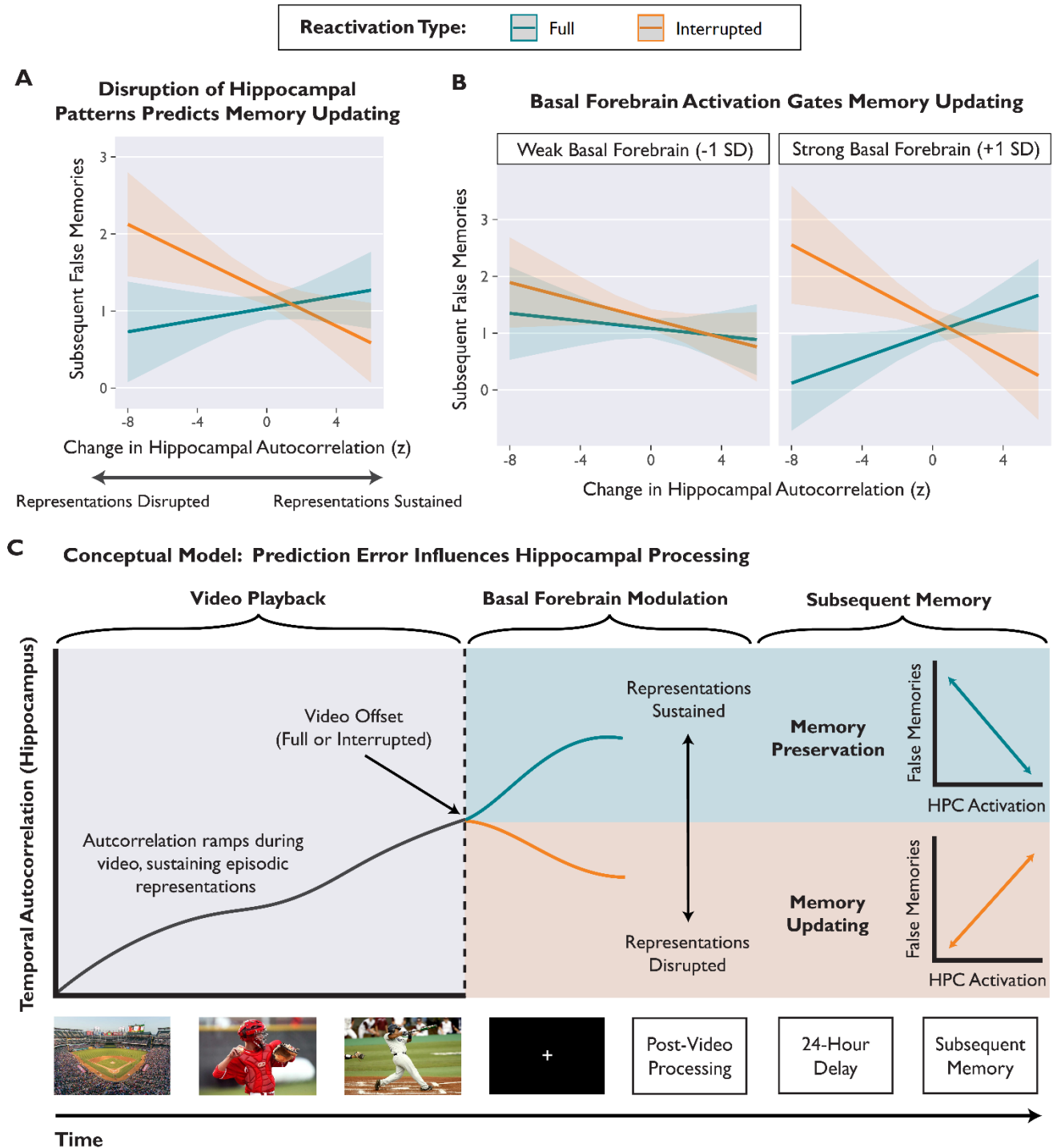
12   memory outcomes.

Figure 5. Prediction errors elicited by Interrupted videos disrupted sustained hippocampal representations, and this disruption predicted memory updating. A) Estimated values from a linear regression model predicting subsequent false memories from the interaction of reactivation type and change in autocorrelation. After Interrupted videos, decreases in autocorrelation were related to increased memory updating. B) The effect of prediction error on hippocampal autocorrelation and subsequent memory depended on concurrent basal forebrain activation.

Basal forebrain activation was binned (weak vs. strong) for visualization, but statistical models used a continuous variable. Shaded bands depict 95% confidence intervals around the regression line. In panels A and B, model-predicted estimates are depicted instead of individual data points in order to show within-subject effects, while controlling for subject and stimulus variability. C) Conceptual schematic depicting the effect of prediction error on hippocampal representations and subsequent memory. During a video, the hippocampus sustains episodic representations over time, consistent with generating ongoing predictions. After video offset, the hippocampus acts to preserve the memory (representations sustained) or prepare for memory updating (representations disrupted). The link between prediction error and memory outcomes depends on co-activation of the hippocampus and basal forebrain during the Post-Video period.

Discussion

1   Here, we show that prediction errors modulate the function of the hippocampus and allow

2   memories to be modified, consistent with an adaptive updating mechanism. In our fMRI

3   paradigm, we elicited mnemonic prediction errors by interrupting familiar narrative videos

4   immediately before the expected conclusions. Prediction errors reversed the relationship between

5   univariate hippocampal activation and subsequent memory: After expected video endings,

6   hippocampal activation was associated with memory preservation, but after prediction errors,

7   hippocampal activation was associated with memory updating. Tracking the stability of

8   hippocampal representations revealed that prediction errors disrupted activation patterns; this

9   pattern disruption predicted memory updating. Crucially, the association between hippocampal

10   activation (both univariate and multivariate) and memory outcomes depended on concurrent

11   basal forebrain activation during the Post-Video period. We conclude that prediction error,

12   coupled with basal forebrain modulation, prompts the hippocampus to abandon ongoing

13   predictions and prepare to update a memory by incorporating other information present in the

14   environment (Figure 5).

15

1 **Prediction Errors Disrupt Hippocampal Representations and Update Memories**

2   Past studies of mnemonic prediction errors have reported an increase in univariate

3 hippocampal activation, but have not examined whether this neural signal affects memory (17–

4 19, 63). For the first time, we show that after prediction errors, hippocampal activation leads to

5 memory updating. Crucially, we demonstrate that univariate measures are *insufficient* for

6 understanding the effect of prediction error on the hippocampus, because the same amount of

7 hippocampal activation can exert opposing effects on memory. Prediction error reversed the

8 relationship between hippocampal activation and subsequent memory, suggesting a shift in

9 hippocampal processing (Figure 3). After expected endings (Full videos), hippocampal activation

10 protected against false memories, consistent with the idea that the hippocampus reinforces

11 memory after the conclusion of an episode (49, 50). In contrast, after surprising endings

12 (Interrupted videos), hippocampal activation predicted *more* false memories, consistent with the

13 idea that prediction errors can destabilize memories and enable updating (1–3). Overall, this

14 reversal supports the idea that the hippocampus acts to *preserve* memories after expected events,

15 but *update* memories after surprising events. These divergent modes of processing parallel prior

16 studies on encoding vs. retrieval modes (20, 27, 64) and internal vs. external attention (28).

17   To test the idea that prediction errors influence hippocampal processing, we tracked

18 hippocampal activation patterns to examine how episodic representations were sustained or

19 disrupted over time. We used *temporal autocorrelation* (the moment-to-moment overlap of

20 activation patterns) as a measure of continuity in hippocampal representations (53, 58, 62). As

21 narratives progressed, autocorrelation increased, reflecting stability and continuity; this increase

22 in autocorrelation suggested that the hippocampus generated predictions (57, 58) and sustained

23 episodic representations over time (49, 65) (Figure 4A). Crucially, prediction errors disrupted the

1   stability of hippocampal representations (Figure 4B, 4C), and this disruption predicted the degree

2   of memory updating (Figure 5A). Overall, we propose that disruption of hippocampal

3   representations indicates a shift in processing: Prediction error prompts the hippocampus to

4   abandon ongoing predictions and prepare to update a memory (Figure 5C).

5   Our univariate findings also diverge from past studies of hippocampal prediction error

6   responses, which have shown that hippocampal activation increases after prediction errors (17–

7   19, 63). In contrast, our single-trial analyses showed no significant main effect of reactivation

8   type on hippocampal activation; the effect of prediction error was revealed only when examining

9   the link between hippocampal activation and subsequent memory. Moreover, our whole-brain

10   mass univariate analyses revealed a significant cluster in left hippocampus that was *less* activated

11   after Interrupted videos than Full videos (Supplementary Information, *Whole-Brain Analysis*).

12   Our task elicits surprise by omitting expected endings, comparable to a *negative* prediction error.

13   Previous studies have elicited surprise by replacing expected stimuli with novel stimuli, thus

14   adding new information, comparable to a *positive* prediction error. Therefore, this reversal of

15   activation (Full > Interrupted trials) is consistent with prior studies of reward (13) and

16   information prediction errors (4, 66), which have shown *increases* in neural activation after

17   positive prediction errors and *decreases* after negative prediction errors. Overall, our results

18   suggest that hippocampal responses may depend on the stimuli and type of surprise, and

19   demonstrate that prediction errors change the effect of hippocampal activation on memory.

20   **Basal Forebrain Activation Relates to Hippocampal Processing**

21   Past studies have suggested that either cholinergic (27–29, 35–37, 60) or dopaminergic

22   (16, 24, 61) modulation could regulate hippocampal function after prediction errors, such as by

23   enhancing plasticity and switching between processing modes. However, mixed evidence

1   supporting both hypotheses has left the question unresolved (25, 26, 31, 33, 34). Here, we

2   investigated whether activation of the basal forebrain or the VTA could explain the relationship

3   between hippocampal activation after prediction error and subsequent memory. We found that

4   the effect of prediction error on memory depended on co-activation of the hippocampus and

5   basal forebrain, suggesting that connectivity between these regions is important for shifting

6   hippocampal processing modes to either preserve or update a memory. Hippocampal activation

7   was associated with memory updating after prediction errors, but only when the basal forebrain

8   was also activated. Likewise, disrupting hippocampal representations led to memory updating

9   after prediction errors, but only when the basal forebrain was also activated. Although fMRI

10   cannot provide direct evidence of neuromodulation, our results are consistent with the idea that

11   cholinergic modulation from the basal forebrain (27, 35–37, 39) influences hippocampal

12   processing and memory outcomes.

13   Our findings are also relevant to the functional relationship between the VTA and

14   hippocampus. Although we found a robust positive correlation between VTA and hippocampal

15   activation during the Post-Video period, VTA activation was unrelated to prediction error and

16   did not link hippocampal activation with memory outcomes. These findings are consistent with

17   our prior proposal that connectivity between the VTA and hippocampus reflects modulation of

18   hippocampal learning states by sustained VTA activity (16, 69) rather than phasic VTA

19   responses (25, 70–72). However, our paradigm was optimized for detecting memory updating

20   instead of midbrain prediction error responses. It is also possible that prediction error signals

21   could be transmitted to the hippocampus from the locus coeruleus (71, 72). Analyses of locus

22   coeruleus in the current data did not reveal any relationships. Future research could disambiguate

1   the roles of the basal forebrain, VTA, and locus coeruleus by examining both event-related and

2   sustained connectivity with the hippocampus and their consequences for memory.

3   **Prediction Error Both Strengthens and Updates Memories**

4   Comparing behavioral results across the Delayed and Immediate groups revealed a

5   dissociation: prediction error both strengthened and updated memories, but over distinct

6   timecourses and likely via different mechanisms (Figure 2). Prediction error increased the

7   number of correct details recalled, both immediately and after a one-day delay. This finding is

8   consistent with recent evidence that mnemonic prediction errors can promote detailed memories

9   immediately (73), possibly by way of enhanced attention and pattern separation. In contrast, we

10   found that prediction error increased false memories only after a delay, consistent with prior

11   studies that have shown that memory updating requires a delay for protein synthesis to occur (1,

12   47, 74). Although it is possible that the relatively few false memories in the Immediate group

13   made it more difficult to detect an effect of prediction error, a control analysis showed that

14   excluding low-variance subjects from the Immediate group did not change our results

15   (Supplementary Information, *Behavioral Variance Control*). Overall, our finding that prediction

16   error increased both correct details and false memories supports the idea that surprise drives

17   adaptive updating.

18   In the present study, we used false memories as an index of memory updating. In the real

19   world, however, memory updating can be adaptive: new information is not "false" *per se*, but a

20   relevant addendum to prior knowledge. In our paradigm, interference from other stimulus videos

21   likely produced false memories because information was integrated across videos. Previously,

22   we found that prediction errors selectively updated memories with semantically-related

23   information from new videos that were specifically chosen to interfere with reactivated

1   memories (48). Here, we showed that videos that shared greater semantic similarity with the rest

2   of the stimulus set produced more false memories (Table S4). This finding suggests that

3   prediction error increases source confusion or integration of information across related

4   memories. Prior studies have shown that memory updating occurs when interference is

5   introduced after a memory trace is reactivated (1, 47, 75). In our paradigm, interference could

6   arise from semantically-related details that result from the subject recalling related memories, or

7   visual input from subsequent videos during the task. This finding accords with prior behavioral

8   studies (1, 47, 75) and computational models of event segmentation (76, 77), which have both

9   shown that interference among related episodes can produce false memories and/or source

10   confusion after prediction error. However, memory updating is beneficial in other situations that

11   require integrating old and new knowledge, or correcting erroneous information.

12   One possible mechanism for the memory updating we observed is *reconsolidation*, the

13   process by which reactivating a memory trace can temporarily render it malleable. However,

14   evidence for cellular reconsolidation processes in humans is lacking. Although previous human

15   studies have used reconsolidation-like paradigms to demonstrate memory malleability (4, 47, 48,

16   75), it remains unknown whether the synaptic mechanisms of reconsolidation are consistent

17   across animals and humans (1). A key prediction of reconsolidation theory is that memory

18   updating effects should only emerge after a delay, because the process of modifying and

19   restabilizing a memory trace requires protein synthesis that occurs over several hours. We found

20   that the behavioral effect of prediction error on memory updating required a delay, which aligns

21   with this theoretical prediction. Overall, our findings are broadly relevant to research on

22   prediction error and memory, and reconsolidation theory offers one possible framework.

23

Limitations and Future Directions

We were unable to directly test whether the relationship between neural activation and memory updating required a delay, because the fMRI participants always completed the delayed memory test and may show effects of memory updating as well as forgetting over time. Although our prior behavioral work demonstrated an effect of prediction error that was not confounded with the delay (48), scanning a group of participants who experience a delay-to-test without memory reactivation would enable an investigation of the neural correlates of immediate vs. delayed memory effects.

In the present study, we elicited surprise by interrupting videos before their expected narrative endings. However, we were not able to directly measure subjective surprise, because asking participants to rate surprise after each video would have disrupted other cognitive processes and revealed the goal of the manipulation. To ensure that Interrupted videos would continue to elicit surprise after the first few trials, we included the following features in our experimental design: 1) Full-length videos were presented twice during encoding to set strong expectations; 2) Interrupted videos violated strongly-expected action-outcome contingencies (e.g., a baseball batter halted mid-swing); 3) trials were pseudorandomized so that participants could not anticipate whether each video would be Full or Interrupted; and 4) participants did not know *when* each video would be interrupted. Consistent with these design considerations, we found that the effect of prediction error on false memories did not interact with trial number, suggesting that surprise did not diminish over the course of the experiment (Table S2).

Prior studies with humans and animals have used *incomplete reminders* (e.g., a conditioned stimulus without the expected outcome) to elicit prediction error (1, 3, 75). Here, we mimicked this approach by interrupting narrative videos and omitting the expected endings.

1   Incomplete reminders may be particularly effective because they elicit memory reactivation, and

2   memory reactivation supports plasticity (78–80). After Interrupted videos, participants may

3   actively retrieve the missing endings; this memory reactivation could contribute to memory

4   malleability. As discussed in our prior review (1), both prediction error and memory reactivation

5   may interact to support memory updating. Memory reactivation may be a prerequisite for

6   generating a prediction, and experiencing a prediction error may prompt further memory

7   reactivation. Future studies could directly investigate the role of memory reactivation by asking

8   participants to report their ability to recall the missing endings, or by testing encoding-retrieval

9   pattern similarity.

10   Lastly, a limitation of the present study is that we were unable to determine the temporal

11   distribution or the source of each individual false memory. Most details that participants recalled

12   (both correct and incorrect) pertained to the entire video, such as perceptual details about the

13   characters and setting. Many of these perceptual details were also shared across multiple videos

14   (e.g., several characters with green shirts), making it impossible to determine the source of each

15   specific detail. Interestingly, we found that the omitted endings on Interrupted trials were rarely

16   associated with false memories. Recall of the narrative content of the video endings was very

17   accurate: in the Delayed group, only 2% of all false memories specifically pertained to the

18   missing ending. Because the salient action-outcome contingencies define the videos, even when

19   interrupted, these central details were not likely to be affected. Instead, we found that prediction

20   error was more likely to induce a holistic distortion of perceptual content from throughout each

21   video (e.g., details about the setting or a character).

22

23

1 **Conclusion**

2   The brain continually generates predictions based on past experiences. When

3 expectations do not align with reality, memories should be updated with relevant new

4 information. We propose that mnemonic prediction errors prompt the hippocampus to abandon

5 ongoing predictions and update memories by incorporating relevant details from subsequent

6 experiences. In this way, surprise modulates hippocampal processing and determines the fate of

7 episodic memories. This theoretical framework of memory updating bears implications for

8 eyewitness testimony, education, and treating pathological memories (e.g., in Post-Traumatic

9 Stress Disorder). Beyond memory research, our results offer new insights for theories on the

10 whole-brain predictive processes that govern attention, perception, action, and decision-making.

Author Contributions: AHS and MDB developed the study design. AHS programmed the study, collected data, conducted analyses, and drafted the manuscript. GMM contributed substantially to data collection and IKB contributed to autocorrelation analyses. MDB and RAA contributed to the analysis approach and interpretation of results. All authors contributed to revising the manuscript and approved the final version.

Declaration of Interests: The authors have no competing interests to declare.

1   **Methods**

2   **Data, Code, and Materials**

3   Brief descriptions of the stimulus videos are provided in Table S13. The full set of

4   stimulus videos, along with derivative data and code necessary to reproduce results, are provided

5   online in the project repository hosted by the Open Science Framework (https://osf.io/xb7sq/).

6   **Participants**

7   We recruited 55 paid participants from the University of Toronto community (Delayed

8   group: $70, Immediate group: $40). Seven participants were excluded (see Supplementary

9   Information, *Exclusions*), yielding a final sample of 48 participants. The sample size was

10   determined *a priori* to satisfy the following conditions: (1) reproduce the sample size and

11   achieve at least 90% power to detect the interaction effect found in a prior study ($\eta_p^2 = 0.17$)

12   (48), and (2) evenly allocate participants to 6 pseudorandomized trial order lists. Participants

13   were healthy young adults (age: $M = 22.42$, $SD = 2.41$, range [18, 30]; gender: 75% female, 25%

14   male). Inclusion criteria were as follows: between the ages of 18-30, normal or corrected-to-

15   normal vision and hearing, no history of neurological or psychiatric disorders, and fluency in

16   English. fMRI participants were all right-handed. All participants provided informed consent,

17   and the study was approved by the University of Toronto Institutional Review Board, Protocol

18   #00035787.

19   In consideration of the effects of sleep on consolidation, we also asked participants to

20   report approximate hours of sleep over the course of the study. Participants slept an average of

21   7.28 hours ($SD = 1.31$) between the Day 1 and Day 2 sessions, and Delayed group participants

22   slept an average of 7.02 hours ($SD = 1$) between the Day 2 and Day 3 sessions.

1 **Stimuli**

2   Stimulus videos were sourced from movies, TV, and YouTube clips. We chose 70 videos

3 that featured distinct narrative events (duration $M = 30$ sec, $SD = 7$ sec). Semantic similarity

4 varied across videos (e.g., several videos featured sporting events), but there were no

5 overlapping scenes or characters. During pilot testing, we ensured that the videos would be

6 infrequently recognized by our participants. The 70 videos used in the experiment are described

7 in Table S13 and publicly available on the Open Science Framework (https://osf.io/xb7sq/). The

8 Interrupted version of each video ended abruptly at the narrative climax, omitting the salient

9 ending and violating expectations (duration $M = 25$ sec, $SD = 4$ sec).

10 **Procedure**

11   During the **Encoding session**, participants viewed all 70 videos in full-length form

12 (randomized order). Each video was presented twice in a row to ensure that participants had

13 strong expectations about the narrative outcomes for each video, a prerequisite for eliciting

14 prediction error later.

15   During the **Reactivation session**, participants viewed each video again a single time (35

16 Full videos, 35 Interrupted videos). Videos were played in a pseudorandom order (six trial order

17 lists, counterbalanced across participants) such that there were never more than two consecutive

18 Interrupted videos. Participants could not reliably anticipate whether each video would be

19 Interrupted, or where the interruption might occur. We also performed eye-tracking during the

20 Encoding and Reactivation sessions for participants in both the Delayed and Immediate groups

21 (EyeLink v.1000+, SR-Research). Eye-tracking was used to monitor alertness during the task,

22 but these data are not discussed further.

1   Lastly, the **Test session** involved a structured interview-style recall test about details

2   from each of the videos. Participants were cued with the name of each video and prompted to

3   recall the narrative. The experimenter then probed the participant for more information with a

4   pre-determined list of questions (e.g., "Can you describe the setting or context of the video?",

5   "Can you describe what the character looked like? Do you remember gender, age range, hair

6   color, or clothing?"). Participants were instructed to answer based on their memory of the Full-

7   length videos that had been originally presented during encoding. Because we were interested in

8   false memories as a measure of memory modification, we instructed participants not to guess and

9   permitted them to skip details they could not recall.

10   Overall, the experiment took place over three days for participants in the Delayed group

11   (24-hour delays between Encoding, Reactivation, and Test), or over two days for participants in

12   the Immediate group (24-hour delay between Encoding and Reactivation, no delay between

13   Reactivation and Test). Only the Delayed group underwent neuroimaging.

14   Consistent with past studies (81–83), we maintained consistent contextual factors

15   between Encoding, Reactivation, and Test sessions. Delayed group participants completed the

16   encoding session in a mock scanner (shell of a retired 1.5T Siemens Avanto scanner), while

17   recorded MRI sounds were played in the background. The mock scanner room was adjacent to

18   the real scanner room, and was very similar in size and appearance. Delayed group participants

19   completed the Reactivation session in the real fMRI scanner and the Test session at a desk in the

20   mock scanner room. Participants in the Immediate group completed all three sessions in the same

21   behavioral testing room. In both groups, participants completed all three sessions with the same

22   experimenter.

1 **Statistical Analysis**

2   For both behavioral and neural data, we conducted trial-wise analyses with linear mixed-

3 effects regression models. Detailed information about the random effects specification is

4 provided in the Supplementary Information (*Supplementary Methods, Linear Mixed-Effects*

5 *Regression*).

6 *Univariate fMRI Analyses*

7   Whole-brain mass univariate results are reported in the Supplementary Information

8 (*Whole-Brain Analysis*, Figure S2, Table S5). Details about preprocessing steps and region-of-

9 interest masks are provided in the Supplementary Information (*Supplementary Methods, fMRI*

10 *Preprocessing, Region of Interest Masks*). The primary findings reported in the main text reflect

11 a single-trial modelling approach that estimated hippocampal responses to each video during the

12 task. In order to isolate responses on each trial, we employed the Least Squares-Single approach

13 and constructed a separate GLM for each trial (84, 85). We modelled each trial as a 2s impulse in

14 the Post-Video period, convolved with the canonical double-gamma hemodynamic response

15 function and phase-shifted 2s after video offset. This 2s shift targets the peak hippocampal

16 response previously identified in studies of post-video processing (49, 50). Within each GLM,

17 the target trial (2s event) was isolated as one regressor, and all other events were modelled with a

18 separate regressor for each type of event (e.g., video playback, video name cues, other fixation

19 periods). For each trial, we masked the processed data and averaged across voxels within each

20 ROI to yield an average activation value.

21 *Multivariate fMRI Analyses*

22   Multivariate temporal autocorrelation analyses (53, 58) were conducted on preprocessed

23 data (prior to single-trial GLM analysis). We extracted the whole-run timeseries from every

1   voxel within each ROI using the *fslmeants* utility. For control analyses (white matter and iLOC

2   ROIs), autocorrelation was calculated on 200 contiguous voxels, approximately matching the

3   size of the hippocampal ROIs. Temporal autocorrelation was defined as the Pearson product-

4   moment correlation between all voxel activation values at timepoint T and timepoint T+1s. This

5   method yielded an autocorrelation value for every second of each functional run, excluding the

6   final TR. Autocorrelation values were standardized (Fisher's *z*) prior to statistical analysis.

7   Next, we aligned multivariate timeseries data with event onset and duration markers.

8   Comparable to past research, we phase-shifted the timeseries by 4 seconds in order to account for

9   HRF lag (86). This manual shifting is necessary because event onset regressors have not yet been

10   convolved with the HRF (unlike in standard GLM analyses used for our univariate analyses).

11   Note that our univariate analyses involved an additional 2s shift in addition to the standard HRF

12   shift; this allowed us to target the peak post-video hippocampal response on each trial, but was

13   not necessary for the autocorrelation analyses that yield TR-by-TR values.

14   After alignment, we calculated average autocorrelation values that were time-locked to

15   events. For statistical analyses, autocorrelation values were averaged across 5-second bins during

16   and after each video. To analyze signal history over the course of video playback, we related the

17   video segment number (5s bins) to average autocorrelation values. For each video, we included

18   the first five seconds (timepoints 0-4), the next four middle segments (timepoints 5-9, 10-14, 15-

19   19, and 20-24), and the last five seconds (variable depending on the length of the video). This

20   binning scheme spans the average video length of 30 seconds; additional middle segments from

21   videos that were longer than 30 seconds were omitted. Lastly, to analyze Post-Video changes in

22   autocorrelation, we calculated trial-by-trial difference scores by subtracting the average

23   autocorrelation value for the 5-second bin immediately *before* video offset from the average

1   autocorrelation value for the 5-second bin immediately *after* video offset. Autocorrelation values

2   and difference scores for each trial were then submitted to linear mixed effects regression.

References

1. A. H. Sinclair, M. D. Barense, Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation. *Trends Neurosci.* **42**, 727–739 (2019).

2. R. N. Henson, P. Gagnepain, Predictive, Interactive Multiple Memory Systems. *Hippocampus* **20**, 1315–1326 (2010).

3. M. T. J. Exton-Mcguinness, J. L. C. Lee, A. C. Reichelt, Updating memories—The role of prediction errors in memory reconsolidation. *Behav. Brain Res.* **278**, 375–384 (2015).

4. A. Pine, N. Sadeh, A. Ben-Yakov, Y. Dudai, A. Mendelsohn, Knowledge acquisition is governed by striatal prediction errors. *Nat. Commun.* **9**, 1–14 (2018).

5. G. Kim, K. A. Norman, N. B. Turk-Browne, Neural differentiation of incorrectly predicted memories. *J. Neurosci.* **37**, 2022–2031 (2017).

6. J. A. Quent, R. N. Henson, A. Greve, A predictive account of how novelty influences declarative memory. *Neurobiol. Learn. Mem.* **179**, 107382 (2021).

7. D. L. Schacter, D. R. Addis, The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 773–786 (2007).

8. L. F. Barrett, W. K. Simmons, Interoceptive predictions in the brain. *Nat. Rev. Neurosci.* **16**, 419–429 (2015).

9. K. Friston, The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).

10. G. B. Keller, T. D. Mrsic-Flogel, Predictive Processing: A Canonical Cortical Computation. *Neuron* **100**, 424–435 (2018).

11. M. W. Spratling, A review of predictive coding algorithms. *Brain Cogn.* **112**, 92–97 (2017).

12. R. S. Sutton, A. G. Barto, *Reinforcement learning: An introduction*, 2nd Ed. (MIT Press, 1998).

13. M. Watabe-Uchida, N. Eshel, N. Uchida, Neural circuitry of reward prediction error. *Annu. Rev. Neurosci.* **40**, 373–394 (2017).

14. N. E. Wheeler, *et al.*, Ideology and predictive processing: coordination, bias, and polarization in socially constrained error minimization. *Curr. Opin. Behav. Sci.* **34**, 192–198 (2020).

15. N. Hindy, E. Avery, N. Turk-Browne, Hippocampal-neocortical interactions sharpen over time for predictive actions. *Nat. Commun.* **10** (2019).

16. D. Shohamy, R. A. Adcock, Dopamine and adaptive memory. *Trends Cogn. Sci.* **14**, 464–472 (2010).

17. J. Chen, P. A. Cook, A. D. Wagner, Prediction strength modulates responses in human area CA1 to sequence violations. *J. Neurophysiol.* **114**, 1227–1238 (2015).

18. K. Duncan, N. Ketz, S. J. Inati, L. Davachi, Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus* **22**, 389–398 (2012).

19. D. Kumaran, E. A. Maguire, An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS Biol.* **4**, e424 (2006).

20. O. Bein, K. Duncan, L. Davachi, Mnemonic prediction errors bias hippocampal states. *Nat. Commun.* **11**, 3451 (2020).

21. J. Chen, R. K. Olsen, A. R. Preston, G. H. Glover, A. D. Wagner, Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. *Learn. Mem.* **18**, 523–528 (2011).

22. K. Duncan, C. Curtis, L. Davachi, Distinct memory signatures in the hippocampus: intentional states distinguish match and mismatch enhancement signals. *J. Neurosci.* **29**, 131–139 (2009).

23. K. C. Dickerson, J. Li, M. R. Delgado, Parallel contributions of distinct human memory systems during probabilistic learning. *NeuroImage* **55**, 266–276 (2011).

24. J. E. Lisman, A. A. Grace, The Hippocampal-VTA Loop: Controlling the Entry of Information into Long-Term Memory. *Neuron* **46**, 703–713 (2005).

25. J. E. Lisman, N. A. Otmakhova, Storage, recall, and novelty detection of sequences by the hippocampus: elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus* **11**, 551–568 (2001).

26. J. Schomaker, M. Meeter, Short- and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neurosci. Biobehav. Rev.* **55**, 268–279 (2015).

27. M. Hasselmo, B. Wyble, G. Wallenstein, Encoding and Retrieval of Episodic Memories: Role of Cholinergic and GABAergic Modulation in the Hippocampus. *Hippocampus* **6** (1996).

28. C. J. Honey, E. L. Newman, A. C. Schapiro, Switching between internal and external modes: A multiscale learning principle. *Netw. Neurosci.* **1**, 339–356 (2017).

29. M. Meeter, J. M. J. Murre, L. M. Talamini, Mode shifting between storage and recall based on novelty detection in oscillating hippocampal circuits. *Hippocampus* **14**, 722–741 (2004).

30. B. E. Sherman, N. B. Turk-Browne, Statistical prediction of the future impairs episodic encoding of the present. *Proc. Natl. Acad. Sci.* **8** (2020).

31. M. F. Carr, L. M. Frank, A single microcircuit with multiple functions: state dependent information processing in the hippocampus. *Curr. Opin. Neurobiol.* **22**, 704–708 (2012).

32. M. M. Chun, J. D. Golomb, N. B. Turk-Browne, A taxonomy of external and internal attention. *Annu. Rev. Psychol.* **62**, 73–101 (2011).

33. H. Tarder-Stoll, M. Jayakumar, H. R. Dimsdale-Zucker, E. Günseli, M. Aly, Dynamic internal states shape memory retrieval. *Neuropsychologia* **138**, 107328 (2020).

34. C. E. Wideman, K. H. Jardine, B. D. Winters, Involvement of classical neurotransmitter systems in memory reconsolidation: Focus on destabilization. *Neurobiol. Learn. Mem.* **156**, 68–79 (2018).

35. E. L. Newman, S. N. Gillet, J. R. Climer, M. E. Hasselmo, Cholinergic blockade reduces theta-gamma phase amplitude coupling and speed modulation of theta frequency consistent with behavioral effects on encoding. *J. Neurosci.* **33**, 19635–19646 (2013).

36. L. M. Giocomo, M. E. Hasselmo, Neuromodulation by glutamate and acetylcholine can change circuit dynamics by regulating the relative influence of afferent input and excitatory feedback. *Mol. Neurobiol.* **36**, 184–200 (2007).

37. M. E. Hasselmo, The role of acetylcholine in learning and memory. *Curr. Opin. Neurobiol.* **16**, 710–715 (2006).

38. C. Kemere, M. F. Carr, M. P. Karlsson, L. M. Frank, Rapid and Continuous Modulation of Hippocampal Network State during Exploration of New Places. *PLOS ONE* **8**, e73114 (2013).

39. A. L. Decker, K. Duncan, Acetylcholine and the complex interdependence of memory and attention. *Curr. Opin. Behav. Sci.* **32**, 21–28 (2020).

40. N. Bunzeck, E. Düzel, Absolute Coding of Stimulus Novelty in the Human Substantia Nigra/VTA. *Neuron* **51**, 369–379 (2006).

41. B. C. Wittmann, N. Bunzeck, R. J. Dolan, E. Düzel, Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *NeuroImage* **38**, 194–202 (2007).

42. V. P. Murty, R. A. Adcock, Enriched encoding: reward motivation organizes cortical networks for hippocampal detection of unexpected events. *Cereb. Cortex N. Y. N 1991* **24**, 2160–8 (2014).

43. D. Shohamy, A. D. Wagner, Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron* **60**, 378–389 (2008).

44. R. A. Adcock, A. Thangavel, S. Whitfield-Gabrieli, B. Knutson, J. D. E. Gabrieli, Reward-Motivated Learning: Mesolimbic Activation Precedes Memory Formation. *Neuron* **50**, 507–517 (2006).

45. A. Tompary, K. Duncan, L. Davachi, Consolidation of associative and item memory is related to post-encoding functional connectivity between the ventral tegmental area and different medial temporal lobe subregions during an unrelated task. *J. Neurosci.* **35**, 7326–7331 (2015).

46. K. Nader, E. O. Einarsson, Memory reconsolidation: an update. *Ann N Acad Sci* **1191**, 27–41 (2010).

47. A. Hupbach, R. Gomez, L. Nadel, "Episodic memory reconsolidation: An update" in *Memory Reconsolidation*, C. M. Alberini, Ed. (Elsevier Academic Press, 2013), pp. 233–247.

48. A. H. Sinclair, M. D. Barense, Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learn. Mem.* **25**, 369–381 (2018).

49. A. Ben-Yakov, N. Eshel, Y. Dudai, Hippocampal Immediate Poststimulus Activity in the Encoding of Consecutive Naturalistic Episodes. *J. Exp. Psychol. Gen.* **142**, 1255–1263 (2013).

50. A. Ben-Yakov, Y. Dudai, Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. *J. Neurosci.* **31**, 9032–9042 (2011).

51. R. A. Cooper, M. Ritchey, Progression from feature-specific brain activity to hippocampal binding during episodic encoding. *J. Neurosci.* **40**, 1701–1709 (2020).

52. Z. M. Reagh, A. I. Delarazan, A. Garber, C. Ranganath, Aging alters neural activity at event boundaries in the hippocampus and Posterior Medial network. *Nat. Commun.* **11**, 1–12 (2020).

53. I. K. Brunec, *et al.*, Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans. *Curr. Biol.* **28**, 2129-2135.e6 (2018).

54. K. B. Kjelstrup, *et al.*, Finite scale of spatial representation in the hippocampus. *Science* **321**, 140–143 (2008).

55. C. Baldassano, *et al.*, Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709-721.e5 (2017).

56. M. Zubair, *et al.*, Divergent Whole Brain Projections from the Ventral Midbrain in Macaques. *Cereb. Cortex N. Y. N 1991* **31**, 2913–2931 (2021).

57. R. U. Haque, S. K. Inati, A. I. Levey, K. A. Zaghloul, Feedforward prediction error signals during episodic memory retrieval. *Nat. Commun.* **11**, 6075 (2020).

58. R. Paz, *et al.*, A neural substrate in the human hippocampus for linking successive events. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6046–6051 (2010).

59. A. Tambini, L. Davachi, Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19591–6 (2013).

60. K. H. Jardine, *et al.*, Activation of cortical M1 muscarinic receptors and related intracellular signaling is necessary for reactivation-induced object memory updating. *Sci. Rep.* **10**, 9209 (2020).

61. J. I. Rossato, *et al.*, State-dependent effect of dopamine D1/D5 receptors inactivation on memory destabilization and reconsolidation. *Behav. Brain Res.* **285**, 194–199 (2015).

62. R. Raut, A. Snyder, M. Raichle, Hierarchical dynamics as a macroscopic organizing principle of the human brain. *PNAS* **117**, 20890–20897 (2020).

63. D. Kumaran, E. A. Maguire, Match-mismatch processes underlie human hippocampal responses to associative novelty. *J. Neurosci.* **27**, 8517–8524 (2007).

64. K. Duncan, A. Tompary, L. Davachi, Associative encoding and retrieval are predicted by functional connectivity in distinct hippocampal area CA1 pathways. *J. Neurosci.* **34**, 11188–11198 (2014).

65. S. DuBrow, L. Davachi, Temporal binding within and across events. *Neurobiol. Learn. Mem.* **134**, 107–114 (2016).

66. E. S. Bromberg-Martin, O. Hikosaka, Lateral habenula neurons signal errors in the prediction of reward information. *Nat. Neurosci.* **14**, 1209–1218 (2011).

67. V. P. Murty, I. C. Ballard, R. A. Adcock, Hippocampus and Prefrontal Cortex Predict Distinct Timescales of Activation in the Human Ventral Tegmental Area. *Cereb. Cortex N. Y. NY* **27**, 1660–1669 (2017).

68. J. Lisman, A. A. Grace, E. Duzel, A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends Neurosci.* **34**, 536–547 (2011).

69. A. A. Grace, S. B. Floresco, Y. Goto, D. J. Lodge, Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends Neurosci.* **30**, 220–227 (2007).

70. C. G. McNamara, Á. Tejero-Cantero, S. Trouche, N. Campo-Urriza, D. Dupret, Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nat. Neurosci.* **17**, 1658–1660 (2014).

71. C. G. McNamara, D. Dupret, Two sources of dopamine for the hippocampus. *Trends Neurosci.* **40**, 383–384 (2017).

72. K. A. Kempadoo, E. V. Mosharov, S. J. Choi, D. Sulzer, E. R. Kandel, Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proc. Natl. Acad. Sci.* **113**, 14835–14840 (2016).

73. O. Bein, N. A. Plotkin, L. Davachi, Mnemonic prediction errors promote detailed memories (2021) https:/doi.org/10.31234/osf.io/b5z8a (May 11, 2021).

74. K. Nader, Memory traces unbound. *Trends Neurosci.* **26**, 65–72 (2003).

75. C. Forcato, M. L. C. Rodríguez, M. E. Pedreira, H. Maldonado, Reconsolidation in humans opens up declarative memory to the entrance of new information. *Neurobiol. Learn. Mem.* **93**, 77–84 (2010).

76. N. T. Franklin, K. A. Norman, C. Ranganath, J. M. Zacks, S. J. Gershman, Structured Event Memory: A neuro-symbolic model of event cognition. *Psychol. Rev.* **127**, 327–361 (2020).

77. G. A. Radvansky, J. M. Zacks, Event boundaries in memory and cognition. *Curr. Opin. Behav. Sci.* **17**, 133–140 (2017).

78. V. J. H. Ritvo, N. B. Turk-Browne, K. A. Norman, Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends Cogn. Sci.* **23**, 726–742 (2019).

79. D. Stawarczyk, C. Wahlheim, J. Etzel, A. Snyder, J. Zacks, Aging and the encoding of event changes: The role of neural activity pattern reinstatement. *bioRxiv* (2020) https:/doi.org/https://doi.org/10.1101/809806.

80. D. Stawarczyk, M. A. Bezdek, J. M. Zacks, Event Representations and Predictive Processing: The Role of the Midline Default Network Core. *Top. Cogn. Sci.* (2019) https:/doi.org/10.1111/tops.12450.

81. A. M. Capelo, P. B. Albuquerque, S. Cadavid, Exploring the role of context on the existing evidence for reconsolidation of episodic memory. *Memory* **27**, 280–294 (2019).

82. A. Hupbach, R. Gomez, L. Nadel, Episodic memory updating: The role of context familiarity. *Psychon. Bull. Rev.* **18**, 787–797 (2011).

83. A. Hupbach, R. Gomez, O. Hardt, L. Nadel, Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learn. Mem.* **14**, 47–53 (2007).

84. J. A. Mumford, B. O. Turner, F. G. Ashby, R. A. Poldrack, Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage* **59**, 2636–2643 (2012).

85. J. A. Mumford, T. Davis, R. A. Poldrack, The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage* (2014) https:/doi.org/10.1016/j.neuroimage.2014.09.026.

86. T. Sadeh, J. Chen, Y. Goshen-Gottstein, M. Moscovitch, Overlap between hippocampal pre-encoding and encoding patterns supports episodic memory. *Hippocampus* (2019) https:/doi.org/10.1002/hipo.23079.