



Contextual incongruency triggers memory reinstatement and the disruption of neural stability

Xiongbo Wu^{a,b,c,1,*}, Pau A. Packard^d, Josué García-Arch^{a,b,c}, Nico Bunzeck^{e,f},
Lluís Fuentemilla^{a,b,c}

^a Cognition and Brain Plasticity Group, Bellvitge Institute for Biomedical Research, Hospitalet de Llobregat 08907, Spain

^b Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona 08035, Spain

^c Institute of Neurosciences, University of Barcelona, Barcelona 08035, Spain

^d Multisensory Research Group, Center for Brain and Cognition, Pompeu Fabra University, Barcelona, Spain

^e Department of Psychology, University of Lübeck, Lübeck 23562, Germany

^f Center of Brain, Behavior and Metabolism (CBBM), University of Lübeck, Lübeck 23562, Germany

ARTICLE INFO

Keywords:

Episodic memory
EEG
Representational similarity analysis
Neural reactivation
N400
Congruence

ABSTRACT

Schemas, or internal representation models of the environment, are thought to be central in organising our everyday life behaviour by giving stability and predictiveness to the structure of the world. However, when an element from an unfolding event mismatches the schema-derived expectations, the coherent narrative is interrupted and an update to the current event model representation is required. Here, we asked whether the perceived incongruence of an item from an unfolding event and its impact on memory relied on the disruption of neural stability patterns preceded by the neural reactivation of the memory representations of the just-encoded event. Our study includes data from two different experiments whereby human participants ($N = 33$, 26 females and $N = 18$, 16 females, respectively) encoded images of objects preceded by trial-unique sequences of events depicting daily routine. We found that neural stability patterns gradually increased throughout the ongoing exposure to a schema-consistent episode, which was corroborated by the re-analysis of data from two other experiments, and that the brain stability pattern was interrupted when the encoding of an object of the event was incongruent with the ongoing schema. We found that the decrease in neural stability for low-congruence items was seen at ~ 1000 ms from object encoding onset and that it was preceded by an enhanced N400 ERP and an increased degree of neural reactivation of the just-encoded episode. Current results offer new insights into the neural mechanisms and their temporal orchestration that are engaged during online encoding of schema-consistent episodic narratives and the detection of incongruencies.

1. Introduction

Experience is guided by internal representation models of the environment, or knowledge schemas, with an impact on perception and memory (Gilboa and Marlatte, 2017). Schemas are thought to be central in organising our everyday life behaviour by giving stability and predictability to the structure of the world (Gershman et al., 2014). Thus, despite the ever-changing sequence of inputs of our experience, schemas bring relatedness and comprehension of unfolding events by anticipating stereotyped or congruent-like elements to encounter next. A computational advantage to the memory systems is, therefore, that schema-consistent items can be added to an existing schema without requiring alterations or extensions to it (McClelland et al., 2020).

Accordingly, when elements of the unfolding experience are congruent with expected representations from a currently activated schema, they are rapidly integrated into the memory model of the event (McClelland et al., 2020; Tse et al., 2007, 2011; van Kesteren et al., 2014). However, when our predictions are incorrect, we must update our internal models of the world to support adaptive behaviour. Nevertheless, the neural mechanisms that support memory integration and updating of an unfolding event remain unclear.

Previous fMRI results showed activity invariance during the encoding of different, but schema-consistent episodes (Baldassano et al., 2018), lending support to the notion that event encoding activates knowledge schema representations about the expected structure of the world. Thus, if an internal memory representation (or a schema) is

* Corresponding author at: Department of Psychology, Ludwig-Maximilians-University, Leopoldstraße 13, 80802 Munich, Germany.
E-mail address: Xiongbo.Wu@psy.lmu.de (X. Wu).

¹ Current address: Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

stable during the unfolding experience of an event, then some properties of its underlying neural implementations may also exhibit invariance during encoding. Indeed, neuroimaging studies in humans observed stable brain patterns of activity during the encoding of continuous streams of schema consistent audio-visual inputs and that shifts in neural stability are coincident with the detection of unexpected elements in the unfolding stream (Baldassano et al., 2017). Similarly, Sinclair et al. (2021) recently showed that hippocampal activation patterns stabilised during the encoding of a narrative episode, akin to sustained representations accumulated during an unfolding schema-congruent event. Intriguingly, this study also revealed that when the narrative was suddenly interrupted, the ongoing stability of the neural activity became disrupted, reflecting the need to update the sustained representation of the event model.

However, the notion that the detection of incongruencies of an unfolding event engenders a disruption of the ongoing representation challenges a set of findings that found no memory disturbance or even improvement for surprising events (e.g., Chen et al., 2015; Frank et al., 2018, 2020; Greve et al., 2017, 2019; Pine et al., 2018; Quent et al., 2021; Rouhani et al., 2018). This literature relies on the idea that mnemonic prediction error enhances hippocampal biases toward encoding (Bein et al., 2020) and that this shift in encoding strategy reflects the need to evaluate and, if necessary, update the representational content of the ongoing experience with the current incongruent event.

How does the brain accommodate these two seemingly opposite lines of research evidence, namely, that mnemonic prediction errors disrupt ongoing neural representations of unfolding events and, at the same time, promote the update of the ongoing memory model during encoding? Here, we asked whether this process is supported by distinct brain mechanisms that occur rapidly (in the order of milliseconds) but are sequentially orchestrated over time. More specifically, drawing on past theoretical (McClelland et al., 2020) and empirical research (Silva et al., 2019; Sols et al., 2017; Wu et al., 2022), we hypothesise that the subjective degree of an item's congruence with an unfolding experience is determined by an evaluation process guided by a rapid reactivation of the encoded event. The concomitant representation of the new element and the reactivated memory of the just-encoded event would promote the effective and rapid assessment of the extent to which the novel element matches, or mismatches expectations driven by the unfolding event. As a result, the brain would be able to either assimilate the new item with the ongoing memory representation by preserving a stable state of neural pattern of activity or, alternatively, disrupt it to promote its update.

To test this hypothesis, we recorded scalp electrophysiological (EEG) activity while healthy participants encoded images of target objects preceded by trial-unique sequences of four pictures of events depicting a routine from everyday life (Fig. 1a). The sequence of pictures preceding the target object image was thought to mimic a realistic unfolding episodic event with the aim to provide a gradual schema-consistent narrative that determined whether specific target objects matched or mismatched expected occurrences within that context. Importantly, participants were instructed to rate the perceived congruence of the item in relation to the previously encoded event sequence episode, thereby allowing us to assess the degree of perceived congruence of the target object for every single trial at an individual level. To examine how object congruence shaped memory for the target object, a surprise recognition memory test was administered to the participants after the encoding phase. In Experiment 1, we first asked participants to indicate whether a label word referred to object pictures encoded in the previous phase, and if so, to recognize which of two very similar pictures was the exact one presented during the encoding phase (Fig. 1b). This later test allowed assessment of the extent to which encoding congruence detailed memory representation (e.g., Bein et al., 2020). We ran one additional follow-up behavioural experiment on a separate sample of participants to further scrutinise congruence-shaped long-term memory. This experiment consisted of a similar structure and materials used in Experiment 1 but differed in the format of the recognition phase. In Experiment 2,

we asked participants to indicate whether object pictures were encoded together with a selected picture of the episodic sequence, thereby allowing us to assess the ability of the participants to bind the object to the preceding episodic context (Fig. 1c). This test allowed examining whether the perceived congruence of the object within an episodic narrative influenced how the two become associated in long-term memory.

2. Material and methods

2.1. Participants

Participants were healthy college students from the University of Barcelona who had normal or corrected-to-normal vision and reported no history of medical, neurological or psychiatric disorders. Thirty-three participants (26 females, $M = 20.94$ years, $SD = 3.24$ years) were recruited and were paid €10/h for their participation in Experiment 1. Four participants were excluded due to loss of EEG data for technical reasons. Eighteen participants (16 female, $M = 23.05$ years, $SD = 6.55$) were recruited and paid €5/hour for their participation in the follow up Experiment 2. All participants signed informed consent, approved by the University of Barcelona Ethics Committee.

2.2. Stimuli

Experimental stimuli consisted of 160 photographs of household objects and 80 episodic sequences, each comprising 4 photographs. There were 80 different household objects included each with two slightly different versions, for a total of 160 photographs. Episodic sequences consisted of 4 snapshots in temporal order depicting the same person moving around and interacting with the surroundings in different rooms from the same house during a short period of time. Each sequential episode was designed to match one of the 80 household objects. The object images were taken from the Stark lab set of stimuli, freely available at (<http://faculty.sites.uci.edu/starklab/mnemonic-similarity-task-mst/>). The pictures of the episodic sequences did not actually contain the matching household object. Instead, the sequences were designed in a way that the matching object could fit in or make sense within the given sequence. In other words, the matching object could be expected to be encountered in the situation depicted in the episodic sequence. Each episodic sequence was designed to be congruent with its specific corresponding (congruent) object.

2.3. Experimental design

The experiment design of the two studies consisted of an encoding phase and a test phase. The encoding phase was the same in the two studies. Participants were presented in trial with an episodic sequence followed by a picture object. Participants were asked to rate the degree of congruence on a scale from 1 (i.e., does not fit in) to 4 (i.e., fits in very well) of the target picture object in relation to the context formed by the succession of the 4 preceding episodic sequence images (Fig. 1a). The encoding phase included a total of 80 trials, each consisting of an episodic sequence followed by an object. Two versions of the encoding phase were constructed so that 40 of the episodic sequences could be perceived as highly congruent with picture objects by the participant. This yielded a total of an *a priori* possible 40 high and 40 low congruent sequence-object pairings. However, we followed the empirical participant's ratings when assigning degree of object-episode congruency in our analysis. The order of the trials at encoding was randomised for each participant.

Each trial started with the appearance of a fixation cross on the screen for a random duration of 2000 to 4000 ms. Afterwards, an episodic sequence consisting of four photographs was presented. Each of the four photographs was presented on a white background for 2000 ms, one at a time in temporal order, separated by the presentation of a fixation cross for 500 ms. After the episodic sequence was presented, a

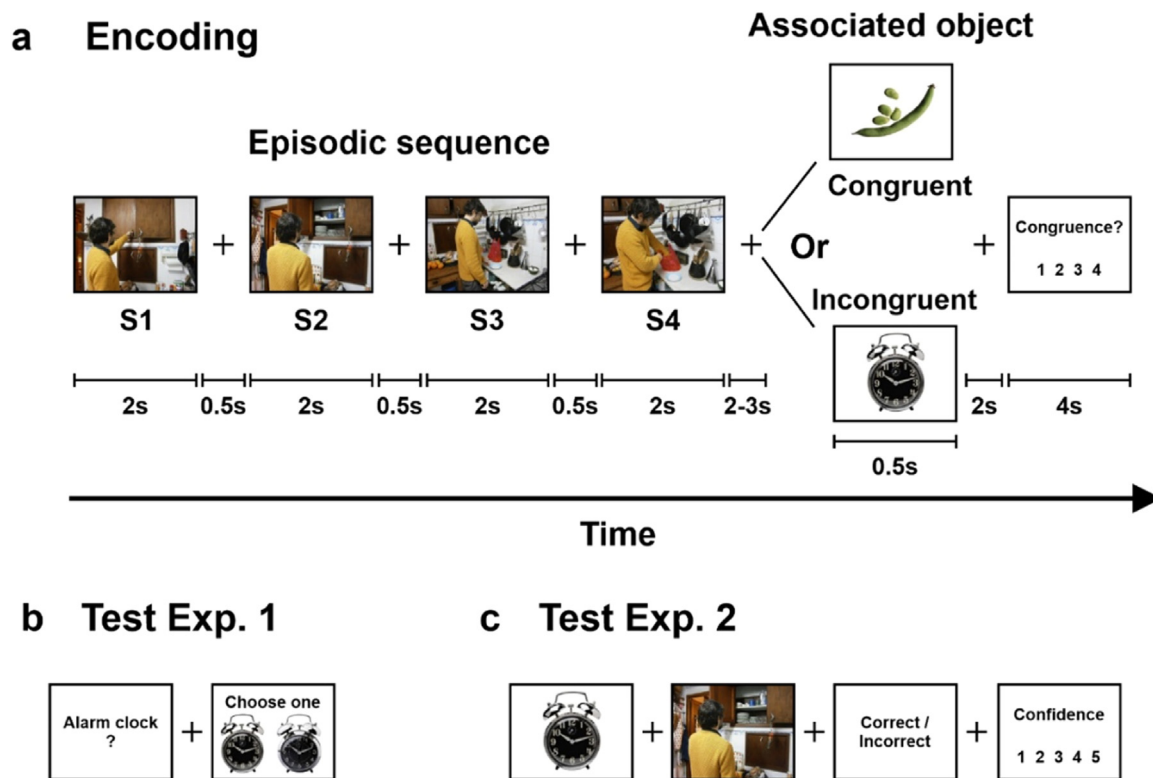


Fig. 1. Encoding and recognition memory task design for Experiments 1 and 2. (a) During the encoding phase for all experiments, participants encoded episodic sequences composed of 4 photographs depicting a routine domestic episode. These were followed by highly congruent or low congruent/incongruent object pictures. Participants indicated the degree of congruence between the episode and the object. (b) In Experiment 1, memory of object pictures was tested by the object word label followed by a true and a lure item of the same object. Participants had to indicate the correct picture presented during the encoding phase. (c) In Experiment 2, each object picture presented during the encoding phase was displayed together with one image from an episodic sequence. Participants were required to indicate whether the object and episodic image picture corresponded with the episodic + object picture presented in the same trial during the encoding phase.

fixation cross appeared on the screen for 2000 to 3000 ms, separating the episodic sequence from the presentation of the following object. The picture of the object was presented on a white background for 500 ms, followed by the appearance of a fixation cross for 2 s. Finally, a screen was presented with the word ‘Congruence?’ and the digits ‘1–2–3–4’ below, upon which participants had to indicate, within a maximum of 4 s, the degree of congruence between the object and the just-encoded episodic sequence by pressing 1, 2, 3 or 4 in the keyboard. Participants were previously instructed to respond thoughtfully as quickly and accurately as possible. As soon as they responded, a fixation cross was presented for 500 ms, and the next trial began. The encoding phase lasted around 30 min. After the 80 episodic sequences and objects were presented, the encoding phase was finished.

The encoding phase was followed by a ~10 min interference task consisting of choosing the correct answer to simple additions and subtractions that appeared on the screen. Participants were told to respond as quickly as possible, although no time limit was imposed. The distraction task ensured the participants would not rehearse the pictures they had previously seen.

2.4. Recognition memory test

After a break of ~10 min, a recognition memory test was presented unexpectedly to the participants in the two studies.

In Experiment 1, the test included 160 object words and 160 object picture pairs of each word. Eighty words and object pictures corresponded to previously presented objects in the encoding phase (‘Old’), whereas the other 80 were non-related (‘New’) objects (Fig. 1b). Each picture pair depicted the same object but with small changes in specific features between each other (e.g., orientation, colour, etc.). Each trial

began with a fixation cross lasting from 3 to 5 s at random. Subsequently, each word was presented for a maximum of 6 s with a question mark below ‘?’. Participants were instructed to press ‘1’ on the keyboard if they thought the word referred to an object presented during encoding (Old) or ‘2’ if not (New). If the participant responded ‘Old’ to a word, then two pictures of the object word were presented on the screen for a maximum of 6 s (Fig. 1b). Pictures were presented on the computer screen, with one item to the left and one to the right of fixation. The left/right assignment was randomly chosen in each trial. The two pictures from each pair were almost identical, but only one corresponded to the exact one presented in the encoding phase (true), whereas the other one served as a lure. For example, if the participant had seen the photograph of a hammer and later, during the test, correctly identified the word ‘hammer’ as one of the objects she saw, then two photographs of similar-looking hammers appeared. The participant was instructed to identify the picture object that was exactly like the one presented in the encoding phase by pressing 1 if it was the left photograph or 2 if it was the right one. The picture pair presented when a participant misclassified as ‘Old’ a new word was almost identical, though neither of the two pictures had been seen in the encoding phase. The order of the presentation of word + picture pairs in the test was randomised before each participant started the test. The recognition memory test lasted ~20 min.

In the test phase of Experiment 2, the 80 object pictures presented during the encoding phase were included in the test. In each trial, each object picture was presented together with the first or the second image from each of the episodic sequences presented in the encoding phase (Fig. 1c). Participants were requested to indicate whether the object picture and episodic image picture matched the episodic sequence and object picture presented during the encoding phase. Half of the

object+episode picture pairs presented in the test matched the encoding ones, whereas the others were randomly paired with each other. The total set of 80 picture pairs was constructed so that 40 old ones included 20 object+episode pairs encoded in the high-congruency condition and 20 in the low-congruency condition. The same distribution pattern was used to construct the set of object+episode new picture pairings (i.e., those that do not match the trials presented at encoding). Two different versions of 40 old and 40 new sets of picture pairings were created by controlling so that in one version, a picture object was paired with a matched image from the encoded episodic sequence (Old) and to an unmatched image (New) in the other. The two versions were assigned randomly to the participant's sample. Each trial began with a fixation cross lasting from 3 to 5 s at random. The object and the episodic sequence picture were then presented on the screen. Participants were instructed to answer whether both photographs had been presented together in the same trial during the encoding phase (by pressing '1' on the keyboard) or not (by pressing '2'). Participants were asked to rate their confidence in their previous response from 1 ('no confidence') to 5 ('absolute confidence') with the same numbers on the keyboard. Object and episodic pictures were presented on the computer screen, with one item to the left and one to the right of fixation. The left/right assignment was randomly chosen on each trial.

2.5. Behavioural data analyses

Paired Student *t*-test was used to compare participants' performance (measured in percentage) between conditions. Repeated measures ANOVA was used to statistically assess differences between participants' performance when they included more than two variables. Statistical significance threshold was set at $p < 0.05$.

2.6. EEG recording and preprocessing

In study 1, EEG was recorded during the task using a 32-channel system at a sampling rate of 500 Hz, with an online band-pass filter set at 0.01–100 Hz, using a BrainAmp amplifier and tin electrodes mounted in an electrocap (Electro-Cap International) located at 29 standard positions (Fp1/2, Fz, F7/8, F3/4, Fcz, Fc1/2, Fc5/6, Cz, C3/4, T3/4, Cp1/2, Cp5/6, Pz, P3/4, T5/6, P01/2, Oz) and at the left and right mastoids. An electrode placed at the lateral outer canthus of the right eye served as an online reference. EEG was re-referenced offline to the average of all channels. Vertical eye movements were monitored with an electrode at the infraorbital ridge of the right eye. Electrode impedances were kept below 5 k Ω . EEG was low-pass filtered offline at 30 Hz. We applied the Parks-McClellan Notch filter using the toolbox ERPLAB (<http://erpinfo.org/erplab>).

For each participant, we extracted the EEG epochs for each encoding image. Epochs had a duration of 2000 ms for images from the episodic sequence and 2500 ms for images of the picture object and were baseline corrected to the pre-stimulus interval (–100 to 0 ms). Epochs with maximum absolute amplitude over 100 μ V were discarded for further analysis. For later analysis, all the epochs were smoothed by averaging data via a moving window of 100 ms (excluding the baseline period) and then downsampled by a factor of 5.

2.7. Event-Related potentials (ERP) analysis

ERPs elicited by object picture were studied by extracting a trial-by-trial event-locked EEG epochs of 2600 ms starting at 100 ms before the presentation of the image. A baseline from –100 ms to stimulus onset was used. ERP trials with amplifier saturation, or trials with a shift exceeding 100 μ V/s were automatically rejected offline.

2.8. Representational similarity analysis (RSA)

To account for whether the ongoing encoding of an episodic sequence of pictures elicited a gradual increase in stable brain activity

patterns, we implemented a temporally resolved RSA using Pearson correlation coefficients, which are insensitive to the absolute amplitude and variance of the EEG response. The correlation analysis on EEG data was made at the individual level and to each time point separately and included spatial (i.e., scalp voltages from all the 29 electrodes) features of the resulting EEG single trials. Cross-correlation outputs were converted to Fisher's *z* scores for statistical analysis. To examine how a schema-consistent sequence's unfolding modulated the stabilisation of activity patterns, we correlated, for each participant and at single trial level, the EEG patterns of activity elicited by the 1st and the 2nd, the 2nd and the 3rd and the 3rd and the 4th pictures of each episodic sequence.

To examine whether neural stability was disrupted by objects perceived as incongruent with the just encoded episodic sequence, compared to congruent objects, we correlated, for each participant and at single trial level, the EEG patterns of activity elicited by the 4th pictures of each episodic sequence and the EEG patterns elicited by the subsequent associated picture object.

We implemented a cluster-based permutation test to account for neural stability differences between picture order within the episodic sequence and between High and Low congruence conditions (Maris and Oostenveld, 2007). It identifies clusters of significant points in the resulting 1D matrix in a data-driven manner and addresses the multiple-comparison problem by using a nonparametric statistical method based on cluster-level randomisation testing to control the family-wise error rate. Statistics were computed for every time point, and the time points whose statistical values were larger than a threshold ($p < 0.05$, two-tail) were selected and clustered into connected sets based on adjacency points in the 1D matrix. The observed cluster-level statistics were calculated by taking the sum of the statistical values within a cluster. Condition labels were then permuted 1000 times to simulate the null hypothesis, and the maximum cluster statistic was chosen to construct a distribution of the cluster-level statistics under the null hypothesis. The nonparametric statistical test was obtained by calculating the proportion of randomised test statistics that exceeded the observed cluster-level statistics.

RSA was also conducted between the EEG signal of each encoding image (i.e., the image at 1st, 2nd, 3rd, 4th position in a sequence) with the EEG signal of the corresponding object encoding period (i.e., the matching object with the sequence and the followed fixation cross). We termed the output of this analysis as neural reactivation, instead of neural stability, because the resulting similarity values may not be directly explained by a direct timepoint-to-timepoint correspondence between the each of the neural patterns, akin to possible temporal compression of memory reactivation seen in previous rodent (e.g., Foster and Wilson, 2006) and human (e.g., Kurth-Nelson et al., 2016) studies. RSA was performed timepoint-to-timepoint at trial level and upon spatial features (i.e., scalp voltages from all the 29 electrodes) (Silva et al., 2019; Wu et al., 2022). Point-to-point Pearson correlation values were calculated, resulting in a 2D similarity matrix with the size of 200×250 (each time point represents 10 ms, given the 500 Hz EEG recording sampling rate and a down-sampling factor of 5). The *x*-axis of the matrix represented the object and the followed fixation cross time points, and the *y*-axis represented the time points of image encoding. The output 2D matrix represents the overall degree of neural pattern similarity between EEG elicited for each pair of encoding images and its corresponding object encoding period.

We also explored the neural similarity between the EEG signal of each encoding image (i.e., the image at 1st, 2nd, 3rd, 4th position in a sequence) with the EEG signal of the offset period immediately following the last image of the sequence. To do so, we extracted EEG signal of the first 2 s of event offset ERPs. After smoothing and down-sampling each epoch following the same parameters used in the previous analysis, we then performed the RSA timepoint-to-timepoint at trial level between each encoding image with its corresponding event offset. This returned a 2D similarity matrix with a size of 200×200 , where the *x*-axis of the matrix represented the event offset time points, and the *y*-axis

represented the time points of image encoding. The output 2D matrix represents the overall degree of neural pattern similarity between EEG elicited for each pair of encoding images and its corresponding event offset period.

2.9. Linear-mixed effect model

To further explore in the trial base how the neural pattern was associated with the subjective feeling of the congruence and its impact on subsequent memory of the object image, we applied a Linear Mixed-Effect Model (LMM) on data from the neural stability analysis, on ERPs elicited by the picture object and during the offset period after that picture, and on EEG RSA data between each of the images of the sequence and the immediately following offset period preceding the object picture (i.e., sequence offset) and between each of the images of the sequence and the encoding of the object picture (i.e., object encoding).

To complement the cluster-based statistical approach implemented to assess for statistical significance of modulations of neural similarity described in the previous section, we constructed a point-to-point LMM with neural stability value as the dependent variable. To evaluate the extent to which neural stability fluctuated throughout the encoding of the episodic sequence, we introduced in the model picture pair order in the sequence as a fixed-effect variable: 1st and 2nd pictures, 2nd and 3rd pictures and 3rd and 4th pictures. Subject was included in the model as the grouping variable, with random intercept and a fixed slope for the fixed-effect variable. To evaluate whether the degree of perceived congruence of the object picture with the encoded episode changed the ongoing EEG activity state of stability, 2 fixed-effect variables were introduced into the model: the rating of the object's congruence to the sequences (on a scale of 1 to 4) and the memory for the object (whether the word and the image were correctly recognised during the test). We z -transformed the neural stability value and congruence rating score across all trials. The subject was then introduced into the model as the grouping variable, with random intercept and a fixed slope for each fixed-effect variable.

For ERPs elicited by the encoding of the object picture, we first grouped 29 electrodes into 6 scalp regions. To obtain more stable spatial patterns, border electrodes between regions were included in each neighbouring region (Lu et al., 2015; Sols et al., 2017). As a result, the 6 regions were defined as follows: region 1 (FP1, Fz, F3, F7, FCz, FC1, FC5); region 2 (FP2, Fz, F4, FCz, FC2, FC6); region 3 (FCz, FC1, FC5, Cz, C3, T3, Cp1, Cp5); region 4 (FCz, FC2, FC6, Cz, C4, T4, CP2, CP6); region 5 (CP1, CP5, Pz, P3, T5, PO1, Oz); region 6 (CP2, CP6, Pz, P4, T6, PO2, Oz). The LMM was constructed separately for each region. ERPs voltages at each offset time point were averaged across electrodes within the region and then introduced in our LMM as the dependent variable. The rating of the object's congruence to the sequences (on a scale of 1 to 4) and the memory for the object (whether the word and the image were correctly recognised during the test) were included in the model as fixed-effect variables. We z -transformed the ERPs voltage value and congruence rating score across all trials. The subject was then introduced into the model as the grouping variable, with random intercept and a fixed slope for each fixed-effect variable. The statistical significance was adjusted by Bonferroni correction for each fixed-effect variable taking into account both the number of regions and the number of time points, resulting in a corrected alpha level of $\alpha = 3.33 \times 10^{-5}$ (0.05/(6 × 250)).

We also applied LMM on output from RSA. We identified on the resulting 2D similarity matrix the time point of encoding and offset where the pattern similarity reached the peak value. We then averaged at a single-trial level the similarity value across ± 50 ms (11 data points) around the peak encoding time point for each offset time point, resulting in a 1D similarity value with the length of 250 time points, covering the whole offset period. For each time point, we constructed the LMM with similarity value as the dependent variable. Then 3 fixed-effect variables were introduced into the model: the image position in the sequence (1st, 2nd, 3rd, and 4th), congruence rating and memory for the

object. We z -transformed the similarity value, congruence rating score and the image position number across all trials. Subject was included in the model as the grouping variable, with random intercept and a fixed slope for each fixed-effect variable. The statistical significance for each fixed-effect variable was Bonferroni corrected with a thresholded alpha level of $\alpha = 2 \times 10^{-4}$ (0.05/250).

3. Results

3.1. Behavioural results from experiment 1

In Experiment 1, the proportion of trials rated as lowly congruent ('1' or '2') was 45.17% (SD = 8.59%) and 15.43% (SD = 6.42%), respectively, and highly congruent ('3' or '4'), respectively, was 13.45% (SD = 6.36%) and 25.39% (SD = 7.47%) with the episode.

In general, participants were accurate in recognizing words referring to picture objects learnt during the encoding phase (Mean = 71.09%, SD = 12.80%; above chance: $t_{(28)} = 8.41$, $p < 0.001$). They showed greater accuracy for words related to picture objects encoded with high (Mean = 76.07%, SD = 13.27%) than low (Mean = 66.11%, SD = 13.89%) congruency to the episodic context during the encoding phase (paired t -test: $t_{(28)} = 5.89$, $p < 0.001$) (Fig. 2a). For those correctly recognised words, in the next picture recognition test, participants were accurate in correctly identifying the exact object picture presented in the encoding phase (Mean = 73.13%, SD = 8.02%; above chance: $t_{(28)} = 15.53$, $p < 0.001$). They performed similarly for pictures rated as high (Mean = 74.61%, SD = 11.63%) and low (Mean = 72.61%, SD = 8.61%) congruency to the episodic context (paired t -test: $t_{(28)} = 0.39$, $p = 0.878$) (Fig. 2b).

To specify the subsequent memory strength for objects, for all later analyses, we classified the memory performance of a trial based on whether the object image was correctly recognised during the test. Therefore, successful recognition of both gist word and object image was considered as a 'remembered item' condition, while either failing in recognition of gist word or image was considered as a 'forgotten item' condition. This separation also rendered a balanced percentage of trials across conditions, with 53.53% (SD = 11.43%) for 'remembered' condition and 46.47% (SD = 11.43%) for 'forgotten' condition (Wilcoxon signed-rank test: $z = 1.74$, $p = 0.08$).

3.2. Behavioural results from experiment 2

In Experiment 2, we sought to examine whether the retrieval benefits of encoding congruence observed in Experiments 1 were also accompanied by an increased ability of the participants to bind the object to the preceding episodic context. Thus, in Experiment 2, participants were required to judge whether pairs of pictures, including one of the encoded objects and one picture from the sequence episodes, were seen together or not during encoding. As in Experiment 1, the proportion of trials rated as highly ('4') and lowly congruent ('1') were 35.39% (SD = 12.63%) and 37.67% (SD = 10.06%), respectively, and the average proportion of intermediate levels of congruency was much lower ('2': Mean = 12.92%, SD = 9.97%; and '3': Mean = 10.81%, SD = 5.68%). A comparison of the proportion of trials rated as lowly congruent ('1') and highly congruent ('4') revealed they were not statistically different ($t_{(17)} = -0.85$, $p = 0.41$). Therefore, we analysed participants' hits for items classified as '1' or '4' at encoding. The results of this study revealed that participants were accurate in correctly identifying the encoded correspondence between objects and the episodic context (Mean = 0.79, SD = 0.12), but that the accuracy was higher (paired student t -test, $t_{(17)} = 5.59$, $p < 0.001$) and more confident ($t_{(17)} = 6.78$, $p < 0.001$) when the association between the object and the context at encoding was high rather than low congruent (Fig. 2c).

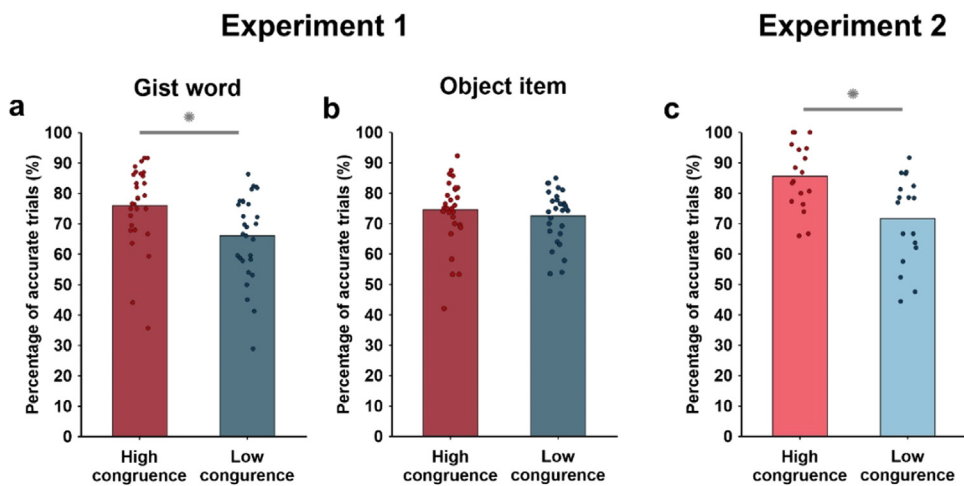


Fig. 2. Behavioural results of Experiments 1 and 2. (a) Participants' test accuracy on recognising the gist word separated by whether the picture object was perceived by the participant as highly or lowly congruent with the episodic sequence during encoding in Experiment 1. (b) Participants' test accuracy on identifying the exact object picture presented in the encoding phase after correctly recognising the 'Old' gist word. The accuracy was separated by whether the picture object was perceived by the participant as highly or lowly congruent with the episodic sequence during encoding in Experiment 1. (c) Participants' percentage of hits in the recognition memory test of Experiment 2. In all plots, dots represent values for an individual subject. * indicates $p < 0.05$.

3.3. Neural stability and episodic congruence

We next examined whether EEG states of stability changed throughout the encoding of a schema-consistent images of the narrative sequence. We hypothesised that schema consistency for picture sequences would promote neural stability, and that this would be observed as an increase in across-item EEG pattern similarity of the EEG patterns elicited throughout the picture sequence. To address this issue, we calculated a time-resolved neural similarity analysis between EEG patterns elicited by picture pairs (1st and 2nd, 2nd and 3rd and 3rd and 4th) across the sequences. Confirming our hypothesis, we found that the 3rd and the 4th pictures elicited more stable patterns of EEG activity than the 1st and the 2nd pictures from 110 to 290 ms (Cluster statistics: $p = 0.04$, corrected for multiple comparisons, sum T -values = 64.75), from 970 to 1350 ms (corrected $p = 0.009$, sum T -values = 119.38) and from 1520 to 2000 ms (corrected $p = 0.007$, sum T -values = 145.28) from stimulus onset. We also found that the 3rd and the 4th pictures showed more stable EEG patterns than the 2nd and the 3rd pictures from 1550 to 1800 ms (corrected $p = 0.03$, sum T -values = 68.11) and that the 2nd and the 3rd picture elicited EEG stability did not differ statistically with the 1st and the 2nd picture ($p > 0.05$) (Fig. 3a). A similar pattern of result was found by the LMM analysis, which confirmed that neural stability increased as the pictures in the sequence increased. The results of the LMM analysis revealed that the neural similarity increases were shown at an early from 90 to 390 ms and at a later time window from 780 to 2000 ms from stimulus onset ($p < 0.05$, uncorrected) (Fig. 3b).

Next, we evaluated whether EEG states of stability were associated with the encoding of object pictures that were highly or lowly congruent with the preceding episode. We hypothesised that congruency induced higher states of stability and that this elevated state of neural stability would be reflected as an increase in neural similarity upon object encoding, rendering them more similar than objects perceived as lowly congruent with the preceding episodic context. To address this issue, we compared the cross-temporal correlation analysis of EEG activity patterns between the last picture of the episodic sequence (i.e., 4th picture) and the picture object for trials in which the object was perceived as high and low congruent. Please note that the assignment of episode-object associations was counterbalanced between participants and, therefore, an object encoded highly congruently with an episode in one participant could be encoded lowly congruently by another participant. The results of this analysis showed that highly congruent items elicited more stable patterns of EEG activity than lowly congruent items with the preceding picture of the episodic sequence from 980 ms to 1220 ms from picture onset (corrected $p = 0.04$, sum T -values = 66.91)

(Fig. 3c). A similar pattern of result was found by the LMM analysis, which confirmed that neural stability was lower for object pictures that were perceived lowly than highly congruent with the preceding episodic sequence from 980 to 1660 ms from stimulus onset ($p < 0.05$, uncorrected) (Fig. 3d). The LMM results revealed however that an opposite pattern of results emerged at early stages of picture object encoding, from 150 to 260 ms from picture onset, whereby neural stability was higher for lowly than highly congruent objects ($p < 0.05$, uncorrected) (Fig. 3d). We found similar opposite relationship between neural stability and memory for the picture object at early stages of encoding, from 190 to 290 ms and from 660 to 680 ms from picture onset ($p < 0.05$, uncorrected) (Fig. 3d). These results suggest that context-dependant neural state of stability may be modulated by episodic congruency at encoding.

3.4. Neural stability in congruent episodes (evidence from Wu et al., 2022)

To further scrutinize for the possibility that the increase in neural stability during the encoding of pictures was due to elements that were consistent with an activated schema, we analysed scalp EEG data from Wu et al. (2022) study. The study involved two groups of participants: one group was asked to encode picture sequences that depicted schema-congruent episodes ($N = 25$, Experiment 1), while the other group was asked to encode schema-unrelated picture series of similar length ($N = 27$, Experiment 2). Each series consisted of 6 pictures, presented once to participants. In Experiment 1, the sequences of pictures depicted day-to-day routine circumstances that were congruent with a schema, such as shopping or cooking. In Experiment 2, we randomized the pictures across sequences from Experiment 1, resulting in sequences that no longer portrayed meaningful or schema-congruent episodes (Fig. 4a). We analysed the scalp EEG data using the same RSA approach as in the main study and implemented a point-to-point LMM analysis to assess for modulation of neural stability as a function of picture order. We found that only the schema-congruent group exhibited a gradual increase in neural stability between pictures in the sequence (Fig. 4b-e). We then directly compared changes in neural stability between groups as a function of picture order with a repeated measures ANOVA and confirmed differences in neural stability modulation between groups (sequence order \times group interaction: $F_{(4,200)} = 4.04$; $p = 0.004$; main factor picture order: $F_{(4,200)} = 6.59$; $p < 0.001$; main factor group: $F_{(1,50)} = 0.10$; $p = 0.75$), from RSA averaged on the time window found in the findings in the main study (i.e., 780 to 2000 ms from stimulus onset, Fig. 3b).

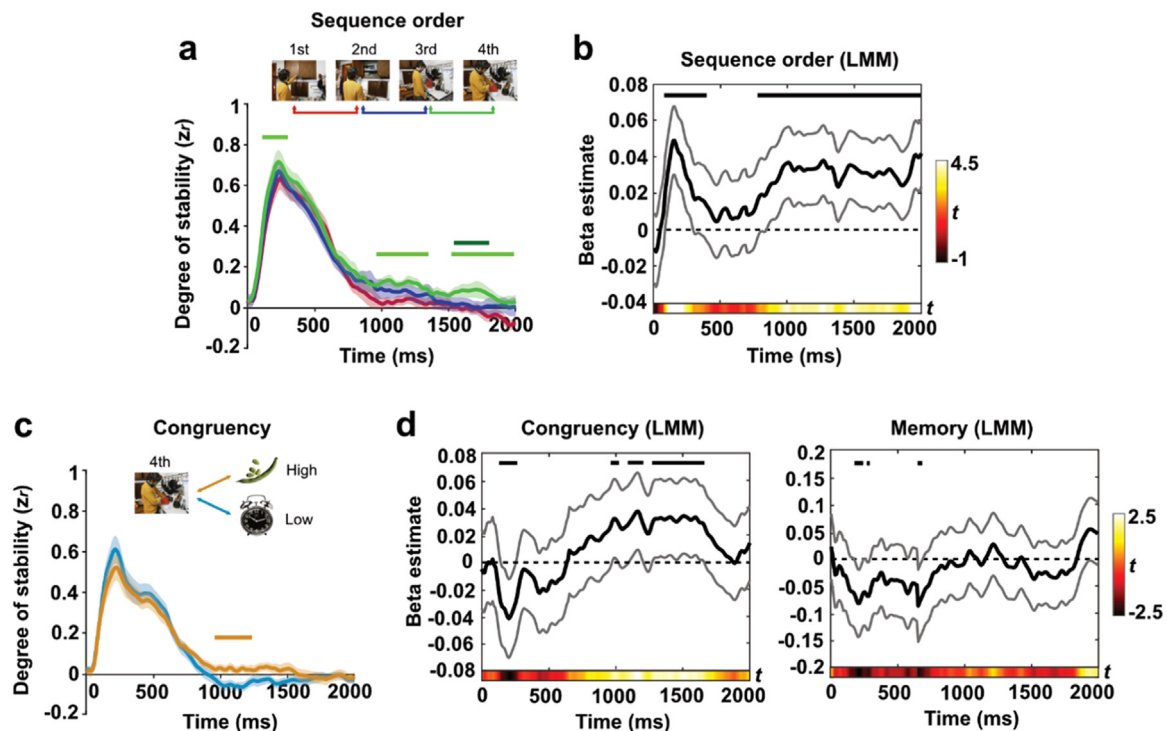


Fig. 3. Neural stability and congruence. (a) Point-to-point participants' average degree of neural stability between the 1st and the 2nd, the 2nd and the 3rd and the 3rd and 4th picture from the episodic sequence. Thick green lines depict the statistically significant ($p < 0.05$, cluster-based permutation test) time periods of increased neural similarity by the 3rd and 4th compared to the 2nd and 3rd (thick dark green) and to the 1st and 2nd (thick light green) pictures. (b) Beta estimates (thick black line), confidence intervals (thick grey lines) and associated t values resulting from the LMM analysis on the neural stability data for pairs of pictures in the episodic sequence. (c) Point-to-point participants' average degree of neural stability between the 4th picture of the episodic sequence and the associated subsequent object picture separately when participants rated it as highly or lowly congruent with the just encoded episode. (d) Beta estimates (thick black line), confidence intervals (thick grey lines) and associated t values resulting from the LMM analysis on the neural stability between the 4th picture of the episodic sequence and the object picture. Thick black lines in b and d depict statistically significant points ($p < 0.05$, uncorrected).

3.5. The N400 signals low congruent items from encoding episodes

We hypothesised that the disruption of neural stability elicited by the detection of low-congruence objects from the unfolding schema-consistent episodic narrative would be preceded by a prediction error signal in the brain. We aimed to identify such an error signal in the EEG by means of a transient increase in the N400 ERP component, which has been widely related to incongruence detection in the literature (Kutas and Federmeier, 2011). To assess this issue in our data, we grouped the 29 electrodes into 6 regions and averaged the epochs across electrodes within each region. We then introduced the voltage value into the LMM as the dependent variable and included participants' ability to correctly recognise the picture image at test congruence rating provided at encoding as the main fixed-effect variables. Subject was included as the grouping variable. This analysis was conducted for each time point and each scalp region separately.

The results of the LMM analysis revealed significant effects at specific scalp regions and temporal points for both memory and congruence. For memory, later forgotten objects elicited significantly more negative ERPs amplitude at 520 ms to 870 ms from picture onset distributed over the frontal scalp region (Bonferroni corrected $p < 0.05$, sum T -values = 171.42) (Fig. 5a). For congruence, we found more substantial negative amplitude with objects rated as less congruent with the preceding episodic context. This effect was significant from 410 ms to 730 ms from picture onset and was distributed over frontocentral scalp regions (Bonferroni corrected $p < 0.05$, sum T -values = 252.20) (Fig. 5b).

3.6. Neural reinstatement induced by low contextual congruency

We next sought to test that prediction error elicited by low-congruence items would be accompanied by more robust reactivation of the just-encoded episodic elements. To address this issue, we implemented a trial-based and temporally resolved neural similarity analysis between EEG data elicited at picture object and EEG data elicited by each image of the preceding episodic context. The results of this approach revealed an increase in neural similarity between EEG patterns from ~100 ms - 700 ms at object picture onset and EEG patterns of activity between ~100 ms - 750 ms from the onset of pictures within the episodic sequence (Fig. 6a).

To assess whether the degree of neural reactivation observed during the encoding of picture objects was modulated by their perceived congruency with the encoding episodic context and their later memory accessibility at test, we ran an LMM analysis taking the neural similarity as dependent variable and participant congruence rating and recognition hit rate at test as fixed-effect factors. We also included a variable accounting for the order of pictures in the episodic sequence to control for the possibility that any effect could be accounted for by the temporal proximity of the encoding pictures of the preceding episode with the picture object. We first identified on the similarity matrix the exact time point where the similarity value reached the peak across participants (i.e., at 240 ms from episodic sequence picture onset and at 220 ms from object picture onset) (Fig. 6a). We then extracted the averaged similarity value ± 50 ms around the peak over the encoding time for each time at object encoding period. This resulted in a 1D vector with a length of 250, where each time point represents 10 ms during object

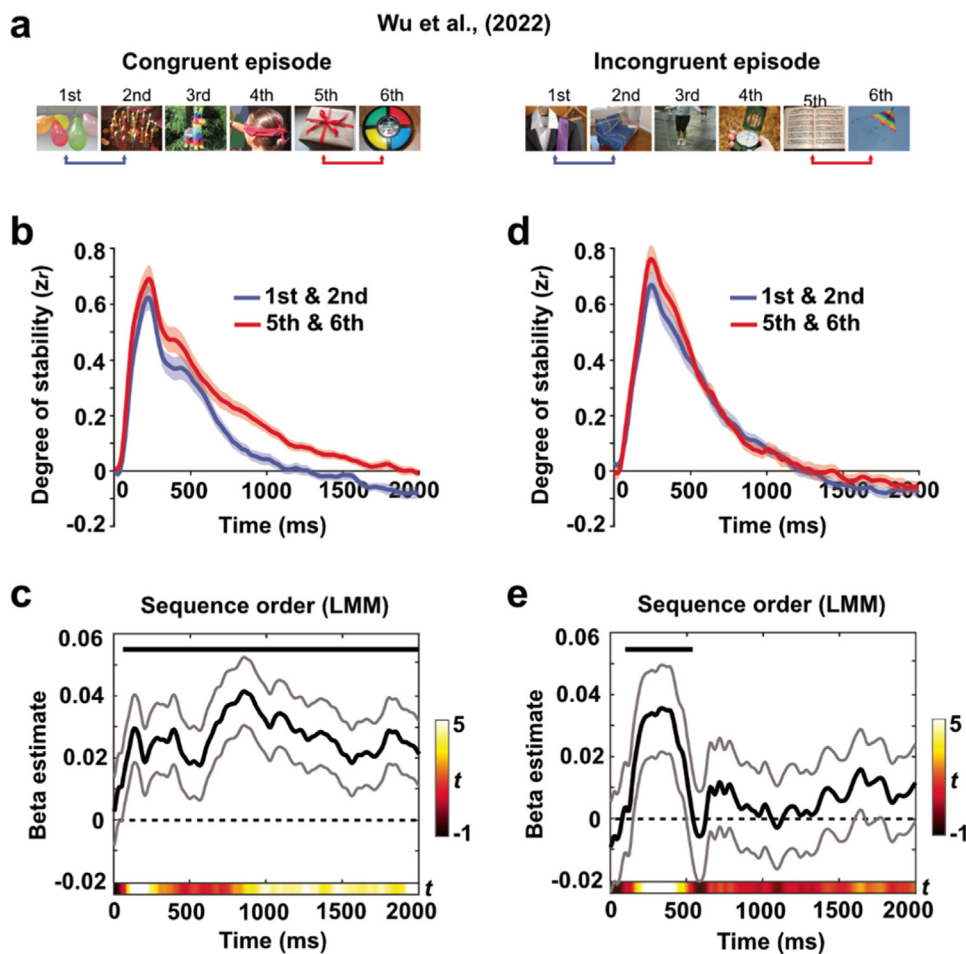


Fig. 4. (a) We implemented RSA to scalp EEG data acquired in Wu et al., 2022. In this experiment, two groups of different healthy participants were asked to encode different series of 6 unique picture sets. In one group of participants (Experiment 1), each picture series depicted a sequence of instances that depicted the unfolding of a realistic and congruent life-like episode, for example a birthday party. In the other group of participants (Experiment 2), pictures series from the first group were scrambled and presented in a way that series were meaningless or incongruent between them. (b) Point-to-point participants' average degree of neural stability between first and last pairs of pictures from episodic sequence in EEG data from group of participants that were asked to encode picture sequences depicting a schema-congruent episode in Wu et al. (2022) study. (c) Beta estimates (thick black line), confidence intervals (thick grey lines) and associated t values resulting from the LMM analysis on the neural stability data for pairs of pictures from EEG data in Experiment 1. (d) Point-to-point participants' average degree of neural stability between first and last pairs of pictures from episodic sequence in EEG data from group of participants that were asked to encode picture sequences depicting a schema-incongruent episode in Wu et al. (2022) study. (e) Beta estimates (thick black line), confidence intervals (thick grey lines) and associated t values resulting from the LMM analysis on the neural stability data for pairs of pictures from EEG data in Experiment 2. Thick grey lines in c and e depict statistically significant points ($p < 0.05$, uncorrected).

encoding. LMM was conducted at each time point with all trials included. We found that the degree of neural reactivation correlated negatively with the participants' ability to correctly recognise an item at test. The predictive negative relationship between neural reactivation and memory started to be significant at 270 ms from picture-object onset and remained significant throughout almost the entire epoch until 2100 ms (Fig. 6b, upper). In addition, we also found a negative correlation between neural similarity and congruence rating. However, the significant effects were more transient but comparably distributed along with the object-picture encoding epoch. Notably, the first significant time point was very early, at 190 ms from picture onset, which preceded the relationship effects between neural similarity and memory (Fig. 6b, lower). A more persistent negative correlation was also found later, between 490 ms - 720 ms and at 1340 ms from picture onset. Finally, no significant point exceeded the statistical significance threshold for the variable sequence order, indicating that the significant relationship between neural similarity and memory and congruence was not driven by specific neural similarity measures between picture objects and pictures of the episode.

We examined the possibility that neural similarity could also be seen immediately after the encoding of the sequence, during the offset period preceding the onset of object picture. To address this issue, we computed the RSA between EEG activity elicited by each of the images from each of the episodic sequences with the corresponding EEG activity from the 2000 ms offset period that followed the last image of the sequence. The result of this analysis showed, however, very little, if any, increase in similarity at the offset period after the episodic sequence (Fig. 6c). In addition, the LMM analysis revealed that the degree of neural similarity at the episodic sequence offset period did not statistically vary as a

function of sequence order (Fig. 6d), thereby suggesting that neural similarity at this offset period was not directly associated to the encoding of the preceding episodic event information.

We finally explored whether the increase in RSA between the EEG patterns elicited by the picture sequence and the object picture was trial-specific. To do so, we shuffled the corresponding assignments between the EEG data from the episodic sequence and the object, separately for the High and Low congruency conditions, for each participant. For example, during each shuffling, an episodic sequence from the High congruency condition was paired with an object picture randomly selected from another trial in the High congruency condition. The RSA was then computed on a trial level between each encoding image in the sequence (i.e., the image at the 1st, 2nd, 3rd, and 4th positions in a sequence) and their new random picture epoch. For the resulting 2D similarity matrix, we averaged the similarity value ± 50 ms around the identified peak (i.e., 190 ms - 290 ms, see also Fig. 6a) across the object encoding period. Separately for the High and Low congruency condition, we repeated this procedure 200 permutation times, each with a randomly generated shuffling index. We then averaged the results across permutations and compared them to the true value yielded from the original analysis. No significant difference (all $p > 0.05$, cluster-based permutation test) in the RSA value across the object encoding period was found between the true and shuffled trials for neither the High nor Low conditions after being corrected by the cluster-based permutation test.

4. Discussion

In this study, we tested whether the perceived incongruence of an item from an unfolding event and its impact on memory relied on the

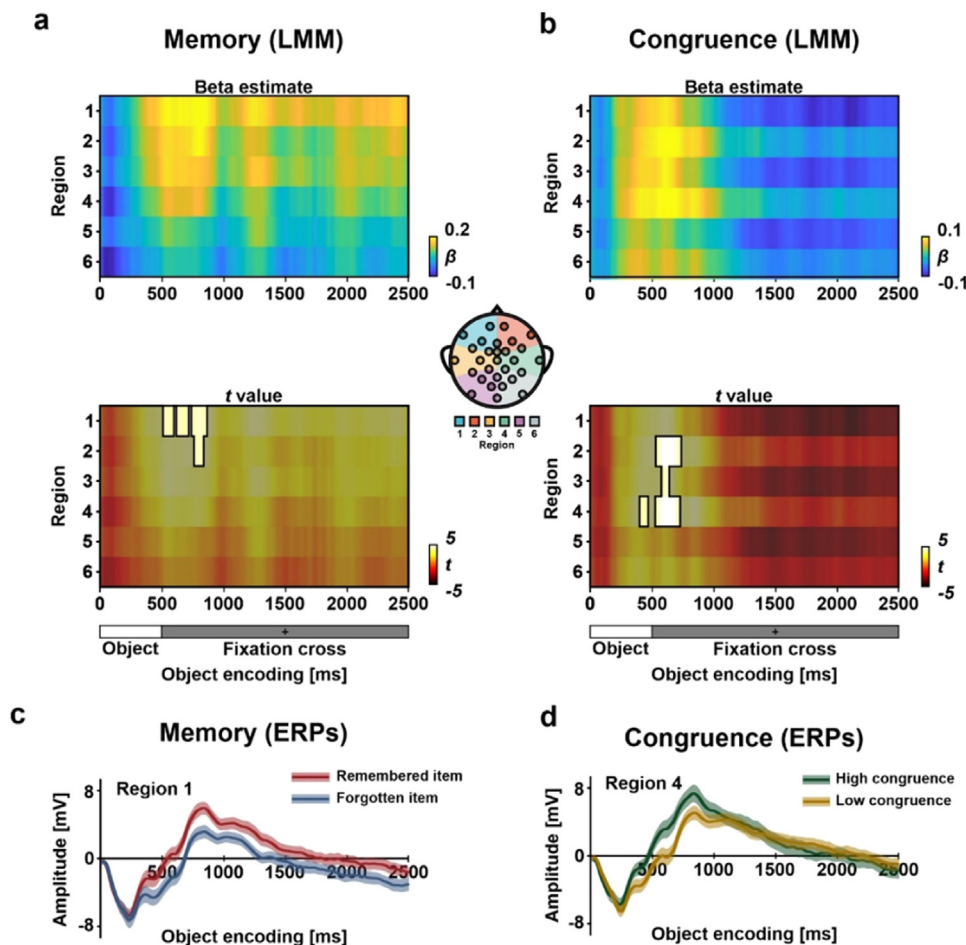


Fig. 5. LMM on ERPs elicited at object encoding. 29 electrodes were grouped into 6 scalp regions, and boundary electrodes were included in either neighbouring region to obtain more stable spatial patterns. (a) Beta estimates and t -value resulting from the LMM analysis at each region and time point for the fixed-effect Memory. (b) Beta estimates and t -value resulting from the LMM analysis at each region and time point for the fixed-effect Congruence. The unmasked area on statistical map marks the corresponding region and time where the t -statistics exceed the significance threshold ($p < 0.05$) with the alpha level adjusted with Bonferroni correction. (c) Participants' averaged ERPs from representative region 1. (d) Participants' averaged ERPs from representative region 4 (note that for visual illustration, trials with congruence ratings of 1 and 2 were grouped and averaged as Low congruence and trials with congruence ratings of 3 and 4 were grouped as High congruence). In (c) and (d), the shaded area represents standard error (SEM) across subjects.

disruption of neural stability patterns preceded by the neural reactivation of the memory representations of the just-encoded event. Our findings, derived from combining behavioural data from 2 different experiments and the implementation of multivariate pattern analysis on EEG signal during encoding of one of them, confirmed our hypothesis by showing that neural stability patterns gradually increase throughout the ongoing exposure to a schema-consistent episodic narrative, and that the brain stability pattern is interrupted when the encoding of an object of the event is lowly congruent within the ongoing schema representation. We found that the decrease in neural stability for low-congruence items was best seen at ~ 1000 ms from object-encoding onset when compared to high-congruence items and that this effect was preceded by an enhanced N400 ERP and an increased degree of neural reactivation of the just-encoded episode for low-congruence items observed between ~ 200 to 1000 ms from picture onset. Current findings provide new insights into the brain mechanisms and their temporal dynamics supporting the encoding of schema-consistent episodic narratives as it unfolds and their involvement upon perceived incongruities.

Central in our findings is that the degree of neural reactivation of the encoded episode by the final target object correlated negatively with the perceived congruence and the participant's ability to later recognise the picture in a memory test. The notion that memory reactivation benefits memory formation is well established in previous research. Most of it showed that the reactivation strength drives long-term memory formation by mimicking neural replay phenomena thought to promote rapid consolidation processes seen in rodent studies (Carr et al., 2011). Other studies have revealed that when novel encoding inputs reactivate previously encoded information that overlaps in content, the long-term memory representations of the two event contents become integrated,

promoting generalisation and adaptive behaviour (Schlichting and Preston, 2015; Shohamy and Wagner, 2008). Despite the notion that memory reactivation may potentially benefit memory formation, another set of findings described the opposite effect. These studies found that memory reactivation of overlapping memories may yield competition between the two, resulting in interference with a negative impact on memory (Kuhl et al., 2011; see for recent reviews: Brunec et al., 2020; Ritvo et al., 2019). In line with this view, we found a negative correlation between object-picture participants' hits and the degree of elicited memory reactivation of the preceding event, suggesting an interference between the encoding of an incongruent item and the simultaneous memory reinstatement of the just encoded event.

Our findings that surprising or unexpected elements of the unfolding experience elicited the rapid reinstatement of the just-encoded picture sequence are in line with previous findings that showed that sudden shifts in an ongoing episodic context (i.e., event boundaries) induce the rapid reactivation of preceding episodic information (Silva et al., 2019; Sols et al., 2017). Event boundaries are thought to represent moments in time whereby a continuous stream of incoming information is segmented into different memory units (Zacks et al., 2011). In this model, the process of event segmentation starts when our current understanding of the world is destabilised by a new observation that does not fit our current expectations. Viewed from this perspective, high levels of surprise (defined here as a high degree of inconsistency of the object picture to the schema or internal model representation activated during the preceding episodic sequence) refer to a substantial change in our understanding of the current inputs from experience. This engenders additional resources to re-evaluate the current internal model in the face of the new observation, which may benefit from the greater reactivation

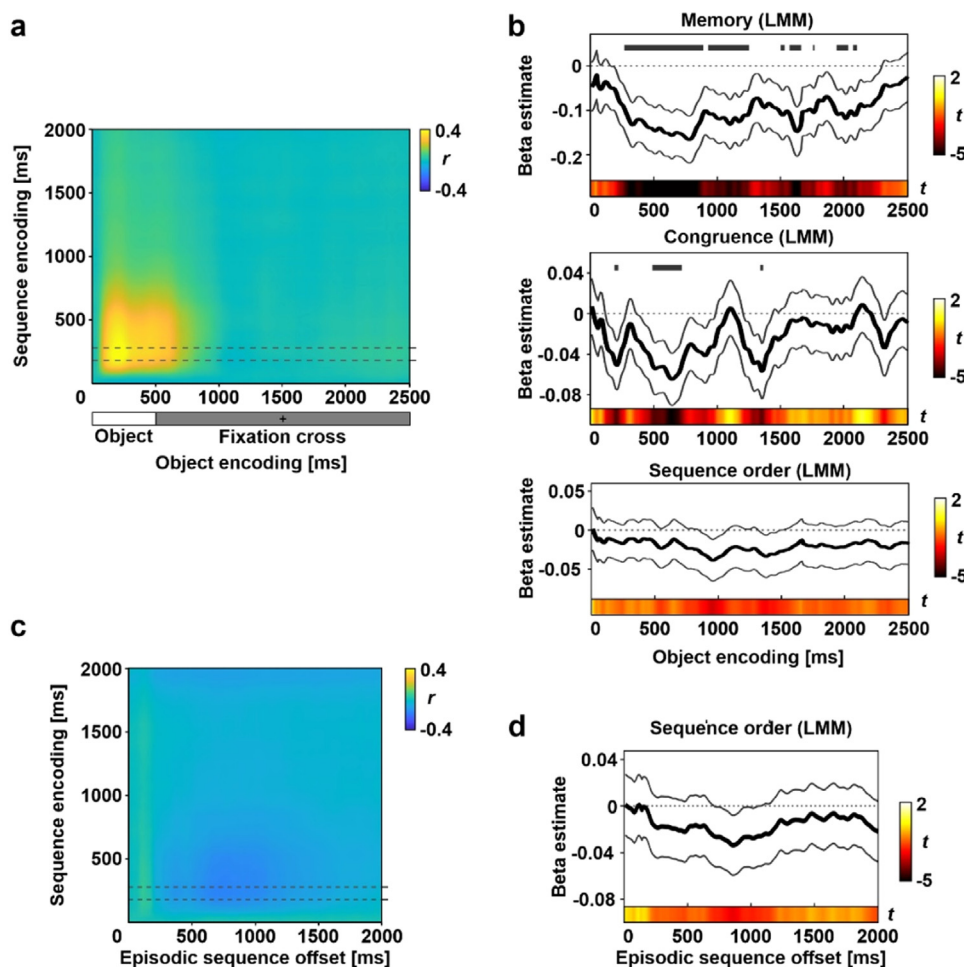


Fig. 6. Neural similarity results. (a) Time-resolved degree of neural similarity between EEG patterns elicited by picture images from the episodic sequence and EEG patterns elicited during the encoding of the associated picture object. Grey dashed line shows the ± 50 ms time window where neural similarity reached the peak (190 ms – 290 ms at encoding time). (b) Beta estimates (thick black line), confidence intervals (grey lines) and associated t values for fixed-effect Memory (upper) Congruence (middle) and Sequence order (lower) resulting from the LMM analysis on averaged similarity value over the peak (190 ms – 290 ms at encoding time) along object encoding period. (c) Time-resolved degree of neural similarity between EEG patterns elicited by picture images from the episodic sequence and EEG patterns elicited during episodic event offset period. (d) Beta estimates (thick black line), confidence intervals (thick grey lines) and associated t values for fixed-effect Sequence order resulting from the LMM analysis on averaged similarity value around the peak (190 ms – 290 ms at encoding time) along the episodic sequence offset period. Thick grey lines on top of (b) marked the time window exceeding the significance threshold ($p < 0.05$) with the alpha level adjusted with Bonferroni correction.

of the memory representations to resolve it. One possible direction for future research is to investigate whether the mechanisms behind neural reactivation induced at event boundaries and during the detection of unexpected events, as observed in the current study, are similar. In this study, we showed that detecting inconsistencies in an ongoing event triggers a rapid reactivation of recently encoded event information, followed by a decrease in neural stability between the encoded episode and the incongruent event. We speculate that the degree of neural reinstatement may reflect the need to re-evaluate ongoing event content, while subsequent changes in neural stability may reflect representational updates that influence how events are ultimately segmented and organized in long-term memory. If this hypothesis is correct, then detecting surprising elements may induce neural reactivation of unfolding event memories, but only changes in neural stability may occur when the surprising element leads to the sectioning off the preceding elements as a separate event.

A limitation of the study is that participants were required to evaluate the coherence between a swiftly displayed picture object and the preceding episodic sequence, which may have increased the strain on their working memory and rekindled the previous sequence, thus constraining the applicability of the outcome of the study to real-life situations. Another limitation of the current study is that increased reactivation of event representations upon the detection of an incongruent element is not trial-specific, thereby indicating that neural reactivation may reflect gist-based or schematic memory representations. This may be explained because the encoding of picture sequences in our study relied heavily on the narrative aspect of each stimulus, which emphasized the gist-like information rather than the perceptual features of each item.

Furthermore, our experimental design included episodic events where the same actor appeared in all episodes, and each depicted different narrative but shared very similar contextual perspectives, such as occurring in the same house layouts, such as kitchen, living room, and bedroom. The high degree of contextual overlap in the encoded material may have made it challenging to identify trial-unique memory reinstatement between episodes and picture objects, complicating the sensitivity of our analysis. Additionally, already encoded event episodes may induce interference with subsequent encoding trials, making it challenging to differentiate between trials. For example, [Koen & Rugg \(2016\)](#) found that using an AB/AC source memory interference paradigm, the reactivation of neural activity features shared across numerous events in the same category was higher for subsequently remembered AB encoding trials. However, pattern similarity analysis only revealed item-level reactivation of AB trials during the AC encoding task when the strength of interference across events was low.

We found that neural activity shifted into a stable state as the number of pictures consistent with an unfolding schema increased and that this took place rapidly, within the first 2000 ms from picture onset. These findings complemented previous fMRI results that showed activity invariance during the encoding of different, but schema consistent episodes ([Baldassano et al., 2018](#)). Leveraging on the fine-grained temporal properties of the EEG signal, the current results, however, extend previous fMRI findings by showing that the shifts towards neural states of stability may be seen at two different time windows within the first 2000 ms from cue onset. First, at an early stage, within the first 250 ms from picture onset, and at a later one, after ~ 1000 ms from picture onset. We interpret these two temporal windows of neural stability as

reflecting different processes. The early time window may be influenced by early perceptual processing, including visual and category properties of the input stimuli. Indeed, temporally resolved decoding approaches in humans have shown that the discrimination of visual images and their classification into categories occurred within the first 250 ms after from their presentation onset (Cichy et al., 2014). In the context of the current study, pictures presented within each episodic sequence shared similar schematic representation, which included visual properties of the stimuli itself (such as the spatial layout) and the stimulus category. For example, images from Fig. 1a, depict the sequence of activities from a person in the same kitchen. This visual image correspondence between pictures from an episodic sequence may gradually promote the gradual facilitation of visual discrimination and categorization as the episode unfolds, which may, in turn, lead to the increased neural state of stability at early temporal windows throughout episodic sequence encoding in our data.

We also found that shifts towards neural states of stability can be seen in a later time window, at ~1000 ms from picture onset, and that neural stability gradually increased as the number of schema consistent picture images were presented and that it was disrupted upon the presentation of a picture object perceived as lowly congruent with the preceding episode. While the exact nature of this later state of neural stability fluctuation is uncertain, our finding that it is disrupted by the detection of a schema incongruent element lends support to ERP literature that found late ERP amplitude modulations (after ~900 ms from stimulus onset) reflecting the results of an expectancy match/mismatch-related processing between the incoming information and the preceding activated schema information (e.g., Hölte et al., 2019). Additionally, while extensive electrophysiological evidence showed that mnemonic processes can take place rapidly in humans, i.e., within the first 1500 ms from a reminder onset (e.g., Rugg and Curran, 2007; Yonelinas, 2002), a recent proposal suggested that visual cues can render the activation of mnemonic associated information via a cascade progression (Staresina and Wimber, 2019). According to this proposal, first, information would enter the hippocampus in a feedforward fashion allowing a fast old/new computation between the current and the available information in memory. This would then be followed by pattern completion mechanisms at approximately 500 ms after a cue onset, which would trigger a sustained activation of an associated memory trace between 500 and 1500 ms. We interpret, thus, our findings in line with the possibility that late stages of neural stability reflected an end point mnemonic processing accounted by several preceding cascade stages of processing, being some of them reflected by the amplitude modulation of the N400 ERP (i.e., reflecting early stages of mismatch detection of an unexpected element from the expected schema-consistent episode) and reinstatement of the just encoded episodic memory information.

In summary, the current study offers three new findings. It shows that the detection of low-congruence elements of an episodic experience elicited a rapid memory reactivation of the just-encoded event information; that this is concomitant to a mnemonic prediction error signal during encoding; and that the result of this computation leverages the disruption of stable patterns of neural activity elicited during the schema-consistent episodic event. These findings inform us about the rapid but sequential structure of the distinct neural mechanisms supporting the detection of incongruencies during encoding and their consequences on memory. We speculate that these same processes may take place in realistic scenarios of our everyday experience.

Credit author statement

P.A.P., N.B. and L.F. designed the experiment. P.A.P. conducted the experiment. X.W., J.G.A. and L.F. analysed the data. X.W. and L.F. wrote the original draft. L.F. provided supervision. L.F. acquired funding.

Declaration of Competing Interest

The authors declare no competing financial interests.

Data availability

Data and MATLAB code implementing all quantitative analyses in this study is available at <https://osf.io/jf8nq/>.

Acknowledgments

We thank Luz Bavassi for their helpful discussions on earlier versions of the manuscript. This work was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades, which is part of Agencia Estatal de Investigación (AEI), through the project PID2019-111199GB-I00 (Co-funded by European Regional Development Fund. ERDF, a way to build Europe), to L.F. We thank CERCA Programme/Generalitat de Catalunya for institutional support.

References

- Baldassano, C., Chen, J., Zadbood, A., Pillow, J.W., Hasson, U., Norman, K.A., 2017. Discovering event structure in continuous narrative perception and memory. *Neuron* 95 (3), 709–721. doi:10.1016/j.neuron.2017.06.041, e5.
- Baldassano, C., Hasson, U., Norman, K.A., 2018. Representation of real-world event schemas during narrative perception. *J. Neurosci.* 38 (45), 9689–9699. doi:10.1523/JNEUROSCI.0251-18.2018.
- Bein, O., Duncan, K., Davachi, L., 2020. Mnemonic prediction errors bias hippocampal states. *Nat. Commun.* 11 (1), 3451. doi:10.1038/s41467-020-17287-1.
- Brune, I.K., Robin, J., Olsen, R.K., Moscovitch, M., Barense, M.D., 2020. Integration and differentiation of hippocampal memory traces. *Neurosci. Biobehav. Rev.* 118, 196–208. doi:10.1016/j.neubiorev.2020.07.024.
- Carr, M.F., Jadhav, S.P., Frank, L.M., 2011. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nat. Neurosci.* 14 (2), 147–153. doi:10.1038/nn.2732.
- Chen, J., Cook, P.A., Wagner, A.D., 2015. Prediction strength modulates responses in human area CA1 to sequence violations. *J. Neurophysiol.* 114 (2), 1227–1238. doi:10.1152/jn.00149.2015.
- Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. *Nat. Neurosci.* 17 (3), 455–462.
- Foster, D.J., Wilson, M.A., 2006. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440 (7084), 680–683. doi:10.1038/nature04587.
- Frank, D., Montaldi, D., Wittmann, B., Talmi, D., 2018. Beneficial and detrimental effects of schema incongruence on memory for contextual events. *Learn. Mem.* 25 (8), 352–360. doi:10.1101/lm.047738.118.
- Frank, D., Montemurro, M.A., Montaldi, D., 2020. Pattern separation underpins expectation-modulated memory. *J. Neurosci.* 40 (17), 3455–3464. doi:10.1523/JNEUROSCI.2047-19.2020.
- Gershman, S.J., Radulescu, A., Norman, K.A., Niv, Y., 2014. Statistical computations underlying the dynamics of memory updating. *PLoS Comput. Biol.* 10 (11), e1003939. doi:10.1371/journal.pcbi.1003939.
- Gilboa, A., Marlatte, H., 2017. Neurobiology of schemas and schema-mediated memory. *Trends Cogn. Sci. (Regul. Ed.)* 21 (8), 618–631. doi:10.1016/j.tics.2017.04.013.
- Greve, A., Cooper, E., Kaula, A., Anderson, M.C., Henson, R., 2017. Does prediction error drive one-shot declarative learning? *J. Mem. Lang.* 94, 149–165. doi:10.1016/j.jml.2016.11.001.
- Greve, A., Cooper, E., Tibon, R., Henson, R.N., 2019. Knowledge is power: prior knowledge aids memory for both congruent and incongruent events, but in different ways. *J. Exp. Psychol. Gen.* 148 (2), 325–341. doi:10.1037/xge0000498.
- Hölte, G., Lubahn, B., Mecklinger, A., 2019. The congruent, the incongruent, and the unexpected: event-related potentials unveil the processes involved in schematic encoding. *Neuropsychologia* 131, 285–293. doi:10.1016/j.neuropsychologia.2019.05.013.
- Koen, J.D., Rugg, M.D., 2016. Memory reactivation predicts resistance to retroactive interference: evidence from multivariate classification and pattern similarity analyses. *J. Neurosci.* 36 (15), 4389–4399. doi:10.1523/JNEUROSCI.4099-15.2016.
- Kuhl, B.A., Rissman, J., Chun, M.M., Wagner, A.D., 2011. Fidelity of neural reactivation reveals competition between memories. *Proc. Natl. Acad. Sci. U.S.A.* 108 (14), 5903–5908. doi:10.1073/pnas.1016939108.
- Kurth-Nelson, Z., Economides, M., Dolan, R.J., Dayan, P., 2016. Fast sequences of non-spatial state representations in humans. *Neuron* 91 (1), 194–204. doi:10.1016/j.neuron.2016.05.028.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi:10.1146/annurev.psych.093008.131123.
- Lu, Y., Wang, C., Chen, C., Xue, G., 2015. Spatiotemporal neural pattern similarity supports episodic memory. *Curr. Biol.: CB* 25 (6), 780–785. doi:10.1016/j.cub.2015.01.055.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190. doi:10.1016/j.jneumeth.2007.03.024.

- McClelland, J.L., McNaughton, B.L., Lampinen, A.K., 2020. Integration of new information in memory: new insights from a complementary learning systems perspective. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 375 (1799), 20190637. doi:10.1098/rstb.2019.0637.
- Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., Mendelsohn, A., 2018. Knowledge acquisition is governed by striatal prediction errors. *Nat. Commun.* 9 (1), 1673. doi:10.1038/s41467-018-03992-5.
- Quent, J.A., Henson, R.N., Greve, A., 2021. A predictive account of how novelty influences declarative memory. *Neurobiol. Learn. Mem.* 179, 107382. doi:10.1016/j.nlm.2021.107382.
- Ritvo, V.J.H., Turk-Browne, N.B., Norman, K.A., 2019. Nonmonotonic plasticity: how memory retrieval drives learning. *Trends Cogn. Sci. (Regul. Ed.)* 23 (9), 726–742. doi:10.1016/j.tics.2019.06.007.
- Rouhani, N., Norman, K.A., Niv, Y., 2018. Dissociable effects of surprising rewards on learning and memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 44 (9), 1430–1443. doi:10.1037/xlm0000518.
- Rugg, M.D., Curran, T., 2007. Event-related potentials and recognition memory. *Trends Cogn. Sci. (Regul. Ed.)* 11 (6), 251–257. doi:10.1016/j.tics.2007.04.004.
- Schlichting, M.L., Preston, A.R., 2015. Memory integration: neural mechanisms and implications for behavior. *Curr. Opin. Behav. Sci.* 1, 1–8. doi:10.1016/j.cobeha.2014.07.005.
- Shohamy, D., Wagner, A.D., 2008. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* 60 (2), 378–389. doi:10.1016/j.neuron.2008.09.023.
- Silva, M., Baldassano, C., Fuentemilla, L., 2019. Rapid memory reactivation at movie event boundaries promotes episodic encoding. *J. Neurosci.* 39 (43), 8538–8548. doi:10.1523/JNEUROSCI.0360-19.2019.
- Sinclair, A.H., Manalili, G.M., Brunec, I.K., Adcock, R.A., Barense, M.D., 2021. Prediction errors disrupt hippocampal representations and update episodic memories. *Proc. Natl. Acad. Sci. U.S.A.* 118 (51), e2117625118. doi:10.1073/pnas.2117625118.
- Sols, I., DuBrow, S., Davachi, L., Fuentemilla, L., 2017. Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Curr. Biol.* 27 (22), 3499–3504. doi:10.1016/j.cub.2017.09.057, e4.
- Staresina, B.P., Wimber, M., 2019. A neural chronometry of memory recall. *Trends Cogn. Sci. (Regul. Ed.)* 23 (12), 1071–1085. doi:10.1016/j.tics.2019.09.011.
- Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M.P., Morris, R.G.M., 2007. Schemas and memory consolidation. *Science* 316 (5821), 76–82. doi:10.1126/science.1135935.
- Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bitto, H., Morris, R.G.M., 2011. Schema-dependent gene activation and memory encoding in neocortex. *Science* 333 (6044), 891–895. doi:10.1126/science.1205274.
- van Kesteren, M.T.R., Rijpkema, M., Ruiters, D.J., Morris, R.G.M., Fernández, G., 2014. Building on prior knowledge: schema-dependent encoding processes relate to academic performance. *J. Cogn. Neurosci.* 26 (10), 2250–2261. doi:10.1162/jocn_a_00630.
- Wu, X., Viñals, X., Ben-Yakov, A., Staresina, B.P., Fuentemilla, L., 2022. Post-encoding reactivation is related to learning of episodes in humans. *J. Cogn. Neurosci.* 35 (1), 74–89. doi:10.1162/jocn_a_01934.
- Yonelinas, A.P., 2002. The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46 (3), 441–517. doi:10.1006/jmla.2002.2864.
- Zacks, J.M., Kurby, C.A., Eisenberg, M.L., Haroutunian, N., 2011. Prediction error associated with the perceptual segmentation of naturalistic events. *J. Cogn. Neurosci.* 23 (12), 4057–4066. doi:10.1162/jocn_a_00078.