nature human behaviour

Article

https://doi.org/10.1038/s41562-025-02315-1

Emotional arousal enhances narrative memories through functional integration of large-scale brain networks

Received: 20 March 2025

Accepted: 5 September 2025

Published online: 13 October 2025

Check for updates

Jadyn S. Park[1,6] ✉, Kruthi Gollapudi[1], Jin Ke[2], Matthias Nau[3], Ioannis Pappas[4] & Yuan Chang Leong[1,5,6] ✉

Emotional events tend to be vividly remembered. While growing evidence suggests that emotions have their basis in brain-wide network interactions, it is unclear whether and how these whole-brain dynamics contribute to memory encoding. Here we combined functional MRI, graph theory, text analyses and pupillometry in a naturalistic context where participants recalled complex narratives in their own words. Across three independent datasets, emotionally arousing moments during narrative perception were associated with an integrated brain state characterized by increased cohesion across functional modules, which in turn predicted the fidelity of subsequent recall. Network integration mediated the influence of emotional arousal on recall fidelity, with consistent within- and between-network interactions supporting the mediation across datasets. Together, these results suggest that emotional arousal enhances memory encoding via strengthening functional integration across brain networks. Our findings advance a cross-level understanding of emotional memories that bridges large-scale brain network dynamics, affective states and ongoing cognition.

Events that induce emotional arousal are often better remembered than those that are emotionally neutral[1–4]. For instance, a person is more likely to recall witnessing a car accident than what they saw on a routine drive home. Enhanced memory for arousing events prioritizes the recall of experiences that are motivationally important, such as those relevant to potential threats and rewards[5–7]. This prioritization facilitates the efficient allocation of cognitive resources towards events that impact well-being, enabling individuals to respond more rapidly and effectively to similar events in the future[8,9].

Emotional arousal is commonly defined as a subjective state of heightened alertness and activation[10–12], often accompanied by physiological changes such as increased pupil dilation and electrodermal activity[13,14]. Prevailing theories propose that emotional arousal triggers the release of norepinephrine in the amygdala, which interacts with other memory-related regions, most prominently the hippocampus, to enhance the encoding of emotional experiences[15–18]. Consistent with this account, human functional MRI (fMRI) studies have found that enhanced memory of emotional stimuli is associated with amygdalar and hippocampal activation[19–23], as well as increased amygdalar–hippocampal connectivity[24,25]. These findings have been recently corroborated in intracranial recordings in patients[26,27]. The administration of the β-adrenergic antagonist propranolol, which interferes with norepinephrine binding, reduces amygdalar and hippocampal responses to emotional stimuli and diminishes the associated memory enhancement, suggesting a causal role of arousal-dependent norepinephrine release in mediating these effects[28–31].

[1]Department of Psychology, University of Chicago, Chicago, IL, USA. [2]Department of Psychology, Yale University, New Haven, CT, USA. [3]Institute for Brain and Behavior Amsterdam, Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [4]Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA. [5]Neuroscience Institute, University of Chicago, Chicago, IL, USA. [6]Institute of Mind and Biology, University of Chicago, Chicago, IL, USA. ✉e-mail: jadynpark@uchicago.edu; ycleong@uchicago.edu

https://doi.org/10.1038/s41562-025-02315-1

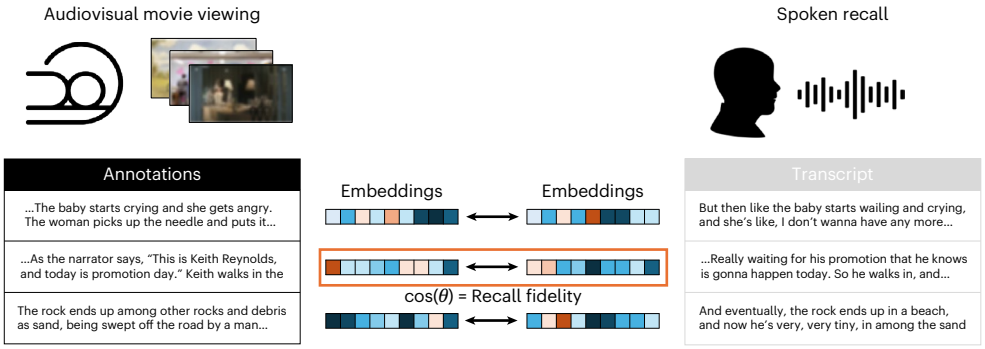


Fig. 1 | Schematic of task and semantic similarity analysis between movie annotations and spoken recall. Participants watched audiovisual clips in the fMRI scanner and were instructed to recall what they had watched from memory. Descriptions of movie events obtained from independent coders and transcriptions of participants' verbal recall were converted to vector embeddings. Recall fidelity was computed as the cosine similarity between the vector of scene annotations from an event and the transcript from the matching event in a participant's recall. The video screen captures are blurred for copyright reasons.

Current models of arousal-dependent memory enhancement have made foundational contributions to our understanding of the brain regions involved, but they generally do not consider the role of the large-scale functional network organization of the brain (but see ref. 32). There is growing recognition that the dynamic reconfiguration of distributed functional brain networks in response to task demands is crucial for adaptive behaviour[33–38] and underlies emotional responses and affective states[12,39–42]. Converging evidence suggests that heightened arousal promotes brain network integration—a brain state characterized by increased connectivity and cohesion across functional brain systems[33,43–48], which have in turn been associated with memory encoding and retrieval[49,50]. Network integration is thought to be supported by the release of norepinephrine, as demonstrated by pharmacological manipulations[46,48], chemogenetic stimulation[45] and pupil dilation[33,43], a widely used proxy of norepinephrine activity[51]. Building on this work, we propose that functional network integration facilitates the enhanced encoding of emotionally arousing memories, above and beyond the contributions of the amygdala and the hippocampus.

While earlier behavioural and pharmacological studies on emotional memories used narratives as stimuli[1,2,31], neuroimaging studies have primarily focused on the encoding and recall of static stimuli such as emotionally arousing images[19,20,24,25] or words[22,23]. Consequently, less is known about the relationship between neural activity and the emotional enhancement of narrative memories. Unlike static stimuli, narratives represent complex experiences where interconnected events unfold dynamically, requiring the accumulation and updating of information over time. Advances in neuroimaging techniques and natural language processing now allow for the study of how the brain encodes and recalls these naturalistic episodes[52]. Here we take advantage of these advances to investigate the neural basis underlying how emotional arousal enhances what is remembered from temporally extended narratives.

We used two publicly available fMRI datasets where participants watched dynamic audiovisual narratives and verbally recalled what they remembered. Arousal ratings of the movies were obtained by analysing annotations of them using large language models (LLMs) and from subjective ratings collected in separate behavioural experiments. Applying graph theoretic analyses to the fMRI data, we computed a dynamic measure of functional integration during narrative perception. We assessed the fidelity of participants' memory by using text embedding models to measure the similarity between their recall and detailed descriptions of the corresponding event. We then tested whether functional integration during encoding mediated the effects of emotional arousal on subsequent recall fidelity. Finally, we replicated our results in a third dataset where we measured arousal using pupillometry while participants listened to and recalled a suspenseful audio story, providing converging evidence of our results with a physiological measure of arousal. Altogether, our work advances an integrative view of arousal-dependent memory enhancement of narrative memories that bridges affective states and ongoing cognition through brain network dynamics.

Results

Our first set of analyses used two publicly available fMRI datasets that, to our knowledge, are the only open datasets combining naturalistic narrative stimuli with free spoken recall. These data allow for fine-grained, event-level assessment of memory fidelity within participants. In the first dataset (Film Festival)[53], participants watched ten audiovisual video clips (range, 2 min 12s–7 min 45 s; average duration, 4 min 32 s; total duration, 45 min 22 s). In the second dataset (Sherlock)[54], participants watched a 50-minute segment from an episode of the British TV show *Sherlock*. In both datasets, the participants were then asked to describe what they remembered from the movies in as much detail as they could. Each movie was divided into a set of events defined by major shifts in the narrative, including changes in topic, location, time, and characters (68 events in Film Festival; average length, 38.4 s; s.d., 18.2 s; 48 events in Sherlock; average length, 57.5 s; s.d., 41.76 s).

We first obtained detailed annotations of the movie events as a ground-truth description of what occurred in each event. To assess how well participants recalled the movie events, we quantified the semantic similarity between participants' recall transcripts and these annotations. We used Google's Universal Sentence Encoder (USE)[55] to convert the annotation and recall transcripts of each event into numerical vectors in a high-dimensional embedding space (Methods). We then calculated the cosine similarity between the vector encoding the annotation and those encoding a participant's recall of an event as a measure of recall fidelity[56–58] (Fig. 1). A high recall fidelity score indicated that a participant's recall closely matched the annotation of an event. Extended Data Fig. 1 displays an example event in each of the two datasets with the transcripts and fidelity scores of the recall from two different participants.

Functional integration during movie viewing is associated with higher memory fidelity

We first examined how dynamic changes in brain network organization during movie viewing were associated with subsequent memory fidelity. To that end, we parcellated the brain into 216 cortical and subcortical regions. For each event in each participant, we computed the functional connectivity between each pair of brain regions as the Fisher *z*-transformed Pearson correlation between the

https://doi.org/10.1038/s41562-025-02315-1

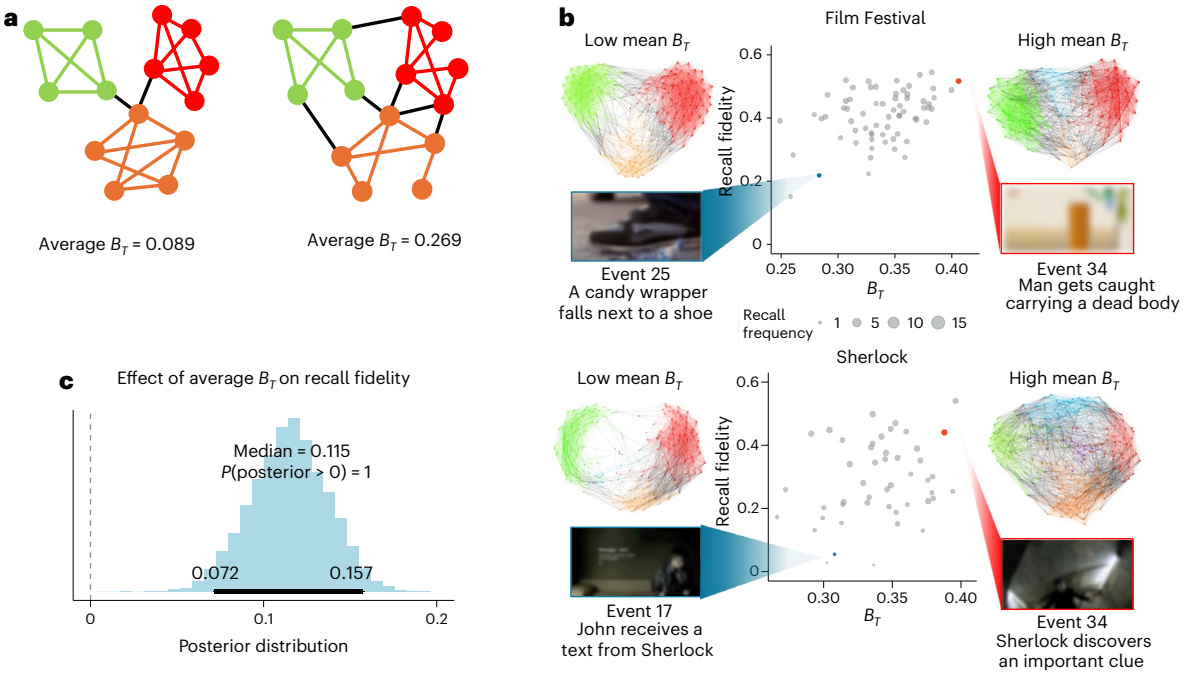


Fig. 2 | Functional integration during movie viewing is associated with higher recall fidelity. a, Graphical representation of average B_T. Each community is shown in a different colour. Both networks contain the same number of nodes and edges, but the network on the right has more connections between communities, resulting in a higher average B_T. **b**, Scatter plots depicting the relationship between average B_T, average recall fidelity and recall frequency. Each data point is an event. The size of each circle depicts the number of participants that recalled that event. Red and blue circles correspond to example events with high (red) or low (blue) average B_T. The connectivity plots depict functional connections for the respective events, with colours denoting module assignment. Each connectivity plot contains the same number of connections, but there are more intermodular connections during events when average B_T is high. The video screen captures are blurred for copyright reasons. **c**, Posterior distribution of the regression coefficient when predicting recall fidelity from average B_T, estimated by a Bayesian multilevel model that pooled across Film Festival and Sherlock. The 95% HDI of the distribution is indicated by the bold horizontal line.

blood-oxygenation-level-dependent (BOLD) time courses. The connectivity matrices were thresholded to retain the top 15% of connections and binarized to create sparse graphs, where nodes represent brain regions and edges represent functional connections between regions. This approach allowed us to leverage graph theory to quantify network properties of large-scale neural dynamics and investigate how functional modules in the brain interact with one another during movie watching.

In each graph, brain regions were assigned to communities such that regions within a community were more strongly connected to one another than to regions in other communities[59], thereby partitioning the brain into different functional modules (Methods). We then quantified the diversity of intermodular connections of each brain region at temporal segment T by calculating its participation coefficient (B_T)[60]. A brain region with a high B_T has connections distributed across multiple modules, suggesting that it may play a role in integrating information across modules[61]. Conversely, a brain region with a low B_T has most of its connections within its own module, suggesting more localized processing. High average B_T across brain regions would then reflect high levels of intermodule connectivity, indicating an integrated brain state when there is stronger cohesion across different functional modules.

For each event, we computed the average B_T across brain regions as a measure of whole-brain functional network integration[33]. We note that each graph was thresholded to retain a fixed number of connections, and thus all graphs had the same number of connections. An increase in B_T would thus not be due to a greater number of connections. Instead, graphs with high average B_T had a greater number of connections between functional modules, indicating greater integration across networks (Fig. 2a). We hypothesized that events associated with greater functional network integration during movie viewing are more strongly encoded and are thus remembered with higher fidelity.

We used Bayesian multilevel models to test how functional network integration during movie viewing was related to subsequent event memory. To increase statistical power and estimate reliability, we pooled data across the Film Festival and Sherlock datasets, treating dataset as a random effect to account for between-dataset variability. Consistent with our hypothesis, average B_T when viewing an event was positively associated with recall fidelity ($\beta = 0.11$; 95% highest density interval (HDI), [0.07, 0.16]; $P(\beta > 0) = 1$; Widely Applicable Information Criterion (WAIC) = 4,975; Fig. 2b,c). Model comparison showed that this model outperformed a null model without average B_T ($WAIC_{null} = 4{,}996$; Bayes factor (BF_{10}) = 7,739), indicating strong evidence for the contribution of network integration to recall fidelity.

Functional integration mediates the effects of emotional arousal on narrative recall

Given the well-established role of emotional arousal in enhancing memory encoding[4,17,18], we next sought to examine the impact of emotional arousal on functional integration and narrative recall. To this aim, we obtained ratings of emotional arousal based on text descriptions of each narrative event using an open-access LLM[62] (Methods). Additionally, we collected emotional arousal ratings from 30 participants for comparison and validation. Our approach parallels those of normed affective datasets such as the Affective Norms for English Words[63] and the International Affective Picture System[64], which use average ratings of arousal and valence across participants to characterize the emotional content of words or images. Similarly, we treated both LLM- and behaviour-derived ratings as normative, event-level estimates of emotional arousal of the narrative content.

https://doi.org/10.1038/s41562-025-02315-1

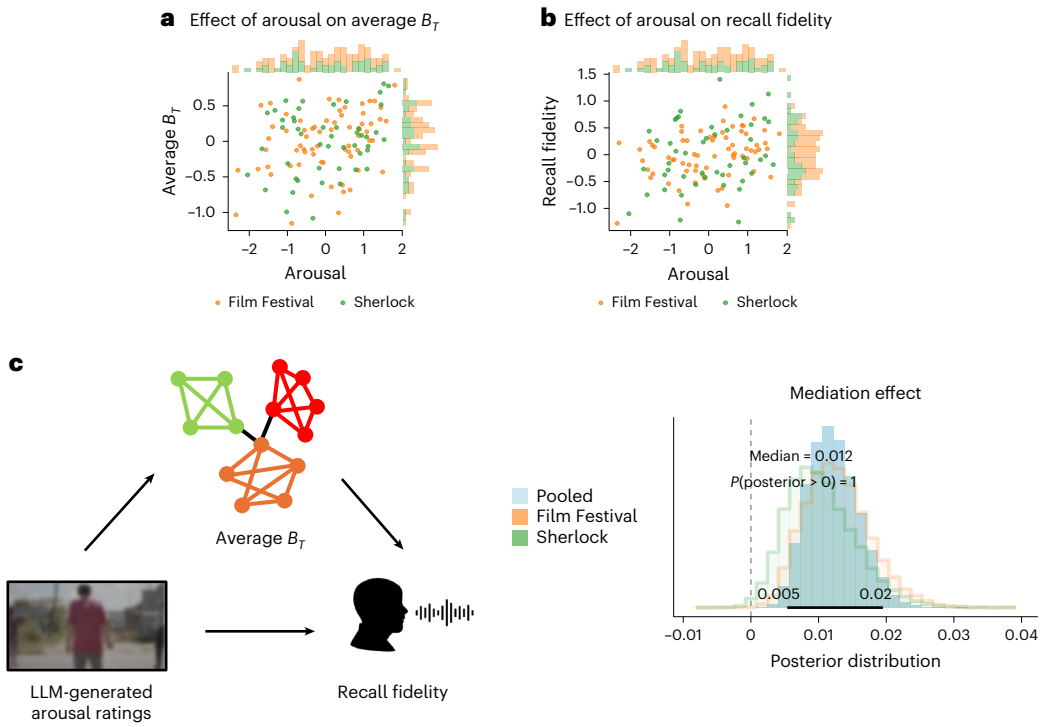


Fig. 3 | Functional network integration mediates the effects of emotional arousal on recall fidelity. a,**b**, Relationship between emotional arousal and average B_T (**a**) and recall fidelity (**b**). All variables were *z*-scored across participants. Each point corresponds to an event, with the datasets differentiated by colour. The marginal histograms display the distribution of arousal and average B_T or recall fidelity for each dataset. The corresponding posterior distributions are shown in Extended Data Fig. 3. **c**, Mediation analysis testing whether average B_T during an event mediates the effect of emotional arousal on recall fidelity. The overlaid histograms show the posterior distributions of the mediation (indirect) effect from Bayesian multilevel models estimated separately for each dataset and for the pooled model. The 95% HDI of the regression coefficient estimated using the pooled model is indicated by the bold horizontal line. The video screen capture is blurred for copyright reasons.

Human behavioural ratings (Film Festival: median = 3.2; range = 1.3–4.7; Sherlock: median = 3; range = 1.5–4.7) were highly correlated across participants (Film Festival: one-to-average $r = 0.72$, non-parametric $P < 0.001$; Sherlock: one-to-average $r = 0.69$, non-parametric $P < 0.001$), indicating consistency across individuals on the level of emotional arousal of each event. LLM-generated ratings (Film Festival: median = 3.3; range = 1.6–4.9; Sherlock: median = 3.5; range = 1.9–4.5) correlated with the average behavioural ratings, providing convergent validity of the arousal measures (Film Festival: Spearman's $\rho_{66} = 0.60$, $P < 0.001$; Sherlock: Spearman's $\rho_{46} = 0.75$, $P < 0.001$; Extended Data Fig. 2). Subsequent analyses yielded similar results for both the LLM-generated and average behavioural ratings. For brevity, we report the results using the LLM-generated ratings here; the results based on behavioural ratings are reported in 'Robustness checks' in the Supplementary Information.

We next sought to relate the arousal ratings to our graph measures of network integration. A Bayesian multilevel model pooling across Film Festival and Sherlock indicated that average B_T was higher when participants viewed events rated as more emotionally arousing ($\beta = 0.12$; 95% HDI, [0.07, 0.16]; $P(\beta > 0) = 1$; WAIC = 5,078; $\text{WAIC}_{\text{null}} = 5{,}105$; $\text{BF}_{10} = 4.6 \times 10^4$; Fig. 3a and Extended Data Fig. 3a), and that events rated as more emotionally arousing were recalled with greater fidelity ($\beta = 0.14$; 95% HDI, [0.1, 0.18]; $P(\beta > 0) = 1$; WAIC = 4,955; $\text{WAIC}_{\text{null}} = 4{,}996$; $\text{BF}_{10} = 3.4 \times 10^7$; Fig. 3b and Extended Data Fig. 3b).

Furthermore, average B_T while viewing an event mediated the effects of emotional arousal on recall fidelity ($\beta = 0.01$; 95% HDI, [0.01, 0.02]; $P(\beta > 0) = 1$; WAIC = 4,943; Fig. 3c), which accounted for 9% of the total effect of arousal on recall fidelity. Here we do not compare against a null model, as removing the mediator tests the unique variance it explains in the outcome rather than the presence of mediation. Instead, we rely on the proportion of the posterior distribution of the indirect effect that is greater than zero to assess the strength of the evidence for mediation[65]. To further assess the robustness of our results, we show that the mediation effect replicates with global efficiency, an alternative measure of whole-brain functional integration ($\beta = 0.02$; 95% HDI, [0.01, 0.03]; $P(\beta > 0) = 1$; WAIC = 4,918).

When event duration, visual intensity and audio intensity were included as covariates, the mediation effect was attenuated but remained statistically credible, indicating that network integration explained unique variance beyond these features (average B_T: $\beta = 0.004$; 95% HDI, [0, 0.01]; $P(\beta > 0) = 0.989$; WAIC = 4,833; global efficiency: $\beta = 0.005$; 95% HDI, [0, 0.01]; $P(\beta > 0) = 0.994$; WAIC = 4,832). Additionally, the mediation effect replicated independently in each dataset, with functional connectivity matrices computed after global signal regression, different parcellation schemes and network thresholding parameters (see 'Robustness checks' in the Supplementary Information). Altogether, these findings provide consistent evidence in support of our core hypothesis that emotional arousal enhances memory encoding by facilitating increased functional integration across brain networks.

Network efficiency within and between canonical brain networks mediates the effects of emotional arousal on narrative recall

While average B_T provides a measure of the overall levels of network integration across functional modules in the brain, it is agnostic to the brain regions that drive this integration. Thus, different configurations of large-scale network organization might underlie the observed effects. Are there specific network interactions that mediate the effects of emotional arousal on memory encoding? To address this question,

https://doi.org/10.1038/s41562-025-02315-1

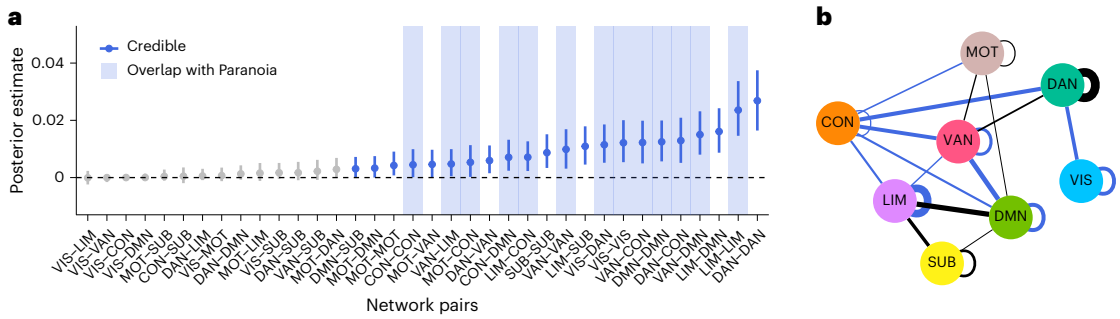


Fig. 4 | Within- and between-network efficiency mediates the effects of emotional arousal on recall fidelity in audiovisual datasets. a, The dots and whiskers indicate the median and 95% HDI of the posterior distribution of the mediation effect for each network connection, estimated from the pooled sample of 32 participants. Data points are presented in blue if the mediation effect was credible after controlling for multiple comparisons (expected PEP < 0.01). Blue shading indicates network connections that were also identified in the Paranoia dataset. Network connections are ordered by the median. VIS, visual network; MOT, somatomotor network; DAN, dorsal attention network; VAN, ventral attention network; LIM, limbic network; CON, control network; DMN, default mode network; SUB, subcortical network. **b**, Each node represents a functional network. The edges represent the credible connections after controlling for multiple comparisons, and edge weights denote the median of the posterior estimate. Black edges indicate connections that were identified only in the audiovisual datasets, while blue edges indicate connections that were also identified in the Paranoia dataset.

we ran additional analyses to characterize the contributions of connectivity patterns within and between canonical functional brain networks.

We grouped the brain regions on the basis of their assignment to seven functional networks defined by the Schaefer atlas (dorsal attention, ventral attention, control, default mode, visual, motor and limbic), with the 16 subcortical regions of interest (ROIs) grouped together as one network. For each network, we calculated the within-network efficiency[66]. For each pair of networks, we calculated the between-network efficiency[49] (Methods). The within- and between-network efficiency provided measures that are often interpreted as reflecting the potential efficiency of information transfer within and across different brain networks, respectively[59,66].

We tested whether within- and between-network efficiency mediated the effects of emotional arousal on recall fidelity. To correct for multiple comparisons, we thresholded the results at an expected posterior error probability (PEP) of 0.01 (ref. 67) (Methods). Within-network efficiency in all eight functional networks and between-network efficiency in 14 out of 28 possible network pairs mediated the effects of emotional arousal on recall fidelity (Fig. 4a,b; see Extended Data Table 1 for the full statistical details). The between-network connections included connections across all eight functional networks, suggesting that the enhanced encoding of emotionally arousing experiences is supported by brain-wide integration.

Amygdalar and hippocampal engagement during emotionally arousing events

Given the notable contributions of individual brain regions to memory and emotional processing, we focused our subsequent analyses on two specific ROIs commonly implicated in enhanced memory encoding[21,27]: the amygdala and the hippocampus. The events in our study spanned 30 seconds to over a minute (Film Festival mean, 38.4 s; Sherlock mean, 57.5 s). As averaging amygdalar and hippocampal activity over the course of an event fails to capture the time-locked neural responses evoked by dynamic stimuli, we employed inter-participant correlation (IPC) analyses[68] to identify shared, stimulus-locked neural responses across participants. Prior work using the Sherlock dataset found that events with higher hippocampal IPC during encoding were more likely to be remembered[54]. Here we calculated the IPC of the amygdala and hippocampus for each event and tested their relationship with arousal, integration and recall fidelity using Bayesian multilevel models that pooled data across both datasets (Fig. 5a,b).

Amygdalar and hippocampal IPC were positively associated with emotional arousal, consistent with the engagement of these regions by emotionally arousing content (amygdala: $\beta = 0.07$; 95% HDI, [0.02, 0.11]; $P(\beta > 0) = 0.998$; WAIC = 5,197; $\text{WAIC}_{\text{null}} = 5{,}203$; $\text{BF}_{10} = 1.5$; hippocampus: $\beta = 0.09$; 95% HDI, [0.04, 0.13]; $P(\beta > 0) = 1$; WAIC = 5,193; $\text{WAIC}_{\text{null}} = 5{,}206$; $\text{BF}_{10} = 33$). Amygdalar IPC was positively associated with average B_T ($\beta = 0.07$; 95% HDI, [0.02, 0.11]; $P(\beta > 0) = 0.998$; WAIC = 5,094; $\text{WAIC}_{\text{null}} = 5{,}105$; $\text{BF}_{10} = 13$), while hippocampal IPC was not ($\beta = 0.02$; 95% HDI, [−0.03, 0.06]; $P(\beta > 0) = 0.78$; WAIC = 5,102; $\text{WAIC}_{\text{null}} = 5{,}106$; $\text{BF}_{10} = 0.3$).

Both amygdalar and hippocampal IPC were positively associated with recall fidelity (amygdala: $\beta = 0.07$; 95% HDI, [0.02, 0.11]; $P(\beta > 0) = 0.999$; WAIC = 4,987; $\text{WAIC}_{\text{null}} = 4{,}996$; $\text{BF}_{10} = 4.5$; hippocampus: $\beta = 0.07$; 95% HDI, [0.03, 0.11]; $P(\beta > 0) = 1$; WAIC = 4,986; $\text{WAIC}_{\text{null}} = 4{,}996$; $\text{BF}_{10} = 10$). Additionally, amygdalar and hippocampal IPC mediated the effects of arousal on recall fidelity (amygdala: $\beta = 0.004$; 95% HDI, [0, 0.01]; $P(\beta > 0) = 0.992$; WAIC = 4,950, 3% of total effect; hippocampus: $\beta = 0.004$; 95% HDI, [0, 0.01]; $P(\beta > 0) = 0.996$; WAIC = 4,948, 4% of total effect). Although inference was based on the pooled model to maximize statistical power, the results from individual datasets were directionally consistent in all cases. Most, but not all, dataset-specific estimates met the predetermined credibility threshold of 0.95 (see 'Dataset-specific IPC results' in the Supplementary Information).

These results are consistent with prior findings indicating the importance of the amygdala and hippocampus to the enhancement of emotional memories. The mediation effect of network integration on the relationship between emotional arousal and recall fidelity was robust to controlling for amygdalar and hippocampal IPC ($\beta = 0.01$; 95% HDI, [0, 0.02]; $P(\beta > 0) = 1$; WAIC = 4,933), as well as average activity in the amygdala, hippocampus and amygdala–hippocampus functional connectivity ($\beta = 0.01$; 95% HDI, [0.01, 0.02]; $P(\beta > 0) = 1$; WAIC = 4,937), indicating a unique contribution of large-scale functional integration above and beyond that explained by amygdalar and hippocampal engagement.

Functional integration mediates the effects of pupil dilation on narrative recall

Thus far, we have demonstrated that emotional arousal, as assessed through subjective ratings and LLM-derived measures, is linked to heightened network integration and enhanced recall fidelity. Next, we assessed whether our findings would generalize to a physiological measure of arousal. We collected a new dataset ($n = 27$) where participants listened to a suspenseful auditory story while their pupil size was continuously measured as a physiological measure of arousal.

https://doi.org/10.1038/s41562-025-02315-1

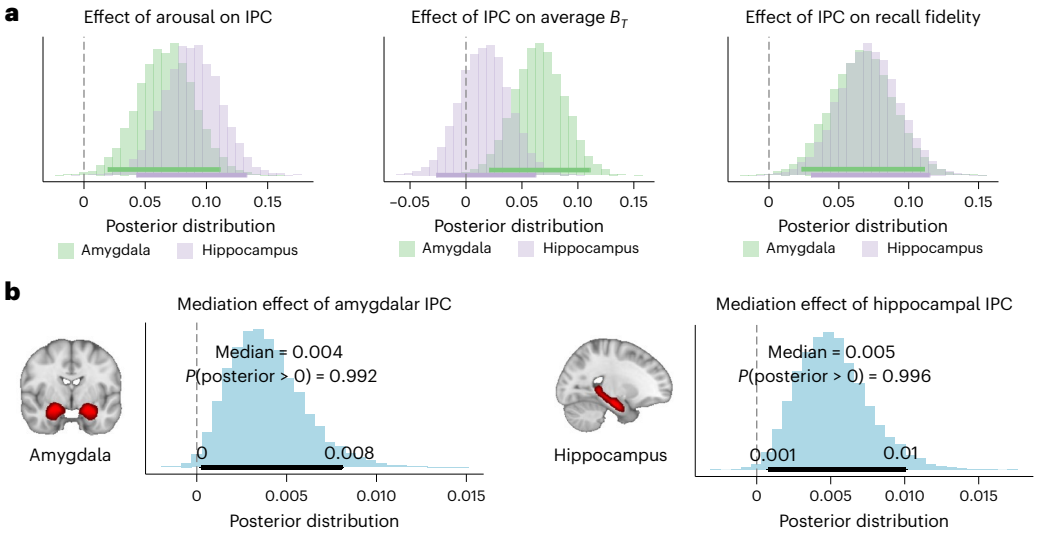


Fig. 5 | Emotional arousal was associated with hippocampal and amygdalar engagement during encoding. a, Emotionally arousing events were associated with greater IPC in the amygdala and hippocampus. The overlaid histograms show the posterior distributions from each model for both brain regions, with the 95% HDIs shown as bold horizontal lines. In the amygdala, greater IPC was associated with increased arousal (left), stronger integration across functional modules (middle) and higher recall fidelity (right). In the hippocampus, greater IPC was linked to increased arousal and recall fidelity, but not integration. **b**, Mediation analyses testing whether IPC in the amygdala (left) and hippocampus (right) mediates the effect of emotional arousal on recall fidelity. The 95% HDIs of the posterior distributions are indicated by the bold horizontal lines.

We further used an open dataset (Paranoia, $n = 22$) of participants listening to the same story while undergoing fMRI[69]. We chose an audio-only stimulus to eliminate luminance-related confounds during pupillometry and to isolate the effects of narrative-driven arousal. Emotional intensity fluctuates over the course of the story[57], making it well suited for examining how moment-to-moment changes in arousal relate to event-level memory encoding. Furthermore, the story's moderate duration (~20 min) made it feasible to include both narrative listening and free recall in a single session.

The story was segmented into 24 events, and pupil size was averaged for each event. Pupil size across events was correlated between participants (one-to-average $r = 0.60$, non-parametric $P < 0.001$), indicating that the story evoked arousal in a reliable manner across participants. Pupil size was associated with greater recall fidelity ($\beta = 0.13$; 95% HDI, [0.05, 0.2]; $P(\beta > 0) = 0.999$; WAIC = 1,794; $BF_{10} = 15.34$; Fig. 6a), indicating that arousing events were better remembered. Pupil size was also positively associated with greater functional network integration ($\beta = 0.33$; 95% HDI, [0.26, 0.4]; $P(\beta > 0) = 1$; WAIC = 1,743; $BF_{10} = 94 \times 10^{13}$; Fig. 6b). Consistent with our earlier results, functional network integration at encoding was associated with subsequent recall fidelity ($\beta = 0.21$; 95% HDI, [0.13, 0.28]; $P(\beta > 0) = 1$; WAIC = 1,776; $BF_{10} = 11 \times 10^4$; Fig. 6c).

Finally, we found that functional network integration mediated the effects of pupil dilation on recall fidelity ($\beta = 0.06$; 95% HDI, [0.03, 0.09]; $P(\beta > 0) = 1$; WAIC = 1,775; Fig. 6d), which accounted for 46% of the total effect of arousal on recall fidelity. A post hoc robustness analysis confirmed that the mediation effect was reliable at the current sample size (Extended Data Fig. 4). The mediation effect replicated when we used global efficiency as the measure of network integration ($\beta = 0.09$; 95% HDI, [0.06, 0.12]; $P(\beta > 0) = 1$; WAIC = 1,762) and was robust to controlling for event duration and audio intensity (average B_T: $\beta = 0.05$; 95% HDI, [0.02, 0.07]; $P(\beta > 0) = 1$; WAIC = 1,690; global efficiency: $\beta = 0.06$; 95% HDI, [0.03, 0.09]; $P(\beta > 0) = 1$; WAIC = 1,685). These results demonstrate that our earlier findings generalize to a physiological measure of arousal.

Within-network efficiency in five of the eight functional networks (ventral attention, control, default mode, visual and limbic) and between-network efficiency in 10 out of 28 possible network pairs mediated the effects of emotional arousal on recall fidelity (Fig. 6e; see Extended Data Table 2 for the full statistical details). Of these 15 connections, 13 overlap with those that were identified in the pooled analysis of the Film Festival and Sherlock datasets (Fig. 7). We note that differences between Paranoia and the audiovisual datasets should be interpreted with caution as they may be related to lower statistical power in the smaller Paranoia sample. Nevertheless, an overlap in the majority of connections across datasets suggests that there is a core set of network interactions supporting the effects of arousal on memory encoding.

Discussion

Emotionally arousing experiences are often remembered with great clarity and detail[1,2]. In the present study, we combined fMRI, graph theory and a naturalistic memory paradigm to study the emotional enhancement of narrative memories. Across three independent fMRI datasets, we demonstrated that whole-brain functional network integration during narrative perception mediated the effects of heightened emotional arousal on the fidelity with which participants subsequently remembered an event. The results replicated across measures of arousal obtained from text analyses of the content, behavioural ratings and pupil dilation, as well as across movies and audio stories. The mediation effect was observed in interactions across multiple large-scale brain networks, indicating that emotional arousal may enhance memory by promoting coordinated activity across diverse brain regions. Together, our findings suggest that functional network integration supports the enhanced encoding of emotional memories by facilitating coordinated activity across the brain.

Our findings extend the neuroscientific investigation of emotional memories to complex, temporally extended narratives, which are fundamental to how people make sense of experiences, share knowledge and connect with others[52,70,71]. Prior work has demonstrated that emotionally arousing content enhances the encoding of narrative memories[1,2,31]. Here we show that this memory enhancement is associated with increased network integration during narrative encoding. Encoding narratives involves the processing of multisensory information[72], integration with existing knowledge and schemas[73], and alignment with ongoing goals and internal states[74]. An integrated brain state, where different functional modules are highly interconnected, facilitates information exchange and coordination across

https://doi.org/10.1038/s41562-025-02315-1

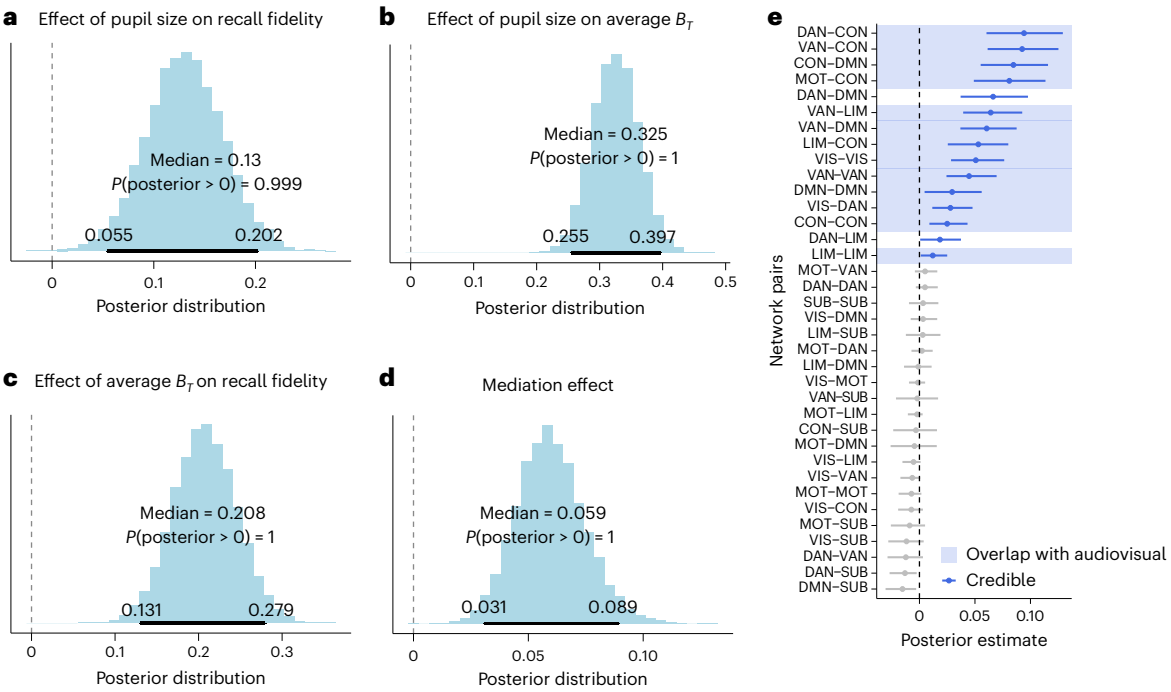


Fig. 6 | Functional network integration mediates the effects of pupil dilation on recall fidelity. a,**b**, Pupil size during story listening was associated with recall fidelity (**a**) and average B_T (**b**). **c**,**d**, Average B_T during story listening was associated with recall fidelity (**c**) and mediated the effects of pupil size on recall fidelity (**d**). Posterior distributions were estimated by a Bayesian multilevel model. The 95% HDI of each distribution is indicated by a bold horizontal line. **e**, The dots and whiskers indicate the median and 95% HDI of the posterior distribution of the mediation effect for each network connection, estimated from a sample of 22 participants. Data points are presented in blue if the mediation effect was credible after controlling for multiple comparisons (expected PEP < 0.01). Blue shading indicates network connections that were also identified in the audiovisual datasets (Film Festival and Sherlock). Network connections are ordered by the median.

regions[36,75,76]. The enhanced connectivity across perceptual, associative and memory-related regions may support the binding of narrative events into a coherent memory representation, allowing for better encoding of event details and their broader context.

The current results contribute to the growing body of literature relating arousal to brain network organization. Our findings are consistent with previous studies that demonstrated changes in network organization following shifts in arousal states. For example, Kinnison and colleagues[47] found increased functional integration across the brain while participants viewed cues indicating potential threats (for example, a shock) or monetary rewards. Similarly, electrophysiological studies in animal models and fMRI studies in humans have revealed that fluctuations in autonomic arousal induce changes in functional network topology[33,41,43,45,77]. Building on these findings, our study demonstrates the behavioural consequences of these arousal-dependent network dynamics. Specifically, we show that arousal may prioritize information for encoding by facilitating communication across different functional brain networks, leading to greater fidelity during subsequent recall. These findings highlight the potential importance of arousal-dependent network interactions in shaping ongoing cognition.

One of the strengths of graph theory is its ability to extract network-level metrics that provide insights into how the brain functions as an interconnected system. This approach allows us to capture emergent properties, such as the level of integration across functional modules, that become apparent only when considering the entire network. As a network-level property, integration can arise from various configurations of interactions between functional modules. By framing these network interactions within the broader concept of integration, our approach provides a unifying framework to explain the diverse neural pathways that strengthen emotional memories. Within this framework, we can also examine the specific within- and between-network interactions that reliably support emotional memory across contexts, identifying common network configurations that facilitate the encoding of emotionally meaningful information.

To that end, we identified a core set of network interactions that consistently mediated the effects of emotional arousal on memory. These include interactions between the control network and the ventral attention network (that is, the "salience" network[78]), consistent with prior work showing increased connectivity between these networks during emotionally arousing experiences[42]. The ventral attention network supports the detection of salient information[79], and increased efficiency between these networks may enhance the allocation of cognitive resources to emotionally charged narrative moments. We also observed consistent mediation effects involving the default mode and the ventral attention networks, which may facilitate the incorporation of externally emotionally salient information into internally constructed narrative models[80,81]. Additionally, mediation by the efficiency between the limbic network and both the control and ventral attention networks suggests that arousal-related memory enhancement may depend on coordination between affective processing systems and networks supporting executive control and attentional allocation[82].

Arousal-related memory enhancement was also mediated by the efficiency between the dorsal attention and control networks. The involvement of the dorsal attention network, which is typically associated with the top-down control of attention, aligns with prior evidence that the enhanced encoding of emotional stimuli depends on attentional processes[83]. From this perspective, attention and arousal are interdependent processes that work in concert to prioritize the encoding of emotionally charged narrative moments. This attentional modulation is potentially supported by the locus coeruleus–norepinephrine system, which is activated by arousing stimuli and is known to enhance network integration[33,44,84] and attention control[85,86]. Consistent with this account, our study found that pupil dilation, a correlate of

https://doi.org/10.1038/s41562-025-02315-1

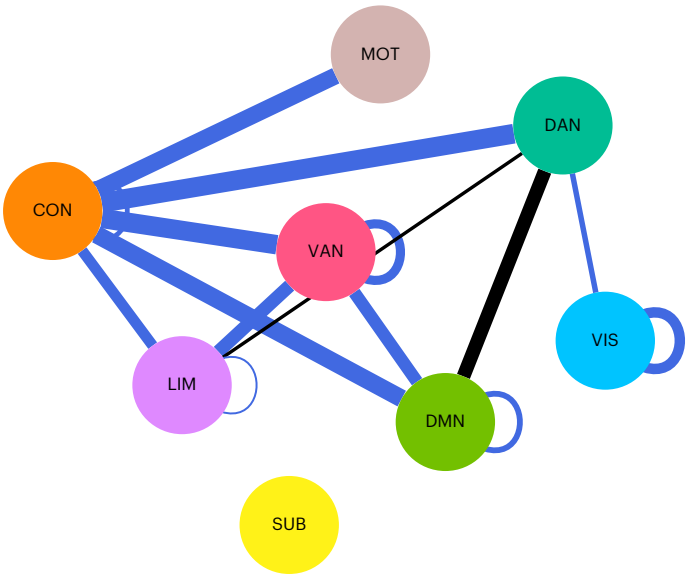


Fig. 7 | Credible within- and between-network pairs that mediated the effects of pupil size on recall fidelity. Each node represents a functional network. The edges represent the credible connections after controlling for multiple comparisons, and edge weights denote the median of the posterior estimate. Black edges indicate connections that were identified only in the Paranoia dataset, while blue edges indicate connections that were also identified in the pooled audiovisual dataset.

locus coeruleus–norepinephrine activity[51], was associated with both network integration and enhanced memory encoding.

The observed mediation effect between the somatomotor and control networks aligns with a recent meta-analysis identifying the pre-supplementary motor area, situated at the intersection of these networks, as one of two regions consistently associated with arousal states (the other being the anterior insula)[87]. We also found that efficiency within the visual network and between the visual and dorsal attention networks mediated the effects of arousal on memory, even in Paranoia, an audio-only story. Prior work has shown that listening to stories activates early visual areas[88] and that visual cortex activity encodes schema information from the stories[74]. These findings are often interpreted as evidence of mental imagery during narrative listening. Thus, one interpretation of the involvement of the visual cortex in Paranoia is that arousing moments elicit vivid imagery and are selectively prioritized by top-down attention, thereby enhancing memory encoding. Future studies using decoding methods can directly test this account.

While our findings emphasize the importance of network dynamics in enhancing memory encoding, it is equally important to consider the roles of specific brain regions within these larger networks. The amygdala and hippocampus, for instance, have been extensively studied for their roles in enhancing emotional memories[19–22,24–27]. Here we found that emotional arousal engages the amygdala and hippocampus in a stimulus-locked manner. While amygdalar engagement was positively associated with network integration, hippocampal engagement was not, suggesting distinct roles for these regions in coordinating arousal-dependent large-scale network dynamics. Both amygdalar and hippocampal engagement were positively associated with recall fidelity and mediated the effects of arousal on memory encoding, extending prior work on images, words and sounds[19–27] to temporally extended narratives. Importantly, the mediation effects of network integration remained significant when controlling for amygdalar and hippocampal engagement, indicating that the effects of emotional arousal on memory encoding cannot be fully explained by activity in these regions alone.

One potential concern about our design is that the narrative events may not have been sufficiently arousing to engage neural processes implicated in emotional memory enhancement. We consider this unlikely for several reasons. Emotional words have been shown to reliably elicit emotion-related memory effects[22,23], engaging similar neural processes as highly aversive images. All three narratives in our study contained highly arousing scenes; for example, the *Sherlock* episode included a battlefield sequence and a moment where a main character's life was under threat—events that are arguably more arousing than isolated words. Indeed, behavioural ratings of emotional arousal in the Sherlock and Film Festival datasets spanned nearly the full range of the 1–5 scale, suggesting meaningful variability in event-level arousal. Furthermore, heightened arousal was associated with increased amygdalar engagement, replicating prior work using images and words. In the Paranoia dataset, pupil dilation was synchronized across participants, indicating stimulus-driven fluctuations in arousal that were shared across listeners. These converging behavioural, physiological and neural measures indicate that the narrative events elicited robust and reliable arousal responses.

Nevertheless, it is important to recognize that watching a movie or listening to a story may differ qualitatively from personally relevant emotional experiences. Extending this work to autobiographical experiences may help clarify how these mechanisms contribute to the formation of vivid and enduring memories formed around personally meaningful emotional events. This may be particularly relevant in the context of trauma, where emotional memories are often intensely vivid and disruptions in network integration have been observed[89]. Another limitation is that the three datasets were acquired under different conditions and at different sites. While this heterogeneity strengthens the generalizability of our findings, it also complicates cross-dataset comparisons. Sex was relatively balanced across the audiovisual and audio datasets, but the sample sizes precluded meaningful examination of sex-related differences. We note also that the participants were primarily young adults, and future work is needed to assess how these findings extend to older populations, particularly given evidence that locus-coeruleus-linked arousal responses[90] and neural responses related to event perception undergo age-related changes[91].

The mediation effects we observed were small and were constrained by the modest direct association between emotional arousal and memory fidelity. This is possibly due to memory of complex, temporally extended narratives being influenced by many factors beyond emotional arousal, including prior knowledge[92,93], the extent to which an event aligns with existing schemas[73,94,95], and causal and semantic relationships[53,96,97]. Additionally, our arousal measures were not collected from the fMRI participants, which may have obscured meaningful individual differences in emotional responses. Incorporating participant-specific measures would yield stronger direct and mediation effects. Nonetheless, even modest effects can reflect meaningful contributions in naturalistic settings, where memory formation is subject to overlapping influences. Here we show that mediation via network integration was larger in magnitude than mediation via either hippocampal or amygdalar engagement alone. Rather than suggesting that arousal-dependent changes in network integration are the sole driver of memory encoding, our findings identify it as one notable and previously underexplored contributor.

Taken together, our findings provide robust, converging evidence that emotional arousal enhances narrative memories through the functional integration of large-scale brain networks. While our study focused on memory, emotional arousal is also known to influence a broad range of other cognitive processes, including attention[98,99], perception[100,101] and decision-making[102,103]. Our work lays the foundation for future research examining interactions operating at different scales, from local activity in the amygdala and hippocampus to brain-wide network dynamics, and how they facilitate the influence of emotional arousal on human cognition and behaviour.

Methods

Film Festival dataset

Fifteen participants (10 female, 5 male; ages 21 to 33; mean age, 27.5) watched ten short movie clips over the course of two functional runs (repetition time = 1.5 s; echo time = 39 ms; flip angle, 50°; voxel size, 2 × 2 × 2 mm^3). Each clip was approximately 5 min long (2.15–7.75 min), and the clips varied in content, characters and genre. Following movie watching, the participants were instructed to verbally describe the clips from memory. The participants provided informed consent and were compensated at $20 per hour. Raw anatomical and functional data were downloaded from OpenNeuro[104]. Additional details can be found in the original study[53]. MRI data were preprocessed using FSL/FEAT v.6.00 (FMRIB software library, FMRIB). The steps included motion correction of functional images, removal of low-frequency drifts using a temporal high-pass filter (100 s cut-off) and spatial smoothing (5-mm full width at half maximum). Functional images were registered to participants' anatomical image (rigid-body transformation with 6 degrees of freedom) and to an Montreal Neurological Institute template (affine transformation with 12 degrees of freedom). Following the original authors of this study[53], we discarded the first two volumes of each run to avoid the non-specific transient increase in BOLD response at the beginning of each run. The data from each run were shifted by 4.5 s to account for haemodynamic lag and then normalized by *z*-scoring across time within each participant.

Sherlock dataset

Seventeen participants (7 female, 10 male; ages 19 to 26; mean age, 20.8) watched a segment from an episode of the British TV show *Sherlock* over the course of two functional runs (repetition time = 1.5 s; echo time = 28 s; voxel size, 3 × 3 × 4 mm^3). Following movie watching, the participants were instructed to verbally describe the clips from memory. The participants provided informed consent and were compensated at $20 per hour. Preprocessed functional and anatomical data were downloaded from Princeton University's data storage repository, DataSpace[105]. The data had already undergone motion correction, slice timing correction, linear detrending, high-pass filtering, spatial smoothing, normalization to Montreal Neurological Institute template, *z*-scoring across each run and shifting by 4.5 s to correct for haemodynamic lag, as detailed in the original publication[54]. No additional preprocessing was applied.

Paranoia dataset

Twenty-two participants (11 female, 11 male; ages 19 to 35; mean age, 27.0) listened to a 20-minute-long story involving a mysterious social event over the course of three functional runs (repetition time = 1 s; echo time = 30 ms; flip angle, 8°; voxel size, 2 × 2 × 2 mm^3). The participants provided informed consent and were compensated with $75 for the MRI session. Raw anatomical and functional data were downloaded from OpenNeuro[106]. Each run was approximately 7.2 min long (5.5-8.7 min). Preprocessing steps (motion correction, smoothing and registration) were identical to those of the Film Festival dataset. The data from each run were shifted by five TRs to account for haemodynamic lag and then normalized by *z*-scoring across time within each participant.

Movie event segmentation and annotation

For the Sherlock stimulus, we used the event segmentation and annotation reported in ref. 54. Briefly, the 50-minute episode was divided into 48 events (mean, 57.5 s; s.d., 41.7 s) on the basis of significant shifts in the narrative, such as changes in topic, location, time and characters. An annotator then provided written descriptions about what was happening in the movie during that event. K.G. repeated the procedure with the Film Festival stimulus, generating 68 events (mean, 38.4 s; s.d., 18.2 s) and corresponding written annotations of each event.

The audio intensity of the Film Festival, Sherlock and Paranoia datasets was computed using audioread in MATLAB (v.r2024b, Mathworks). Specifically, we extracted the sound envelope by applying the Hilbert transform and taking the absolute value. The envelope signal was then averaged over the event. Frame-wise visual intensity for the Film Festival and Sherlock datasets was computed by converting each video frame from RGB to HSV colour space using MATLAB's rgb2hsv function. We extracted the 'value' (V) channel, which corresponds to brightness, and averaged the pixel values within each frame. The visual intensity for each event was then obtained by averaging brightness values across all frames within the event.

Recall transcripts

Transcripts of verbal memory recall were obtained from refs. 53 and 54. The transcripts were then segmented and manually matched to the corresponding events on the basis of the annotations that best matched the verbal descriptions. For Sherlock, the onset and offset of the events that were remembered were provided by Chen et al.[54]. For Film Festival, the associated time stamps were identified by K.G.

Recall fidelity

We computed the fidelity of participants' recall by assessing the semantic similarity between participants' recall transcripts and movie annotations. For each event, we converted both the participants' recall transcripts and the corresponding movie annotations into separate 512-dimensional vectors using Google's USE[55]. Sentences with similar meanings are encoded in vectors that are closer in the embedding space. We chose the USE model because it generates a single embedding vector that captures the overall meaning of text spanning multiple sentences. We then calculated the recall fidelity of each event as the cosine similarity between the vector encoding a participant's recall transcript and that encoding the movie annotation. USE has been frequently employed to measure semantic similarity between event descriptions[38,53,58,107]. Our approach draws on prior work computing the cosine similarity between embedding vectors encoding the annotation of an event and participants' recall as measures of memory fidelity[56–58].

LLM measure of emotional arousal

We generated ratings of emotional arousal from the event annotations using StableBeluga-13B (https://huggingface.co/stabilityai/StableBeluga-13B), an open-access LLM[62]. Researchers are increasingly using LLMs as tools for automated text analysis[108]. For example, LLM-generated valence and arousal ratings of text correspond well to those obtained from behavioural participants[109,110]. We provided the model with event annotations and prompted it to rate the arousal of each event on a scale from 1 to 10. To facilitate comparison with human ratings, which were collected on a 1-to-5 scale, we divided the model-generated ratings by 2. As these ratings were *z*-scored prior to analyses, the rescaling would not affect the results. As part of the model prompt, we defined arousal as a state of "feeling very mentally or physically alert, activated, and/or energized" (see 'Large language model (LLM) prompt' in the Supplementary Information for the full model prompt). These instructions were adapted from our earlier work[12] and drew on dimensional models of affect, where arousal reflects the level of psychological activation or intensity associated with a stimulus or experience and is orthogonal to valence, which reflects the degree of pleasantness or unpleasantness[10,11]. We avoided using the anchors 'calm' and 'excited', as employed in some prior studies[25], because these terms may be confounded with positive valence and could introduce bias into the LLM's responses. Due to the inherent stochasticity in the model's responses, we ran 30 iterations for each event, effectively simulating ratings from 30 'participants'. These ratings were *z*-scored within each iteration to normalize the data and then averaged across iterations to generate an average arousal rating for each event.

Behavioural measure of emotional arousal

As a comparison and validation of the LLM-generated arousal ratings, we collected behavioural ratings of emotional arousal from 30 participants (5 male, 24 female, 1 non-binary; ages 21 to 36; mean age, 26.23). The experimental procedures were approved by the University of Chicago Institutional Review Board, and the participants provided informed consent prior to the start of the study. The participants were compensated with $30 for their time. The participants were provided with the same definition of arousal as the LLM and were instructed to rate the arousal of each event on a scale of 1 to 5 after watching each event. While the LLM ratings used a 1-to-10 scale to capture finer-grained distinctions, we opted for a coarser five-point scale for human participants to reduce cognitive load and rating fatigue, given the large number and extended duration of the narrative events. The participants were allowed to pause the video to make their ratings. All participants provided ratings on both Film Festival and Sherlock, with a short break in between. The ratings were *z*-scored within each participant and averaged across participants to generate an average arousal rating for each event. We correlated the LLM and behavioural measures to assess the consistency across the two methods of eliciting arousal ratings.

Generating graphs from functional connectivity matrices

Cortical regions were parcellated into 200 ROIs taken from the Schaefer atlas[111]. Subcortical regions were divided into 16 ROIs on the basis of the Melbourne subcortex atlas[112]. Similar results were observed when we repeated our analyses with the Shen parcellation scheme that included 268 parcels encompassing both cortical and subcortical regions (see 'Robustness checks' in the Supplementary Information for additional justification of the parcellation scheme). We extracted ROI time courses by averaging all voxels within each ROI. We then regressed out mean white matter and mean cerebrospinal fluid time courses, and framewise displacement from each ROI time course to minimize the influence of motion, non-neuronal noise and other nuisance signals on the functional connectivity estimates. Global signal regression was not included in the main analyses as it can introduce spurious correlations between brain regions[113,114]. However, we note that our results remained consistent when global signal regression was applied (see 'Robustness checks' in the Supplementary Information).

For each event in each participant, we calculated the Fisher *z*-transformed Pearson correlation between the BOLD time courses of each pair of ROIs, resulting in a 216 × 216 functional connectivity matrix. Each matrix was then thresholded to retain the top 15% strongest connections and binarized to create an unweighted, undirected graph. To ensure the robustness of our findings, we replicated our results at different thresholds (10%, 20% and 25%; Supplementary Information). All graph theoretic metrics (see below) were then extracted using the Brain Connectivity Toolbox[59].

Average participation coefficient

Consistent with prior studies[33,35,49], brain regions were assigned to communities (that is, modules) using the Louvian algorithm, which seeks to identify a community structure that maximizes the number of within-community connections and minimizes between-community connections. Given the stochastic nature of the algorithm, the community assignment was repeated 1,000 times to obtain a consensus community structure for each participant and event.

We then calculated B_T for each ROI, which quantifies the extent to which the region's connections are distributed across different communities:

$$B_{Ti} = 1 - \sum_{m \in M} \left(\frac{k_{i,m}}{k_i} \right)^2$$

where k_i is the degree (that is, the total number of connections) of region i, and $k_{i,m}$ is the number of connections that region i has with regions in community m. Thus, B_T ranges from 0 to 1, with higher values indicating that a region's connections are evenly distributed across communities, and lower values indicating that a region's connections are predominantly within its assigned community. To assess whole-brain network integration, we averaged B_T across all ROIs for each participant and event. The average B_T was used as our primary measure of network integration, with higher values indicating a more integrated brain state during movie viewing.

Network efficiency measures

We calculated global efficiency (E_g) as an alternative measure of network integration[59]. E_g is defined as the average inverse shortest path length between all pairs of nodes in the network:

$$E_g = \frac{1}{n(n-1)} \sum_{i \neq j \in N} \frac{1}{d_{ij}}$$

where n is the total number of regions, N is the set of all regions in the network and d_{ij} is the length of the shortest path between nodes i and j. Higher E_g is indicative of greater network integration where fewer connections separate any two brain regions.

Within-network efficiency was calculated as the average inverse shortest path length between all pairs of regions within a network[66]. Between-network efficiency of networks 1 and 2 was calculated as the average inverse shortest path length between each region in network 1 and each region in network 2 (ref. 49). For all efficiency measures, the shortest path was defined over all regions in the brain. In other words, when calculating between-network efficiency between networks i and j, the shortest path may include regions outside of the two networks.

Bayesian multilevel models

We used Bayesian multilevel models to analyse the relationship between brain network metrics, emotional arousal and recall fidelity. These models were chosen for their ability to account for the hierarchical structure of the data, where events were nested within participants, and to estimate uncertainty in the parameter estimates directly. All models were implemented using the brms package (v.2.21.0)[115] in R v.4.3.2[116] and were run with the random seed '123' to ensure reproducibility. The models included random intercepts for predictors to account for random variability across participants. For models pooling across the Film Festival and Sherlock datasets, we additionally included dataset as a random intercept to account for between-dataset variability.

Models were specified with weakly informative priors to facilitate convergence while still allowing the posterior distributions to be predominantly shaped by the data (see the Supplementary Information for the full prior specification). Each model was run with four Markov chain Monte Carlo chains, each with 4,000 samples, of which the first 1,000 samples were discarded as burn-in. The remaining samples from all chains were concatenated to form the posterior distribution of each parameter. We checked for model convergence by ensuring that the Gelman–Rubin diagnostic ($\hat{R}$) was less than 1.01 for all parameters. Posterior distributions were summarized by reporting the estimated posterior means and 95% HDIs[117]. We considered parameters to be credibly above zero if more than 95% of the posterior distribution was greater than zero.

To evaluate model fit, we computed the WAIC using the loo package (v.2.7.0)[118]. Lower WAIC values indicate better out-of-sample predictive performance. Where appropriate, we compared models against null models that excluded the predictor of interest. We also computed BF_{10} using the bayestestR package (v.0.13.2)[119] to quantify evidence relative to the null model. BF_{10} values greater than 1 indicate evidence in favour of the alternative model, with larger values reflecting stronger support.

Mediation analyses

Mediation effects were computed using the product-of-coefficients method[120]. Specifically, the posterior distribution of the mediation effect of average B_T was obtained by multiplying the posterior samples of the effect of emotional arousal on average B_T ('path a') by the posterior samples of the effect of average B_T on recall fidelity while controlling for emotional arousal ('path b'). To evaluate evidence for mediation, we examined the proportion of the posterior distribution of the mediation that was greater than zero. A mediation effect was considered statistically credible if more than 95% of the posterior distribution exceeded zero, and we report the 95% HDI to describe uncertainty in the estimated effect. For robustness checks, we fit separate models that included event duration, audio intensity and visual intensity (for the audiovisual datasets) as covariates. All other model specifications and inference criteria were identical to those for the primary models.

We grouped the brain regions into the seven functional networks defined by the Schaefer atlas (dorsal attention, ventral attention, control, default mode, visual, motor and limbic networks), with the 16 subcortical ROIs grouped together as one network. We then assessed the mediation effect of the within-network efficiency of each network, as well as the between-network efficiency of every pair of networks. To account for multiple comparisons across networks and network pairs, we thresholded the results within each dataset at an expected PEP of 0.01, corresponding to a less than 0.01 probability of making an incorrect inference (that is, concluding that a parameter is greater than 0 when it is in fact less than 0)[67]. From a frequentist perspective, this procedure is analogous to controlling for a false discovery rate at $q < 0.01$ (ref. 121). Among the networks and network pairs for which the HDI was above zero, we identified those that survived thresholding at an expected PEP < 0.01.

IPC analyses

We extracted the time-course data from the amygdala and the hippocampus for each event. For each participant and event, we calculated the IPC by correlating an individual's time-course data with the average of all other participants' data. This provided an IPC value for the amygdala and the hippocampus for each participant and each event. We then tested whether the IPC was associated with arousal, memory and B_T.

Paranoia behavioural dataset

A total of 35 participants were recruited from the University of Chicago research participation pool (SONA Systems). The experimental procedures were approved by the Institutional Review Board at the University of Chicago, and all participants provided informed consent at the beginning of the study. Three participants were excluded due to equipment failure; five additional participants were excluded due to noisy pupillometry data (see 'Pupillometry' in the Supplementary Information). Thus, 27 participants in total were included in the analysis (12 male, 15 female; ages 20 to 33; mean age, 22.3). Detailed procedures regarding noise detection and preprocessing are reported in the Supplementary Information. The stimuli were presented using Psychopy (v.2022.2.5)[122].

Pupil size was recorded using the Eyelink 1000 eye tracker (SR Research) at a sampling rate of 500 Hz. Prior to recording, the participants completed a five-point calibration sequence. The participants were seated in front of the computer and instructed to fix their eyes on a cross displayed in front of them. After listening to the story, the participants were allowed a short break before beginning the recall task. For the recall task, the participants were asked to recount the story in their own words. They were encouraged to speak for at least 10 min and were permitted to return if they later recalled additional details.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw MRI data from the Film Festival dataset are available via OpenNeuro at https://openneuro.org/datasets/ds004042/versions/1.0.1. The preprocessed MRI data from the Sherlock dataset are available via DataSpace at https://dataspace.princeton.edu/jspui/handle/88435/dsp01nz8062179. The raw MRI data from the Paranoia dataset are available via OpenNeuro at https://openneuro.org/datasets/ds001338/versions/1.0.0. The behavioural and pupillometry data are available via GitHub at https://github.com/jadynpark/arousal_integration.

Code availability

The custom analysis scripts are available via GitHub at https://github.com/jadynpark/arousal_integration.

References

1. Heuer, F. & Reisberg, D. Vivid memories of emotional events: the accuracy of remembered minutiae. *Mem. Cogn.* **18**, 496–506 (1990).
2. McGaugh, J. L. *Memory and Emotion: The Making of Lasting Memories* (Columbia Univ. Press, 2003).
3. LaBar, K. S. & Cabeza, R. Cognitive neuroscience of emotional memory. *Nat. Rev. Neurosci.* **7**, 54–64 (2006).
4. Rouhani, N., Niv, Y., Frank, M. J. & Schwabe, L. Multiple routes to enhanced memory for emotionally relevant events. *Trends Cogn. Sci.* **27**, 867–882 (2023).
5. Clewett, D. V., Huang, R., Velasco, R., Lee, T.-H. & Mather, M. Locus coeruleus activity strengthens prioritized memories under arousal. *J. Neurosci.* **38**, 1558–1574 (2018).
6. Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B. & Gabrieli, J. D. E. Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* **50**, 507–517 (2006).
7. Montagrin, A., Brosch, T. & Sander, D. Goal conduciveness as a key determinant of memory facilitation. *Emotion* **13**, 622–628 (2013).
8. Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E. & Murty, V. P. Memory consolidation as an adaptive process. *Psychon. Bull. Rev.* **28**, 1796–1810 (2021).
9. Shohamy, D. & Adcock, R. A. Dopamine and adaptive memory. *Trends Cogn. Sci.* **14**, 464–472 (2010).
10. Russell, J. A. Core affect and the psychological construction of emotion. *Psychol. Rev.* **110**, 145–172 (2003).
11. Barrett, L. F. Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cogn. Emot.* **12**, 579–599 (1998).
12. Ke, J., Song, H., Bai, Z., Rosenberg, M. D. & Leong, Y. C. Dynamic brain connectivity predicts emotional arousal during naturalistic movie-watching. *PLoS Comput. Biol.* **21**, e1012994 (2025).
13. Lang, P. J., Greenwald, M. K., Bradley, M. M. & Hamm, A. O. Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology* **30**, 261–273 (1993).
14. Bradley, M. M., Miccoli, L., Escrig, M. A. & Lang, P. J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**, 602–607 (2008).
15. Roozendaal, B. & Hermans, E. J. Norepinephrine effects on the encoding and consolidation of emotional memory: improving synergy between animal and human studies. *Curr. Opin. Behav. Sci.* **14**, 115–122 (2017).
16. McGaugh, J. L. Emotional arousal regulation of memory consolidation. *Curr. Opin. Behav. Sci.* **19**, 55–60 (2018).
17. Dolcos, F. et al. Emerging directions in emotional episodic memory. *Front. Psychol.* **8**, 1867 (2017).
18. Phelps, E. A. Human emotion and memory: interactions of the amygdala and hippocampal complex. *Curr. Opin. Neurobiol.* **14**, 198–202 (2004).
19. Dolcos, F., LaBar, K. S. & Cabeza, R. Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron* **42**, 855–863 (2004).

20. Hamann, S. B., Ely, T. D., Grafton, S. T. & Kilts, C. D. Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nat. Neurosci.* **2**, 289–293 (1999).
21. Murty, V. P., Ritchey, M., Adcock, R. A. & LaBar, K. S. fMRI studies of successful emotional memory encoding: a quantitative meta-analysis. *Neuropsychologia* **48**, 3459–3469 (2010).
22. Kensinger, E. A. & Corkin, S. Two routes to emotional memory: distinct neural processes for valence and arousal. *Proc. Natl Acad. Sci. USA* **101**, 3310–3315 (2004).
23. Richardson, M. P., Strange, B. A. & Dolan, R. J. Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nat. Neurosci.* **7**, 278–285 (2004).
24. Fastenrath, M. et al. Dynamic modulation of amygdala–hippocampal connectivity by emotional arousal. *J. Neurosci.* **34**, 13935–13947 (2014).
25. Ritchey, M., Dolcos, F. & Cabeza, R. Role of amygdala connectivity in the persistence of emotional memories over time: an event-related fMRI investigation. *Cereb. Cortex* **18**, 2494–2504 (2008).
26. Costa, M. et al. Aversive memory formation in humans involves an amygdala–hippocampus phase code. *Nat. Commun.* **13**, 6403 (2022).
27. Qasim, S. E., Mohan, U. R., Stein, J. M. & Jacobs, J. Neuronal activity in the human amygdala and hippocampus enhances emotional memory encoding. *Nat. Hum. Behav.* **7**, 754–764 (2023).
28. van Stegeren, A. H. et al. Noradrenaline mediates amygdala activation in men and women during encoding of emotional material. *Neuroimage* **24**, 898–909 (2005).
29. Strange, B. A. & Dolan, R. J. Adrenergic modulation of emotional memory-evoked human amygdala and hippocampal responses. *Proc. Natl Acad. Sci. USA* **101**, 11454–11458 (2004).
30. Strange, B. A., Hurlemann, R. & Dolan, R. J. An emotion-induced retrograde amnesia in humans is amygdala- and β-adrenergic-dependent. *Proc. Natl Acad. Sci. USA* **100**, 13626–13631 (2003).
31. Cahill, L., Prins, B., Weber, M. & McGaugh, J. L. β-Adrenergic activation and memory for emotional events. *Nature* **371**, 702–704 (1994).
32. Hermans, E. J. et al. How the amygdala affects emotional memory by altering brain network properties. *Neurobiol. Learn. Mem.* **112**, 2–16 (2014).
33. Shine, J. M. et al. The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron* **92**, 544–554 (2016).
34. Braun, U. et al. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc. Natl Acad. Sci. USA* **112**, 11678–11683 (2015).
35. Cohen, J. R. & D'Esposito, M. The segregation and integration of distinct brain networks and their relationship to cognition. *J. Neurosci.* **36**, 12083–12094 (2016).
36. Wang, R. et al. Segregation, integration, and balance of large-scale resting brain networks configure different cognitive abilities. *Proc. Natl Acad. Sci. USA* **118**, e2022288118 (2021).
37. Nau, M., Schmid, A. C., Kaplan, S. M., Baker, C. I. & Kravitz, D. J. Centering cognitive neuroscience on task demands and generalization. *Nat. Neurosci*. https://doi.org/10.1038/s41593-024-01711-6 (2024).
38. Song, H., Park, B., Park, H. & Shim, W. M. Cognitive and neural state dynamics of narrative comprehension. *J. Neurosci.* **41**, 8972–8990 (2021).
39. Pessoa, L. On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* **9**, 148–158 (2008).
40. Barrett, L. F. & Satpute, A. B. Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Curr. Opin. Neurobiol.* **23**, 361–372 (2013).
41. Hermans, E. J., Henckens, M. J. A. G., Joëls, M. & Fernández, G. Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends Neurosci.* **37**, 304–314 (2014).
42. Young, C. B. et al. Dynamic shifts in large-scale brain network balance as a function of arousal. *J. Neurosci.* **37**, 281–290 (2017).
43. Lee, K. et al. Arousal impacts distributed hubs modulating the integration of brain functional connectivity. *Neuroimage* **258**, 119364 (2022).
44. Shine, J. M. Neuromodulatory influences on integration and segregation in the brain. *Trends Cogn. Sci.* **23**, 572–583 (2019).
45. Zerbi, V. et al. Rapid reconfiguration of the functional connectome after chemogenetic locus coeruleus activation. *Neuron* **103**, 702–718.e5 (2019).
46. Hermans, E. J. et al. Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science* **334**, 1151–1153 (2011).
47. Kinnison, J., Padmala, S., Choi, J.-M. & Pessoa, L. Network analysis reveals increased integration during emotional and motivational processing. *J. Neurosci.* **32**, 8361–8372 (2012).
48. Shine, J. M., van den Brink, R. L., Hernaus, D., Nieuwenhuis, S. & Poldrack, R. A. Catecholaminergic manipulation alters dynamic network topology across cognitive states. *Netw. Neurosci.* **2**, 381–396 (2018).
49. Keerativittayayut, R., Aoki, R., Sarabi, M. T., Jimura, K. & Nakahara, K. Large-scale network integration in the human brain tracks temporal fluctuations in memory encoding performance. *Elife* **7**, e32696 (2018).
50. Westphal, A. J., Wang, S. & Rissman, J. Episodic memory retrieval benefits from a less modular brain network organization. *J. Neurosci.* **37**, 3523–3531 (2017).
51. Joshi, S., Li, Y., Kalwani, R. M. & Gold, J. I. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* **89**, 221–234 (2016).
52. Lee, H., Bellana, B. & Chen, J. What can narratives tell us about the neural bases of human memory? *Curr. Opin. Behav. Sci.* **32**, 111–119 (2020).
53. Lee, H. & Chen, J. Predicting memory from the network structure of naturalistic events. *Nat. Commun.* **13**, 4235 (2022).
54. Chen, J. et al. Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).
55. Cer, D. et al. Universal Sentence Encoder. Preprint at https://arxiv.org/abs/1803.11175 (2018).
56. Heusser, A. C., Fitzpatrick, P. C. & Manning, J. R. Geometric models reveal behavioural and neural signatures of transforming experiences into memories. *Nat. Hum. Behav.* **5**, 905–919 (2021).
57. Song, H., Finn, E. S. & Rosenberg, M. D. Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proc. Natl Acad. Sci. USA* **118**, e2021905118 (2021).
58. Musz, E. & Chen, J. Neural signatures associated with temporal compression in the verbal retelling of past events. *Commun. Biol.* **5**, 489 (2022).
59. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
60. Guimerà, R. & Nunes Amaral, L. A. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
61. Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N. & Petersen, S. E. Evidence for hubs in human functional brain networks. *Neuron* **79**, 798–813 (2013).
62. Meet Stable Beluga 1 and Stable Beluga 2, our large and mighty instruction fine-tuned language models. *Stability AI* https://stability.ai/blog/stable-beluga-large-instruction-fine-tuned-models (2023).
63. Bradley, M. M. & Lang, P. J. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings* Technical Report C-1 (Center for Research in Psychophysiology, University of Florida, 1999).

64. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. *International Affective Picture System (IAPS): Technical Manual and Affective Ratings* (NIMH Center for the Study of Emotion and Attention, 1997).
65. Kruschke, J. K. Bayesian analysis reporting guidelines. *Nat. Hum. Behav.* **5**, 1282–1291 (2021).
66. Achard, S. & Bullmore, E. Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* **3**, e17 (2007).
67. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* **7**, 40–44 (2008).
68. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 667–685 (2019).
69. Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A. & Constable, R. T. Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nat. Commun.* **9**, 2043 (2018).
70. Howard, G. S. Culture tales: a narrative approach to thinking, cross-cultural psychology, and psychotherapy. *Am. Psychol.* **46**, 187–197 (1991).
71. Mar, R. A. The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol.* **62**, 103–134 (2011).
72. Ross, L. A., Molholm, S., Butler, J. S., Bene, V. A. D. & Foxe, J. J. Neural correlates of multisensory enhancement in audiovisual narrative speech perception: a fMRI investigation. *Neuroimage* **263**, 119598 (2022).
73. Masís-Obando, R., Norman, K. A. & Baldassano, C. Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *Elife* **11**, e70445 (2022).
74. Soares, A. D. et al. Top-down attention shifts behavioral and neural event boundaries in narratives with overlapping event scripts. *Curr. Biol.* **34**, 4729–4742.e5 (2024).
75. Bertolero, M. A., Yeo, B. T. T. & D'Esposito, M. The modular and integrative functional architecture of the human brain. *Proc. Natl Acad. Sci. USA* **112**, E6798–E6807 (2015).
76. Sporns, O. Network attributes for segregation and integration in the human brain. *Curr. Opin. Neurobiol.* **23**, 162–171 (2013).
77. Raut, R. V. et al. Global waves synchronize the brain's functional systems with fluctuating arousal. *Sci. Adv.* **7**, e70445 (2021).
78. Uddin, L. Q., Yeo, B. T. T. & Spreng, R. N. Towards a universal taxonomy of macro-scale functional human brain networks. *Brain Topogr.* **32**, 926–942 (2019).
79. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 201–215 (2002).
80. Yeshurun, Y., Nguyen, M. & Hasson, U. The default mode network: where the idiosyncratic self meets the shared social world. *Nat. Rev. Neurosci.* **22**, 181–192 (2021).
81. Nau, M. et al. Neural and behavioral reinstatement jointly reflect retrieval of narrative events. *Nat. Commun.* **16**, 7865 (2025).
82. Tucker, D. M. & Luu, P. Adaptive control of functional connectivity: dorsal and ventral limbic divisions regulate the dorsal and ventral neocortical networks. *Cereb. Cortex* **33**, 7870–7895 (2023).
83. Talmi, D. & McGarry, L. M. Accounting for immediate emotional memory enhancement. *J. Mem. Lang.* **66**, 93–108 (2012).
84. Shine, J. M., Aburn, M. J., Breakspear, M. & Poldrack, R. A. The modulation of neural gain facilitates a transition between functional segregation and integration in the brain. *Elife* **7**, e31130 (2018).
85. Unsworth, N. & Robison, M. K. A locus coeruleus–norepinephrine account of individual differences in working memory capacity and attention control. *Psychon. Bull. Rev.* **24**, 1282–1311 (2017).
86. Sara, S. J. The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev. Neurosci.* **10**, 211–223 (2009).
87. Sabat, M., de Dampierre, C. & Tallon-Baudry, C. Evidence for domain-general arousal from semantic and neuroimaging meta-analyses reconciles opposing views on arousal. *Proc. Natl Acad. Sci. USA* **122**, e2413808122 (2025).
88. Saalasti, S. et al. Inferior parietal lobule and early visual areas support elicitation of individualized meanings during narrative listening. *Brain Behav.* **9**, e01288 (2019).
89. Ross, M. C. & Cisler, J. M. Altered large-scale functional brain organization in posttraumatic stress disorder: a comprehensive review of univariate and network-level neurocircuitry models of PTSD. *NeuroImage Clin.* **27**, 102319 (2020).
90. Lee, T.-H. et al. Arousal increases neural gain via the locus coeruleus–noradrenaline system in younger adults but not in older adults. *Nat. Hum. Behav.* **2**, 356–366 (2018).
91. Reagh, Z. M., Delarazan, A. I., Garber, A. & Ranganath, C. Aging alters neural activity at event boundaries in the hippocampus and posterior medial network. *Nat. Commun.* **11**, 3980 (2020).
92. van Kesteren, M. T. R., Fernández, G., Norris, D. G. & Hermans, E. J. Persistent schema-dependent hippocampal–neocortical connectivity during memory encoding and postencoding rest in humans. *Proc. Natl Acad. Sci. USA* **107**, 7550–7555 (2010).
93. Raykov, P. P., Keidel, J. L., Oakhill, J. & Bird, C. M. Activation of person knowledge in medial prefrontal cortex during the encoding of new lifelike events. *Cereb. Cortex* **31**, 3494–3505 (2021).
94. Bower, G. H., Black, J. B. & Turner, T. J. Scripts in memory for text. *Cogn. Psychol.* **11**, 177–220 (1979).
95. Reagh, Z. M. & Ranganath, C. Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nat. Commun.* **14**, 1279 (2023).
96. Omanson, R. C. The relation between centrality and story category variation. *J. Verbal Learn. Verbal Behav.* **21**, 326–337 (1982).
97. Trabasso, T. & van den Broek, P. Causal thinking and the representation of narrative events. *J. Mem. Lang.* **24**, 612–630 (1985).
98. Mather, M. & Sutherland, M. R. Arousal-biased competition in perception and memory. *Perspect. Psychol. Sci.* **6**, 114–133 (2011).
99. Vuilleumier, P. How brains beware: neural mechanisms of emotional attention. *Trends Cogn. Sci.* **9**, 585–594 (2005).
100. Leong, Y. C., Dziembaj, R. & D'Esposito, M. Pupil-linked arousal biases evidence accumulation toward desirable percepts during perceptual decision-making. *Psychol. Sci.* **32**, 1494–1509 (2021).
101. Krishnamurthy, K., Nassar, M. R., Sarode, S. & Gold, J. I. Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nat. Hum. Behav.* **1**, 0107 (2017).
102. Wichary, S., Mata, R. & Rieskamp, J. Probabilistic inferences under emotional stress: how arousal affects decision processes. *J. Behav. Decis. Mak.* **29**, 525–538 (2016).
103. Lempert, K. M., Johnson, E. & Phelps, E. A. Emotional arousal predicts intertemporal choice. *Emotion* **16**, 647–656 (2016).
104. Lee, H. & Chen, J. FilmFestival. *OpenNeuro* https://openneuro.org/datasets/ds004042/versions/1.0.1 (2022).
105. Chen, J. Sherlock movie watching dataset. *Princeton DataSpace* https://dataspace.princeton.edu/handle/88435/dsp01nz8062179 (2016).
106. Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A. & Constable, R. T. ParanoiaStory. Open Neuro https://openneuro.org/datasets/ds001338/versions/1.0.0 (2018).
107. Sava-Segal, C., Richards, C., Leung, M. & Finn, E. S. Individual differences in neural event segmentation of continuous experiences. *Cereb. Cortex* **33**, 8164–8178 (2023).
108. Demszky, D. et al. Using large language models in psychology. *Nat. Rev. Psychol.* **2**, 688–701 (2023).

109. Yang, X., O'Reilly, C. & Shinkareva, S. V. Embracing naturalistic paradigms: substituting GPT predictions for human judgments. Preprint at *bioRxiv* https://doi.org/10.1101/2024.06.17.599327 (2024).
110. Rathje, S. et al. GPT is an effective tool for multilingual psychological text analysis. *Proc. Natl Acad. Sci. USA* **121**, e2308950121 (2024).
111. Schaefer, A. et al. Local–global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
112. Tian, Y., Margulies, D. S., Breakspear, M. & Zalesky, A. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat. Neurosci.* **23**, 1421–1432 (2020).
113. Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B. & Bandettini, P. A. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage* **44**, 893–905 (2009).
114. Saad, Z. S. et al. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. *Brain Connect.* **2**, 25–32 (2012).
115. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
116. R Core Team *R: A Language and Environment for Statistical Computing* (R Core Team, 2023).
117. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Chapman and Hall/CRC, 2016); https://doi.org/10.1201/9781315372495.
118. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput*. https://doi.org/10.1007/s11222-016-9696-4 (2017).
119. Makowski, D. et al. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian Framework. *J. Open Source Softw.* https://doi.org/10.21105/joss.01541 (2019).
120. Imai, K., Keele, L. & Tingley, D. A general approach to causal mediation analysis. *Psychol. Methods* **15**, 309–334 (2010).
121. Storey, J. D. The positive false discovery rate: a Bayesian interpretation and the *q*-value. *Ann. Stat.* **31**, 2013–2035 (2003).
122. Peirce, J. et al. PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* https://doi.org/10.3758/s13428-018-01193-y (2019).

Acknowledgements

We thank M. Rosenberg and D. Gallo for helpful discussions on the work, and J. Chen and H. Lee for sharing the Film Festival and Sherlock datasets with accompanying annotations. We thank E. Finn for sharing the Paranoia dataset. This research was supported by resources provided by the University of Chicago Social Sciences Division, the University of Chicago Research Computing Center and the University of Chicago Data Science Institute. The authors received no specific funding for this work.

Author contributions

J.S.P. and Y.C.L. conceptualized the analysis. J.S.P., K.G. and J.K. analysed the data. J.S.P., M.N., I.P. and Y.C.L. wrote the original draft of the manuscript. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-025-02315-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-025-02315-1.

Correspondence and requests for materials should be addressed to Jadyn S. Park or Yuan Chang Leong.

Peer review information *Nature Human Behaviour* thanks James Shine, Sara Sorella and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

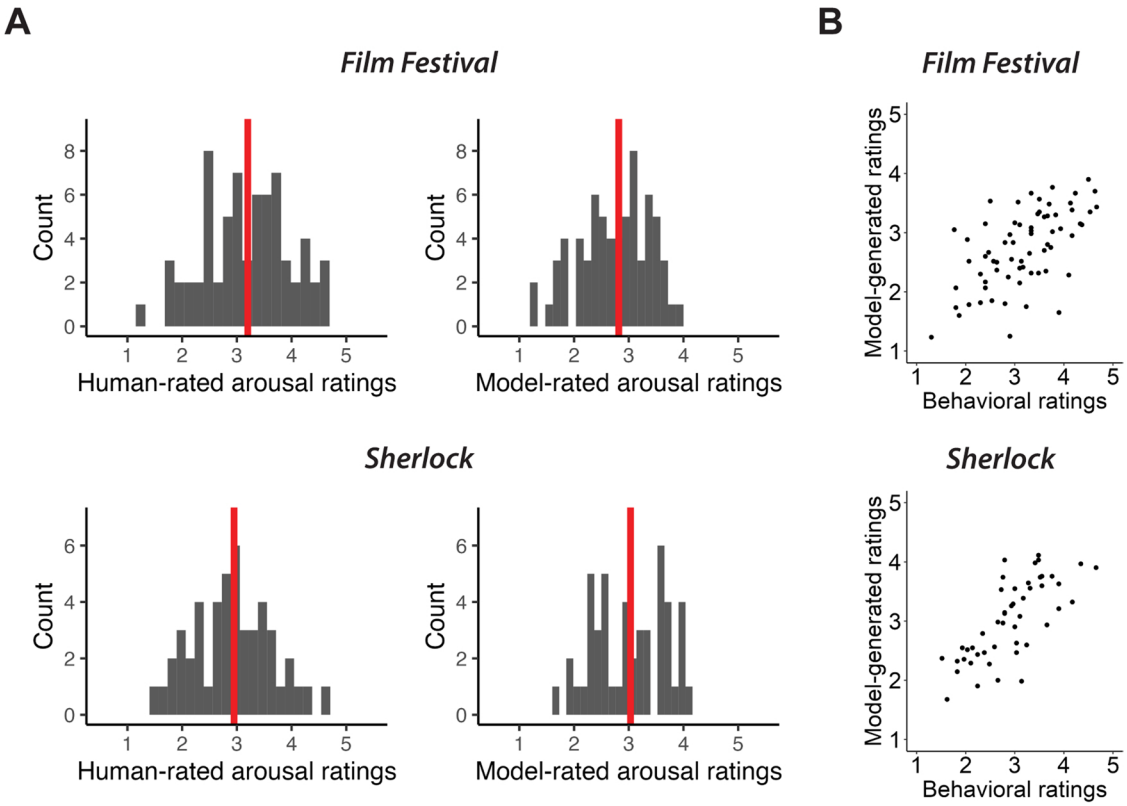
Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

https://doi.org/10.1038/s41562-025-02315-1

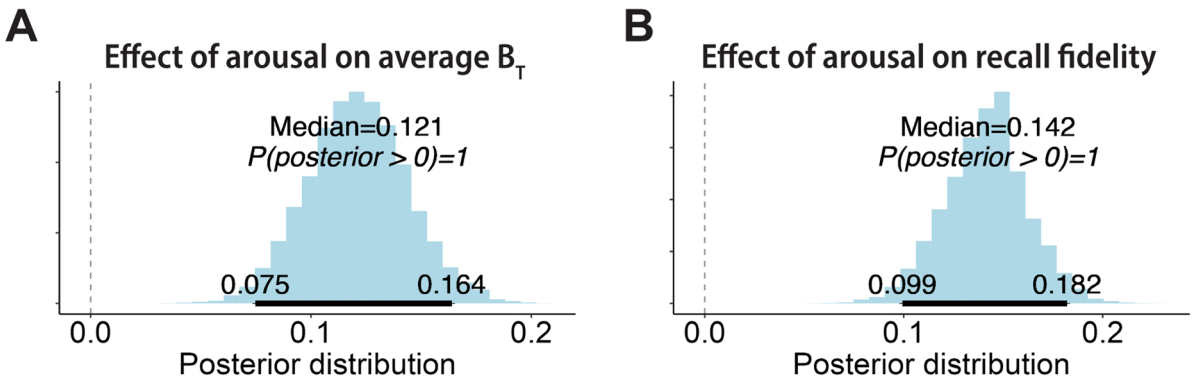
Extended Data Fig. 1 | The semantic similarity between annotations of a movie event and participants' recall transcripts provide a measure of recall fidelity. Excerpts from the annotations of an example event in each dataset are shown with transcripts and fidelity scores of two different participants recalling that event. Colors indicate matching details between the movie and recall, hand-labeled for illustrative purposes.

https://doi.org/10.1038/s41562-025-02315-1



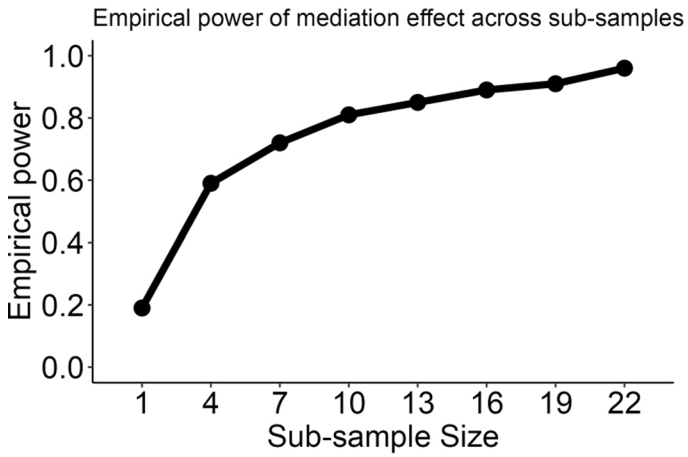
Extended Data Fig. 2 | LLM-generated arousal ratings align with human behavioral ratings. (**A**) Histogram of human- and model-rated arousal ratings. Arousal ratings for each event were averaged across subjects and iterations (LLM ratings). The red vertical line represents the median. (**B**) Scatterplot of model-arousal ratings against behavioral arousal ratings.

https://doi.org/10.1038/s41562-025-02315-1



Extended Data Fig. 3 | Arousal is associated with functional integration and recall fidelity. Posterior distribution of regression coefficient when predicting (**A**) average participation coefficient (BT), and (**B**) recall fidelity, estimated by a Bayesian multilevel model that pooled across Film Festival and Sherlock. The 95% HDI of each distribution is indicated by the bold horizontal line.

https://doi.org/10.1038/s41562-025-02315-1



Extended Data Fig. 4 | Post-hoc robustness analysis of the mediation effect in the Paranoia dataset. We repeated the mediation analysis with randomly drawn subsets of fMRI participants (with replacement), varying the sub-sample size from 1 to 22. For each subsample size, we conducted 100 iterations and computed empirical power, defined as the proportion of iterations in which more than 95% of the posterior distribution of the mediation effect was greater than zero. Power exceeded 0.8 with as few as 10 participants and approached 1.0 at the full sample size, indicating high reliability of the mediation effect.

https://doi.org/10.1038/s41562-025-02315-1

Extended Data Table 1 | Table reporting mediation effect of within- and between-network efficiency on the relationship between emotional arousal and recall fidelity for the audiovisual datasets

Network 1	Network 2	Median	95% HDI	p(b>0)	Adjusted PEP	WAIC
CON	**CON**	**0.00**	**[0, 0.01]**	**0.98**	**0.00**	**4,940**
CON	**DMN**	**0.01**	**[0, 0.01]**	**1.00**	**0.00**	**4,937**
CON	SUB	0.00	[0, 0]	0.69	0.05	4,957
DAN	**CON**	**0.01**	**[0.01, 0.02]**	**1.00**	**0.00**	**4,944**
DAN	**DAN**	**0.03**	**[0.02, 0.04]**	**1.00**	**0.00**	**4,924**
DAN	DMN	0.00	[0, 0]	0.92	0.01	4,954
DAN	LIM	0.00	[0, 0]	0.81	0.02	4,956
DAN	SUB	0.00	[0, 0.01]	0.90	0.01	4,955
DAN	**VAN**	**0.01**	**[0, 0.01]**	**1.00**	**0.00**	**4,947**
DMN	**DMN**	**0.01**	**[0.01, 0.02]**	**1.00**	**0.00**	**4,927**
DMN	**SUB**	**0.00**	**[0, 0.01]**	**0.99**	**0.00**	**4,953**
LIM	**CON**	**0.01**	**[0, 0.01]**	**1.00**	**0.00**	**4,945**
LIM	**DMN**	**0.02**	**[0.01, 0.02]**	**1.00**	**0.00**	**4,925**
LIM	**LIM**	**0.02**	**[0.01, 0.03]**	**1.00**	**0.00**	**4,917**
LIM	**SUB**	**0.01**	**[0, 0.02]**	**1.00**	**0.00**	**4,941**
MOT	**CON**	**0.01**	**[0, 0.01]**	**0.98**	**0.00**	**4,953**
MOT	DAN	0.00	[0, 0.01]	0.98	0.00	4,950
MOT	**DMN**	**0.00**	**[0, 0.01]**	**0.99**	**0.00**	**4,951**
MOT	LIM	0.00	[0, 0.01]	0.88	0.02	4,956
MOT	**MOT**	**0.00**	**[0, 0.01]**	**1.00**	**0.00**	**4,949**
MOT	SUB	0.00	[0, 0]	0.76	0.04	4,957
MOT	**VAN**	**0.01**	**[0, 0.01]**	**0.99**	**0.00**	**4,952**
SUB	**SUB**	**0.01**	**[0, 0.02]**	**1.00**	**0.00**	**4,945**
VAN	**CON**	**0.01**	**[0, 0.02]**	**1.00**	**0.00**	**4,941**
VAN	**DMN**	**0.01**	**[0.01, 0.02]**	**1.00**	**0.00**	**4,929**
VAN	**LIM**	**0.01**	**[0, 0.01]**	**0.99**	**0.00**	**4,953**
VAN	SUB	0.00	[0, 0.01]	0.94	0.01	4,956
VAN	**VAN**	**0.01**	**[0, 0.02]**	**1.00**	**0.00**	**4,948**
VIS	CON	0.00	[0, 0]	0.52	0.07	4,958
VIS	**DAN**	**0.01**	**[0.01, 0.02]**	**1.00**	**0.00**	**4,933**
VIS	DMN	0.00	[0, 0]	0.55	0.06	4,958
VIS	LIM	-0.00	[0, 0]	0.45	0.08	4,957
VIS	MOT	0.00	[0, 0]	0.80	0.03	4,953
VIS	SUB	0.00	[0, 0.01]	0.94	0.01	4,953
VIS	VAN	-0.00	[0, 0]	0.44	0.10	4,958
VIS	**VIS**	**0.01**	**[0.01, 0.02]**	**1.00**	**0.00**	**4,921**

Rows in bold font indicate networks and network pairs where the expected PEP was <0.01 and the 95% HDI were >0. p(b>0): proportion of the posterior draws that were >0. WAIC: Widely Applicable Information Criterion.

https://doi.org/10.1038/s41562-025-02315-1

Extended Data Table 2 | Table reporting mediation effect of within- and between-network efficiency on the relationship between pupil dilation and recall fidelity for the Paranoia dataset

Network 1	Network 2	Median	95% HDI	p(b>0)	Corrected PEP	WAIC
CON	**CON**	**0.03**	**[0.01, 0.04]**	**1.00**	**0.00**	**1,783**
CON	**DMN**	**0.09**	**[0.06, 0.12]**	**1.00**	**0.00**	**1,712**
CON	SUB	-0.00	[-0.02, 0.02]	0.37	0.13	1,797
DAN	**CON**	**0.09**	**[0.06, 0.13]**	**1.00**	**0.00**	**1,757**
DAN	DAN	0.01	[0, 0.02]	0.90	0.01	1,795
DAN	**DMN**	**0.07**	**[0.04, 0.1]**	**1.00**	**0.00**	**1,774**
DAN	**LIM**	**0.02**	**[0, 0.04]**	**0.98**	**0.00**	**1,792**
DAN	SUB	-0.01	[-0.03, 0]	0.00	0.39	1,788
DAN	VAN	-0.01	[-0.03, 0]	0.05	0.29	1,794
DMN	**DMN**	**0.03**	**[0, 0.06]**	**0.99**	**0.00**	**1,717**
DMN	SUB	-0.01	[-0.03, 0]	0.00	0.37	1,790
LIM	**CON**	**0.05**	**[0.03, 0.08]**	**1.00**	**0.00**	**1,713**
LIM	DMN	-0.00	[-0.01, 0.01]	0.43	0.09	1,779
LIM	**LIM**	**0.01**	**[0, 0.03]**	**0.99**	**0.00**	**1,786**
LIM	SUB	0.00	[-0.01, 0.02]	0.66	0.07	1,797
MOT	**CON**	**0.08**	**[0.05, 0.11]**	**1.00**	**0.00**	**1,764**
MOT	DAN	0.00	[-0.01, 0.01]	0.72	0.04	1,788
MOT	DMN	-0.01	[-0.03, 0.02]	0.33	0.15	1,797
MOT	LIM	-0.00	[-0.01, 0]	0.18	0.20	1,795
MOT	MOT	-0.01	[-0.02, 0]	0.04	0.33	1,787
MOT	SUB	-0.01	[-0.03, 0.01]	0.11	0.22	1,768
MOT	VAN	0.01	[0, 0.02]	0.88	0.01	1,795
SUB	SUB	0.00	[-0.01, 0.02]	0.72	0.05	1,796
VAN	**CON**	**0.09**	**[0.06, 0.13]**	**1.00**	**0.00**	**1,741**
VAN	**DMN**	**0.06**	**[0.04, 0.09]**	**1.00**	**0.00**	**1,752**
VAN	**LIM**	**0.06**	**[0.04, 0.09]**	**1.00**	**0.00**	**1,745**
VAN	SUB	-0.00	[-0.02, 0.02]	0.42	0.11	1,797
VAN	**VAN**	**0.04**	**[0.02, 0.07]**	**1.00**	**0.00**	**1,768**
VIS	CON	-0.01	[-0.02, 0]	0.06	0.27	1,794
VIS	**DAN**	**0.03**	**[0.01, 0.05]**	**1.00**	**0.00**	**1,782**
VIS	DMN	0.00	[-0.01, 0.02]	0.74	0.03	1,797
VIS	LIM	-0.01	[-0.02, 0]	0.04	0.32	1,794
VIS	MOT	-0.00	[-0.01, 0.01]	0.28	0.18	1,792
VIS	SUB	-0.01	[-0.03, 0]	0.06	0.25	1,794
VIS	VAN	-0.01	[-0.02, 0]	0.03	0.35	1,791
VIS	**VIS**	**0.05**	**[0.03, 0.08]**	**1.00**	**0.00**	**1,767**

Rows in bold font indicate the networks and network pairs where the expected PEP was <0.01 and the 95% HDI were >0. p(b>0): proportion of the posterior draws that were >0. WAIC: Widely Applicable Information Criterion.

nature portfolio

Corresponding author(s): Yuan Chang Leong

Last updated by author(s): Aug 19, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Psychopy (version 2022.2.5) was used to collect behavioral and pupillometry data.
Data analysis	The preprocessing of fMRI data were done using FSL/FEAT v6.00 https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) and custom bash and MATLAB (R2022b) scripts. For graph theoretical metrics, we used custom MATLAB scripts (https://github.com/jadynpark/arousal_integration/blob/main/sherlock/scripts/1_fmri/1_bct/c_participation_coefficient.m). The functions used in the script are from the Brain Connectivity Toolbox (BCT; https://sites.google.com/site/bctnet/). For creating sentence vector embeddings, we used Google's Universal Sentence Encoder (https://www.tensorflow.org/hub/tutorials/semantic_similarity_with_tf_hub_universal_encoder). For calculating recall fidelity, we used a custom Python script (version 3.9). For statistical analysis, we used the brms package (version 2.21.0) in R version 4.3.2. The models were run with the random seed "123" to ensure reproducibility. Model-generated arousal ratings were created using StableBeluga-13B, an open-source Large Language Model.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw MRI data from the Film Festival dataset are available on OpenNeuro, https://openneuro.org/datasets/ds004042/versions/1.0.1.
Preprocessed MRI data from the Sherlock dataset are hosted at DataSpace, https://dataspace.princeton.edu/jspui/handle/88435/dsp01nz8062179.
Raw MRI data from the Paranoia dataset are available on OpenNeuro, https://openneuro.org/datasets/ds001338/versions/1.0.0
Behavioral and pupillometry data are available at: https://github.com/jadynpark/arousal_integration

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Fifteen participants (10 female, 5 male; ages 21 to 33, mean age=27.5) were included in the Film Festival dataset. Seventeen participants (7 female, 10 male; ages 19 to 26, mean age=20.8) were included in the Sherlock dataset. Twenty-two participants (11 female, 11 male; ages 19 to 35, mean age=27.0) were included in the Paranoia fMRI dataset. For the behavioral arousal experiment, thirty participants (5 male, 24 female, 1 non-binary; ages 21 to 36, mean age=26.23) were included. For the pupillometry experiment, 27 participants were included in the analysis (12 male, 15 female; ages 20 to 33, mean age=22.3).
Reporting on race, ethnicity, or other socially relevant groupings	For the behavioral arousal data that was collected by our group, 56% were Asian, 36% were White, 6% were more than one race. 6% were Hispanic or Latino. For the pupillometry study, of the participants that were included in the analysis, 26% were Asian, 30% were White, 22% were Hispanic or Latino, 11% were Black or African-American, and 11% were more than one race.
Population characteristics	All fMRI participants were right-handed, speak fluent English, and reported normal hearing or corrected-to-normal vision. Participants were primarily college-aged adults participating in the study for either course credit or monetary compensation. We do not believe that there would be self-selection effects. However, as participants were primarily young adults future work is needed to assess how these findings extend to older population.
Recruitment	For the Film Festival and Sherlock datasets, participants were recruited from the Princeton community and compensated at $20/hour. For the Paranoia dataset, participants were recruited from the Yale community and compensated $75 for the MRI session. For all behavioral experiments, participants were recruited from the University of Chicago SONA Research Participant Pool. Participants provided informed consent prior to the start of all studies.
Ethics oversight	The behavioral experiments were conducted in accordance with the protocols approved by the University of Chicago Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	15 participants were included in the Film Festival dataset; 17 were in Sherlock, and 22 were in Paranoia. As these are existing datasets, sample size was pre-determined based on available data. Results replicated independently across Film Festival and Sherlock, providing evidence of the robustness of the results from the pooled sample. A post-hoc robustness analysis indicated that power exceeded 0.8 after 10 participants, and approached 1 at the full sample size of n = 22 in the Paranoia dataset. For the behavioral arousal experiment, 30 participants were recruited. This sample size was based on our previous work (Ke et al., PLOS Comp Biol, 2025) where we found that the reliability of arousal ratings reached a plateau by this sample size. For the pupillometry study, 35 participants were recruited. We show in the manuscript a reliable pupil timecourse across participants at this sample size.
Data exclusions	From the datasets that we collected, no exclusions were made for the behavioral arousal rating study. For the pupillometry study, three

Data exclusions	participants were excluded due to equipment failure, and five additional participants were excluded due to excessive noise in the data. In total, 27 participants were included in the analyses.
Replication	We found that emotionally arousing events were associated with greater integration across functional modules during encoding. Such events, in turn, were more likely to be recalled subsequently. We demonstrate this finding in data pooled across the two audiovisual datasets (Film Festival and Sherlock) as well as the audio-only dataset (Paranoia). The effects were replicated when tested in three datasets independently.
Randomization	Randomization was not performed, as this was an observational study and no experimental manipulations were involved.
Blinding	Blinding was not necessary as our procedures did not involve explicit experimental manipulation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

Magnetic resonance imaging

Experimental design

Design type	The studies used naturalistic movie watching/story listening paradigms.
Design specifications	Participants watched audiovisual movie clips, or listened to an audio narrative, in the MRI scanner. The audiovisual movie clips (Sherlock, Film Festival) were 50-minutes long. The audio story (Paranoia) was 22 minutes long.
Behavioral performance measures	Participants were asked to recall the movie clips/story in the MRI scanner after encoding. For the purpose of the current study, the accompanying neural data were not used.

Acquisition

Imaging type(s)	Functional and structural
Field strength	3T
Sequence & imaging parameters	Structural images were acquired using a high-resolution T1-weighted MPRAGE sequence (1mm3 resolution). Functional BOLD images were collected on a 3T Philips Ingenia
Area of acquisition	Whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	No preprocessing was performed for Sherlock, as the data have already been preprocessed. We used FSL/FEAT v6.00 (FMRIB software library, FMRIB, Oxford, UK) for the preprocessing of functional and structural images for the Film Festival and the Paranoia dataset.
Normalization	The functional images were normalized using affine transformation (12 dof).
Normalization template	The functional images were registered to the MNI 152 template.
Noise and artifact removal	The preprocessing steps include MCFLIRT motion correction, high-pass filtering of the data with a 100-ms cutoff, and spatial smoothing using a Gaussian kernel with a full-width at half-maximum (FWHM) at 5mm.
Volume censoring	Volume censoring was not applied.

Statistical modeling & inference

Model type and settings	We tested which functional networks and network pairs were involved in the mediation effects of arousal on memory. As a result, 36 tests were performed. To account for multiple comparisons across networks and network pairs, we thresholded the results within each dataset at an expected Posterior Error Probability (PEP) of 0.01, corresponding to a less than 0.01 probability of making an incorrect inference. From a frequentist perspective, this procedure is analogous to controlling for a false discovery rate at $q < 0.01$.
Effect(s) tested	We tested which functional networks and network pairs were involved in the mediation effects of arousal on memory.

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	Graph theoretic analyses performed at the whole-brain level, which only involved a single metric. Thus, multiple comparisons was not necessary. Network-based analyses were corrected for multiple comparisons using the Posterior Error Probability. No voxel-wise or cluster-level inference was performed.
Correction	To account for multiple comparisons across networks and network pairs, we thresholded the results within each dataset at an expected Posterior Error Probability (PEP) of 0.01.

Models & analysis

n/a	Involved in the study
☐	☒ Functional and/or effective connectivity
☐	☒ Graph analysis
☒	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	For all three fMRI datasets, we first extracted Blood Oxygen Level Dependent (BOLD) signals. The 200-ROI Schaefer atlas was used for cortical parcellations, and the 16-ROI Melbourne subcortical atlas for subcortex parcellations. In total, there were 216 ROIs. To ensure the robustness of our results, we replicated our methods using a different parcellation scheme (Shen atlas; Shen et al., 2013). We averaged the BOLD time courses of all voxes in each ROI. For each dataset, the time courses were segmented into events, where the event boundaries depicted changes in time, location, and/or narrative of the story. Following previous study led by the original authors of the dataset, Sherlock had 48 events, Film Festival had 68, and Paranoia had 24. Functional connectivity matrices were created for each event.
Graph analysis	For each subject and each event, the functional connectivity matrices were thresholded such that only the strongest 15 percent of functional connections were retained. To ensure robustness, we replicated the results using different threshold values (top 10, 20, and 25 percent). The thresholded matrices were then binarized. Once the graphs were constructed, different metrics were extracted to assess their properties. We used the Brain Connectivity Toolbox (BCT; https://sites.google.com/site/bctnet/) to calculate participation coefficient, a measure of integration across functional modules. Specifically, for each graph, functional modules were first identified using a Louvain algorithm, which was run 1,000 times due to its stochastic nature. We then calculated the participation coefficient for each ROI, and averaged across ROIs to obtain a measure of average functional integration across the whole brain.