Neural signatures of attentional engagement during narratives and its consequences for event memory

Hayoung Song[a,1] ⓘ, Emily S. Finn[b] ⓘ, and Monica D. Rosenberg[a,c,1] ⓘ

[a]Department of Psychology, University of Chicago, Chicago, IL 60637; [b]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755; and [c]Neuroscience Institute, University of Chicago, Chicago, IL 60637

As we comprehend narratives, our attentional engagement fluctuates over time. Despite theoretical conceptions of narrative engagement as emotion-laden attention, little empirical work has characterized the cognitive and neural processes that comprise subjective engagement in naturalistic contexts or its consequences for memory. Here, we relate fluctuations in narrative engagement to patterns of brain coactivation and test whether neural signatures of engagement predict subsequent memory. In behavioral studies, participants continuously rated how engaged they were as they watched a television episode or listened to a story. Self-reported engagement was synchronized across individuals and driven by the emotional content of the narratives. In functional MRI datasets collected as different individuals watched the same show or listened to the same story, engagement drove neural synchrony, such that default mode network activity was more synchronized across individuals during more engaging moments of the narratives. Furthermore, models based on time-varying functional brain connectivity predicted evolving states of engagement across participants and independent datasets. The functional connections that predicted engagement overlapped with a validated neuromarker of sustained attention and predicted recall of narrative events. Together, our findings characterize the neural signatures of attentional engagement in naturalistic contexts and elucidate relationships among narrative engagement, sustained attention, and event memory.

sustained attention | episodic memory | narrative | functional connectivity | predictive modeling

W e engage with the world and construct memories by attending to information in our external environment (1). However, the degree to which we pay attention waxes and wanes over time (2, 3). Such fluctuations of attention not only influence our ongoing perceptual experience but can also have consequences for what we later remember (4).

Changes in attentional states are typically studied with continuous performance tasks (CPTs), which require participants to respond to rare targets in a constant stream of stimuli or respond to every presented stimulus except the rare target (5–7). Paying attention to taxing CPTs, however, often feels different from paying attention in other everyday situations. For example, when we listen to the radio, watch a television show, or have a conversation with family and friends, sustaining focus can feel comparatively effortless. Psychology research has characterized feelings of effortless attention in other contexts, such as flow states of complete absorption in an activity (8). When comprehending narratives, our attention may be naturally captured by the story, causing us to become engaged in the experience. Narrative engagement has been defined as an experience of being deeply immersed in a story with heightened emotional arousal and attentional focus (9, 10). Building on this theoretical definition, we characterize how subjective engagement fluctuates as narratives unfold and test the hypothesis that engagement scales with a story's emotional content as well as an individual's sustained attentional state.

Functional neuroimaging studies have used naturalistic, narrative stimuli to examine how we perceive (11, 12) and remember (13) structured events based on memory of contexts (14–16), prior knowledge or beliefs (17, 18), and emotional and social reasoning (19–21). However, strikingly few neuroimaging studies have directly probed attention during naturalistic paradigms. Among these, Regev et al. (22) examined how selectively attending to narrative inputs from a particular sensory modality (e.g., auditory) while suppressing the other modality (e.g., visual) enhances stimulus-locked brain responses to the attended inputs. With nonnarrative movies, Çukur et al. (23) found that semantic representations were warped toward attended object categories during visual search. However, both of these studies relied on experimental manipulations of attention to elucidate its relationship to brain activity. Work has also reported that intersubject synchrony of functional MRI (fMRI) and electroencephalography (EEG) activity, a measure of neural reliability, relates to heightened attention to stimuli (24–28) and subsequent memory (29, 30). However, the neural signatures of dynamically changing attentional states in naturalistic contexts and their consequences for event memory remain poorly understood.

Previous work suggests that goal-directed focus and attentional control rely on activity in frontoparietal cortical regions comprising a large-scale attention network (31, 32). Regions of the default mode network (DMN) are thought to activate antagonistically to the attention network (33), showing increased activity during off-task thought and mind wandering while deactivating during external attention (34, 35). Interestingly, however, other work provides seemingly counterintuitive evidence that DMN activity characterizes moments of stable and optimal attentional task performance

Significance

The degree to which we are engaged in narratives fluctuates over time. What drives these changes in engagement, and how do they affect what we remember? Behavioral studies showed that people experienced similar fluctuations in engagement during a television show or an audio-narrated story and were more engaged during emotional moments. Functional MRI experiments revealed that changes in a pattern of functional brain connectivity predicted changes in narrative engagement. This predictive brain network not only was related to a validated neuromarker of sustained attention but also predicted what narrative events people recalled after the MRI scan. Overall, this study empirically characterizes engagement as emotion-laden attention and reveals system-level dynamics underlying real-world attention and memory.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

(7, 36–38) and is centrally involved in narrative comprehension and memory (11–13, 39, 40). Given conceptualizations of frontoparietal regions as task positive and DMN regions as task negative in certain contexts—but DMN as task positive in others—what roles do these networks play when our attention waxes and wanes during real-world narratives?

Beyond the canonical DMN and frontoparietal networks, studies have shown that synchrony between a widely distributed set of brain regions reflects changes in attentional states within and across individuals (41, 42). Recent literature showed that functional connectivity (FC), a statistical measure of neural synchrony between pairwise brain regions, predicts attentional state changes during task performance (43, 44). Since these studies characterized attentional states in controlled task conditions, we further ask whether naturalistic attentional engagement is reflected in functional brain connectivity.

The current study characterizes attentional states in real-world settings by tracking subjective engagement during movie watching and story listening. In doing so, we address three primary aims: testing the theoretical conception of engagement as emotion-laden attention, examining how engagement is reflected in large-scale brain dynamics, and elucidating the consequences of engagement during encoding for subsequent memory. We first measured self-reported engagement as behavioral participants watched an episode of television series *Sherlock* or listened to an audio-narrated story, *Paranoia*. Providing empirical support for its theoretical definition, changes in engagement were driven by the emotional contents of narratives and related to fluctuations of a validated FC index of sustained attention during psychological tasks (43). We next related group-average engagement time courses to fMRI activity observed as a separate pool of participants watched *Sherlock* (13) or listened to *Paranoia* (18). Dynamic intersubject correlation analysis (ISC) (45) revealed that activity in large-scale functional networks, especially the DMN, was more synchronized across individuals during more engaging periods of the narratives. Furthermore, using time-resolved predictive modeling (46), we found that patterns of time-resolved FC predicted engagement dynamics and that these same patterns predicted later event recall. Thus, we provide evidence for engagement as emotion-laden sustained attention, elucidate the role of brain network dynamics in engagement, and demonstrate relationships between engagement and episodic memory.
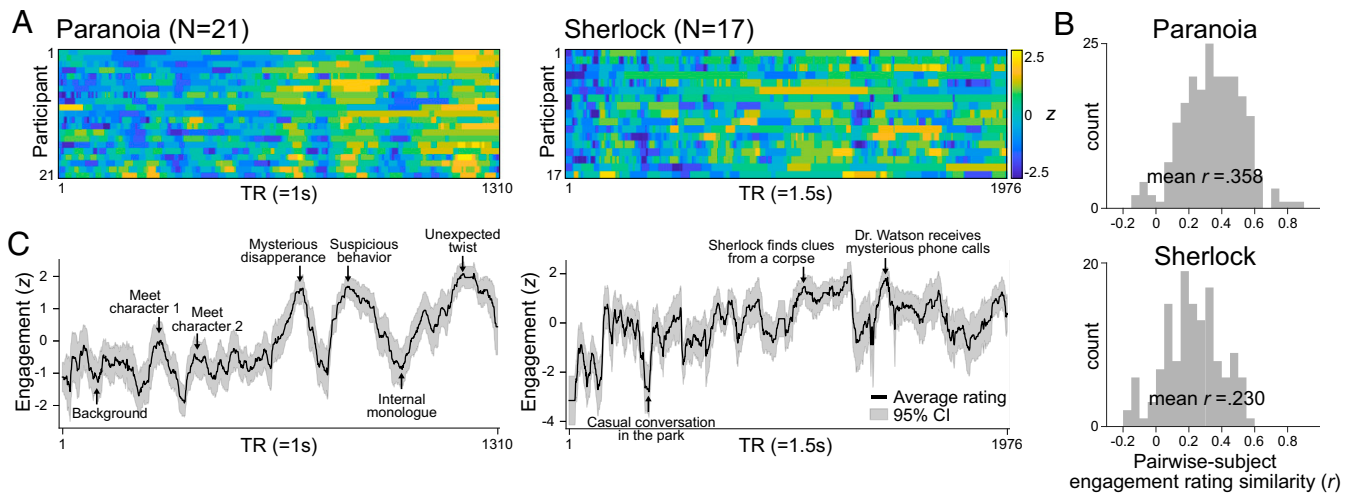
Results

Tracking Dynamic States of Engagement during Movie Watching and Story Listening. We asked how subjective engagement changes over time as individuals comprehend a story and whether changes are synchronized across participants. We measured engagement during two narrative stimuli that were used in previous fMRI research: 1) a 20-min audio-narrated story, *Paranoia* (fMRI dataset $n = 22$) (18), which was intentionally created to induce suspicion surrounding the characters and situations and 2) a 50-min episode of BBC's television series *Sherlock* (fMRI dataset $n = 17$) (13), in which Sherlock and Dr. Watson meet and solve a mysterious crime together. We chose these stimuli because they were both long narratives with complex plots and relationships among characters, and they involved different sensory modalities (one audio only and one audiovisual), which enabled us to investigate high-level cognitive states of engagement that are not specific to sensory modality (47).

In behavioral studies, participants listened to *Paranoia* ($n = 21$) or watched *Sherlock* ($n = 17$) and were instructed to continuously rate how engaging they found the story by adjusting a scale bar from 1 ("Not engaging at all") to 9 ("Completely engaging") using button presses (Fig. 1*A*) (48). The definition of engagement [Table 1; adapted from Busselle and Bilandzic (9)] was explained to participants prior to the task. The scale bar was always visible on the computer monitor so that participants could make continuous adjustments as they experienced changes in subjective engagement.

We first tested whether changes in engagement were similar across participants comprehending the same narratives. Providing initial validation for our self-report task as a measure of stimulus-related narrative engagement, we found significant positive correlations between pairwise engagement time courses (*Paranoia* mean Pearson's $r = 0.358 \pm 0.215$; *Sherlock* $r = 0.230 \pm 0.182$; Fig. 1*B*). Mean r values were computed by averaging Fisher's z–transformed Pearson's correlation coefficients and transforming the mean Fisher's z value back to r. Correlations were significantly positive in 95.24% of *Paranoia* participant pairs and 86.03% of *Sherlock* participant pairs (false discovery rate-corrected P [FDR-P] < 0.05, corrected for number of participant pairs). We averaged all participants' response time courses (Fig. 1*C*) and observed positive linear trends in engagement for both *Paranoia* [$t(1,308) = 40.62$, $r^2 = 0.56$, $P < 0.001$] and *Sherlock* [$t(1,974) = 26.36$, $r^2 = 0.26$, $P < 0.001$], suggesting a gradual increase in engagement as the narrative developed to an end (for detailed behavioral results, see *SI Appendix, Supplementary Text S1*).

Since engagement ratings were similar across participants, we treated the group-average engagement rating (Fig. 1*C*) as a proxy for stimulus-related engagement, common across individuals. We qualitatively assessed moments when participants were, on average, most or least engaged in the narratives (Fig. 1*C*). In *Paranoia*, engagement peaked at moments when a character exhibited suspicious behavior, or when there was an unexpected twist in the story. On the other hand, engagement was generally low when a story setting was being developed or when a protagonist was having an internal thought. In *Sherlock*, engagement peaked at moments when Sherlock was solving a mysterious crime and when events were highly suspenseful. Participants' general engagement decreased during comparatively casual events, less relevant to the crime scenes. The group-averaged engagement time courses were convolved with the hemodynamic response function (HRF) to be applied to a separate pool of individuals who participated in previous fMRI studies.

Engagement Scales with Emotional Arousal of the Narratives. Given that participants' engagement dynamics were time locked to the stimuli, we asked which features of narratives drove changes in engagement. We conducted partial correlations between group-average engagement and four features of the narratives: positive and negative emotional content and auditory and visual sensory information measured at every moment of time (48). We used different ways of extracting the emotional content of the *Paranoia* and *Sherlock* narratives to utilize open resources distributed by the two research groups and assess the robustness of results to a specific extraction approach. For *Paranoia*, emotional arousal was inferred from the occurrence of positive and negative emotion words in the story transcript using Linguistic Inquiry and Word Count text analysis (49) [provided by Finn et al. (18)]. Words indicating positive emotion included accept, adventure, amazingly, and appreciated, and words indicating negative emotion included afraid, alarmed, anxious, and apprehensively (for a full list of emotional words, see *SI Appendix*, Table S1). The frequencies of positive and negative emotion words were counted per sentence, then expanded to match the moments when the sentence was uttered during scans (*SI Appendix*, Fig. S1*A*). Thus, high frequencies indicated high positive or negative emotional arousal. For *Sherlock*, we used four independent raters' emotional ratings [provided by Chen et al. (13)] at each scene of *Sherlock*, containing both valence (positive and negative) and arousal (scale of 1 to 5). Interrater reliability was high, with average of the pairwise raters' Pearson's $r = 0.564 \pm 0.084$, thus we used an averaged rating as a proxy for evolving emotional arousal, again separately for positive and negative emotional content. To quantify auditory salience of the stimulus, we extracted audio envelopes from the sound clips of both *Paranoia* and *Sherlock* using Hilbert transform

Fig. 1. Behavioral experiment results. Ratings of subjective engagement collected as participants listened to the *Paranoia* story (*Left*) or watched the *Sherlock* episode (*Right*). (*A*) Every participant's engagement ratings across time. Ratings were z-normalized across time for each participant. (*B*) Histograms of pairwise participants' response similarities. Engagement rating similarity was calculated by the Pearson's correlation between pairs of participants' engagement ratings across time. (*C*) Average engagement ratings, which are used as proxies for group-level states of engagement. The gray area indicates 95% confidence interval (CI). Event descriptions are given at moments of peak engagement.

(50). For *Sherlock,* the visual salience was represented by global luminance, calculated by averaging pixel-wise luminance values per frame of the video (51).

The occurrence of negative (partial $r = 0.133$, FDR-$P = 0.061$) but not positive (partial $r = -0.019$, FDR-$P = 0.862$) emotion words in *Paranoia* was marginally correlated with changes in engagement when controlling for the rest of the variables (*SI Appendix*, Fig. S1B). In *Sherlock*, both positive (partial $r = 0.530$, FDR-$P = 0.002$) and negative (partial $r = 0.539$, FDR-$P = 0.002$) emotional arousals were correlated with engagement. We observed no significant relationship between auditory envelopes and engagement, both for *Paranoia* (partial $r = -0.018$, FDR-$P = 0.663$) and *Sherlock* (partial $r = -0.103$, FDR-$P = 0.363$). Similarly, there was no significant correlation between visual luminance and engagement for *Sherlock* (partial $r = -0.230$, FDR-$P = 0.124$). For significance tests, observed partial correlation values were compared with null distributions of partial correlations between each factor and 10,000 permuted, phase-randomized engagement ratings [two-tailed test: $P = (1 + $ number of null $|r|$ values $\geq$ empirical $|r|) / (1 + $ number of permutations)]. We created such null distribution because phase randomization retains the same characteristics of temporal dynamics (i.e., frequency and amplitude) but in different phases. Multiple comparisons correction was applied to each dataset separately using FDR correction for number of features tested (i.e., three for *Paranoia* [positive emotion, negative emotion, auditory envelope] and four for *Sherlock* [with the addition of luminance]). The results suggest that engagement scales with emotional narrative content but not with the sensory-level salience of the stimuli themselves.

Neural Synchrony of the DMN Increases during Engaging Moments of the Narratives. We analyzed open-sourced fMRI data from two separate studies (13, 18). The datasets were collected from different participant samples, experimental sites, and research contexts, with different image-acquisition protocols. To assess the reproducibility and generalizability of our results, we applied different analysis pipelines to the two datasets throughout the study. This allowed us to conceptually replicate findings across samples and confirm that results are robust to particular analytic choices. We preprocessed the *Paranoia* fMRI data as described in the *Materials and Methods*, whereas we used the fully preprocessed images of *Sherlock* dataset, provided by Chen et al. (13).

We asked whether stimulus-driven patterns of neural activity scaled with changes in engagement. ISC is a method of isolating shared, stimulus-driven brain activity, assuming that if participants perceive the same stimulus at the same time, their shared variance in fMRI signal is related to stimulus processing (28, 45, 52, 53). Recent studies showed that ISC provides an alternative to identifying brain regions that are entrained to a stimulus with reduced intrinsic noise (52) compared to a conventional general linear model (GLM) that relies on fixed experimental manipulations with a priori hypotheses (45) (reference *SI Appendix*, Fig. S2 for GLM results using group-average engagement as a regressor). We hypothesized that ISC would increase as participants, on average, become more engaged in the story. We used HRF-convolved group-average engagement ratings from the behavioral studies as our index of narrative engagement because participant-specific measures of engagement were not available in the existing fMRI datasets.

Table 1. Definition of engagement instructed to the participants during the behavioral experiments

I find the story engaging when...	I find the story not engaging when...
I am curious and excited to know what's coming up next.	I am bored.
I am immersed in the story.	Other things pop into my mind, like my daily concerns or personal events.
My attention is focused on the story.	My attention is wandering away from the story.
The events are interesting.	I can feel myself dozing off.
	The events are not interesting.

Adapted from Busselle and Bilandzic (9).

Song et al.
Neural signatures of attentional engagement during narratives and its consequences for event memory

PNAS | 3 of 12
https://doi.org/10.1073/pnas.2021905118

To test the hypothesis that ISC varies with engagement, we calculated dynamic ISC using a tapered sliding window approach, where the Fisher's z–transformed Pearson's correlation between pairwise participants' blood oxygen level dependent (BOLD) responses was computed within the temporal window, repeatedly across the entire scan duration (54). We implemented a window size of 40 repetition time [TR] (= 40 s) for *Paranoia* and 30 TR (= 45 s) for *Sherlock* data, following the optimal window size suggested by previous literature (55–57), with a step size of 1 TR and a Gaussian kernel σ = 3 TR (reference *SI Appendix, Supplementary Text S2* for replications with different sliding window sizes). The BOLD time course was extracted from each region-of-interest (ROI) in Yeo et al.'s (58, 59) cortical and Brainnetome atlas's (60) subcortical parcellations (122 ROIs total) (*Materials and Methods*). Dynamic ISC was calculated for all pairs of participants and was averaged per ROI per window. The Pearson's r between the dynamic ISC time course and group-average engagement (smoothed with the same sliding window approach) was calculated for each ROI (Fig. 2*A*).

Fig. 2*B* shows the regions in which dynamic ISC significantly correlated with engagement ratings (two-tailed test nonparametric $P < 0.05$, uncorrected for multiple comparisons). Dynamic ISC in 19 of the 122 Yeo atlas ROIs was significantly correlated with engagement for *Paranoia*, and dynamic ISC in 21 ROIs was significantly correlated with engagement for *Sherlock* (*SI Appendix,* Table S2). In 18/19 and 18/21 of these regions, respectively, dynamic ISC was positively correlated with engagement, suggesting that the cross-participant neural synchrony increased as narrative engagement increased. Three regions—left posterior cingulate cortex [+6.1, +51.0, +31.5] ($r = 0.654$, $r = 0.309$ for *Paranoia* and

Sherlock, respectively), right angular gyrus [−50.9, +56.7, +29.7] ($r = 0.454$, $r = 0.323$), and right superior medial gyrus [−9.8, −49.2, +41.9] ($r = 0.449$, $r = 0.228$)—showed higher synchrony with increasing narrative engagement in both samples. The overlapping regions all corresponded to the DMN. Though the probability of regional overlap between the two samples was not significant (one-tailed $P = 0.681$, nonparametric permutation tests with randomly selected regions, iteration = 10,000), the probability of all overlapping regions corresponding to the DMN was significant above chance ($P = 0.029$). The results replicated when we used a different parcellation scheme, the Shen et al. (61) atlas (*SI Appendix,* Fig. S3). These results suggest that BOLD responses are more entrained to the stimulus when subjective levels of engagement increase, especially the regions of the DMN.

Multivariate Patterns of Activation Magnitude Do Not Predict Changes in Engagement. Recent work suggests that diverse cognitive and attentional states can be predicted from multivariate patterns of fMRI activity (62–64). Thus, we asked whether evolving states of engagement can be predicted from dynamic patterns of brain activity. To this end, we applied dynamic predictive modeling (39, 46) to BOLD time courses of ROIs to predict group-average engagement at every moment of time.

Prediction was conducted using a leave-one-subject-out (LOO) cross-validation within each dataset. Nonlinear support vector regression (SVR) models were trained using fMRI data from all but one participant and applied to the held-out participant's pattern of BOLD activity at every TR to predict the group-average engagement observed at the corresponding TR (Fig. 3*A*). Prediction accuracy was calculated by averaging the Fisher's z–transformed Pearson's correlations between the predicted and observed engagement dynamics across cross-validation folds. We used correlation as an indicator of predictive performance because we were interested in whether the model captures temporal dynamics, rather than the actual values of group-average engagement ratings. Nevertheless, we report both the mean squared error (MSE) and R^2 along with the correlation values. For significance tests, observed prediction accuracy was compared with results from 1,000 permutations where null models were trained and tested with actual brain patterns to predict the phase-randomized engagement ratings. We assumed a one-tailed significance test, with $P = (1 + \text{number of null } r \text{ values} \geq \text{empirical } r) / (1 + \text{number of permutations})$ for Pearson's correlation and R^2, but applied an opposite end of tail for MSE (i.e., the smaller the MSE, the better the prediction).

We examined whether magnitude of BOLD activity in regions selected in the above ISC analysis (i.e., the colored regions in Fig. 2*B*) predict changing levels of engagement. Models trained on BOLD signal magnitude in these regions did not show robust prediction of group-average engagement ratings (*Paranoia*: $r = 0.042$, $P = 0.395$, MSE = 1.051, $P = 0.006$, $R^2 = −0.051$, $P = 0.006$; *Sherlock*: $r = 0.111$, $P = 0.066$, MSE = 1.620, $P = 0.995$, $R^2 = −0.620$, $P = 0.995$). Even when we used magnitude of BOLD time courses of all 122 Yeo atlas ROIs as features, models did not show robust prediction of engagement (*Paranoia*: $r = 0.114$, $P = 0.277$, MSE = 1.012, $P = 0.096$, $R^2 = −0.012$, $P = 0.096$; *Sherlock*: $r = 0.232$, $P = 0.004$, MSE = 1.205, $P = 0.968$, $R^2 = −0.205$, $P = 0.969$). To increase the specificity of the neural features, we conducted additional feature selection in every cross-validation fold so that only the ROIs of which BOLD time courses are consistently correlated with group-average engagement were included as features to the model (one-sample Student's t test, $P < 0.01$) (65). Still, BOLD activation did not consistently predict engagement (*Paranoia*: $r = −0.021$, $P = 0.901$, MSE = 1.120, $P = 0.218$, $R^2 = −0.120$, $P = 0.218$; *Sherlock*: $r = 0.227$, $P = 0.003$, MSE = 1.239, $P = 0.572$, $R^2 = −0.239$, $P = 0.572$). Thus, patterns of BOLD signal magnitude across ROIs are not sufficient for predicting group-average narrative engagement in these samples.
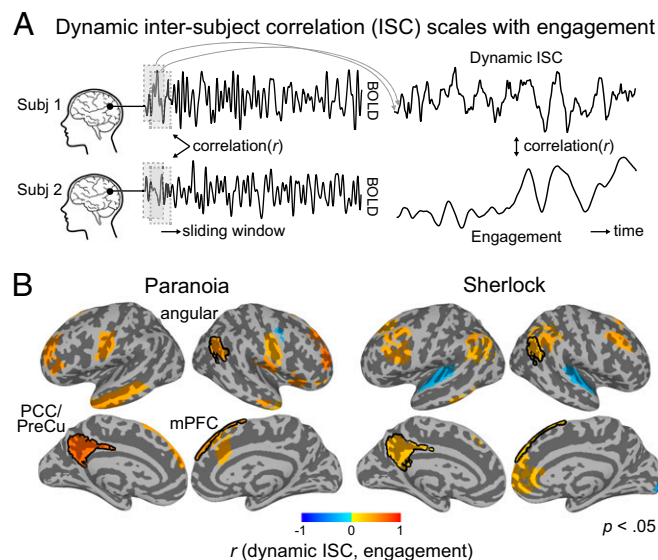
Fig. 2. Dynamic ISC scales with changes in engagement. (*A*) Schematic of dynamic ISC analysis. ISC was calculated by correlating pairwise participants' ROI time courses within a time window. Sliding window analysis was used to measure dynamic changes in ISC. The dynamic ISCs of all pairwise participants were averaged per ROI. The correlation between dynamic ISC and group-average engagement was calculated. Subject images credit: Free-PNGImg, licensed under CC BY-NC 4.0. (*B*) Regions that show significant correlation between dynamic ISC and group-average engagement (uncorrected $P < 0.05$) for *Paranoia* (*Left*) and *Sherlock* (*Right*) datasets. Significance tests were conducted per ROI by comparing the observed correlation value with a permuted distribution in which the engagement rating was phase randomized. Regions selected in both datasets are outlined with black contours. No significant relationship with engagement was observed in subcortical regions. Angular: angular gyrus, mPFC: medial prefrontal cortex, PCC/PreCu: posterior cingulate cortex/precuneus.

Functional Connectivity Predicts Changes in Engagement, Even across Different Stories. Although multivariate patterns of BOLD activity across regions failed to predict changes in engagement, evidence suggests that changes in the statistical interactions of activity in pairs of brain regions—functional connectivity (FC) dynamics—predict cognitive and attentional state changes during task performance (39, 44, 46). Thus, we examined whether whole-brain FC patterns predict evolving states of engagement.

Time-resolved FC was extracted by computing Fisher's z–transformed Pearson's correlations between the time courses of every pair of ROIs using a tapered sliding window (54). Predictive models were first trained and tested within dataset using LOO cross-validation. SVRs were trained to predict group-average engagement from all but one participants' multivariate FC patterns and then applied to dynamic FC patterns from the held-out individual to predict moment-to-moment engagement (Fig. 3A). Feature selection was conducted in every round of cross-validation; functional connections (FCs) significantly correlated with engagement in the training set were selected as features (one-sample Student's t test, $P < 0.01$) (65). Models predicted group-average engagement ratings above chance for both the *Paranoia* ($r = 0.380$, $P = 0.007$; MSE = 0.862, $P = 0.005$; $R^2 = 0.138$, $P = 0.005$) and *Sherlock* ($r = 0.582$, $P = 0.008$; MSE = 0.671, $P = 0.009$; $R^2 = 0.329$, $P = 0.009$) datasets (Fig. 4A). Of note, the null distributions are positively skewed (Fig. 4A), potentially because SVRs were trained and tested on the same behavioral outcome (that is, to generate the null distributions, an SVR model was trained to predict a phase-randomized engagement time course and tested using that same phase-randomized time course on each iteration; null model performance could have been inflated if a stimulus feature that influenced FC were correlated with the randomized engagement time course). Nonetheless, prediction accuracy was significantly above these skewed chance distributions. Prediction performance

was significantly higher than chance when we used a different type of null distribution generated by randomly selecting the same number of FCs from among those not selected in the actual feature selection (*Paranoia*, $P < 0.001$, $P < 0.001$, $P < 0.001$; *Sherlock*: $P < 0.001$, $P < 0.001$, $P < 0.001$, respectively for Pearson's r, MSE, and R^2; iteration = 1,000).
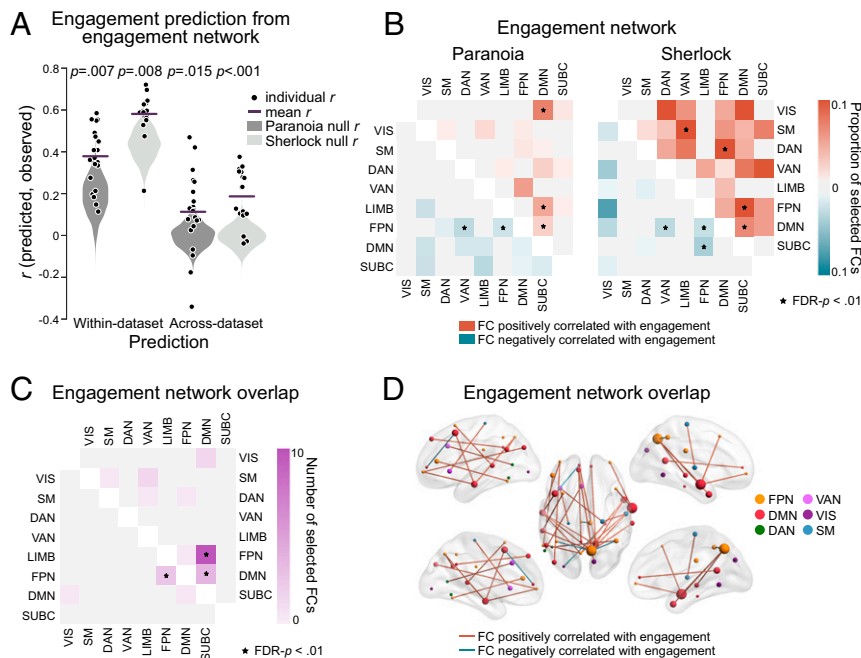
Within-dataset prediction demonstrates that models based on FC dynamics predict stimulus-related engagement within two narratives. To confirm that models are capturing engagement and not other stimulus-specific regularities, we conducted across-dataset prediction (Fig. 3B). SVRs learned the mappings between multivariate patterns of FCs and moment-to-moment engagement using data from all participants in one dataset (e.g., *Paranoia*). The FCs selected in every round of cross-validation during the within-dataset prediction were used as features (Fig. 4B). Next, the model trained in one dataset was applied to predict engagement in the held-out dataset (e.g., *Sherlock*). Across-dataset prediction was successful both when predicting engagement of *Paranoia* from a model trained with *Sherlock* ($r = 0.114$, $P = 0.015$; MSE = 1.062, $P = 0.015$; $R^2 = -0.063$, $P = 0.015$; reference *SI Appendix,* Fig. S4 for interpretation of negative R^2) and when predicting engagement of *Sherlock* from a model trained with *Paranoia* ($r = 0.188$, $P < 0.001$; MSE = 1.022, $P < 0.001$; $R^2 = -0.023$, $P < 0.001$). As expected, the null distribution was not shifted in the across-dataset prediction, suggesting that the model did not learn a spurious relationship between FC and story-specific properties other than subjective engagement (Fig. 4A). Results replicated with additional sliding window sizes (*SI Appendix,* Fig. S5). However, for across-dataset prediction, the prediction performance of the engagement networks was not significantly better than that of size-matched random networks selected from FCs outside the engagement networks (*Paranoia*: $P = 0.492$, $P = 0.458$, $P = 0.458$; *Sherlock*: $P = 0.031$, $P = 0.203$, $P = 0.203$, respectively for Pearson's r, MSE, and R^2), which implies that narrative engagement may be reflected in widely distributed patterns of time-varying FC, rather than being specifically predicted by these particular sets of connections.

To characterize the anatomy of the predictive networks, we visualized the FCs consistently selected in every round of within-dataset cross-validation for the *Paranoia* and *Sherlock* datasets (Fig. 4B). We call these sets of FCs the engagement networks. For *Paranoia*, a set of 205 FCs was consistently selected, with 125 FCs positively and 80 FCs negatively correlated with engagement. For *Sherlock*, 685 FCs were selected, with 583 FCs positively and 102 FCs negatively correlated with engagement. We next asked whether there is significant overlap between the *Paranoia* and *Sherlock* engagement networks. Thirty-one FCs were included in the engagement networks in both datasets, with a significant degree of overlap (using the hypergeometric cumulative distribution function, $P = 0.002$). Twenty-six FCs were positively correlated (significance of overlap, $P < 0.001$), and 5 FCs were negatively correlated ($P < 0.001$) with engagement in both datasets (visualized in Fig. 4 C and D).

To examine whether particular functional networks were represented in the engagement networks more frequently than would be expected by chance, we calculated the proportion of selected FCs relative to the total number of possible connections between each pair of functional networks (one-tailed test nonparametric FDR-$P < 0.01$; Fig. 4 B and C). In both datasets, connections between the DMN and frontoparietal control network (FPN) and connections within the DMN were positively correlated with group-average engagement, whereas connections within the FPN were negatively correlated with engagement. The results suggest that the FCs within and between regions of the DMN and FPN contain representations of higher-order cognitive states of engagement that are not specific to sensory modality (audiovisual versus auditory) or narrative contents (different stories).
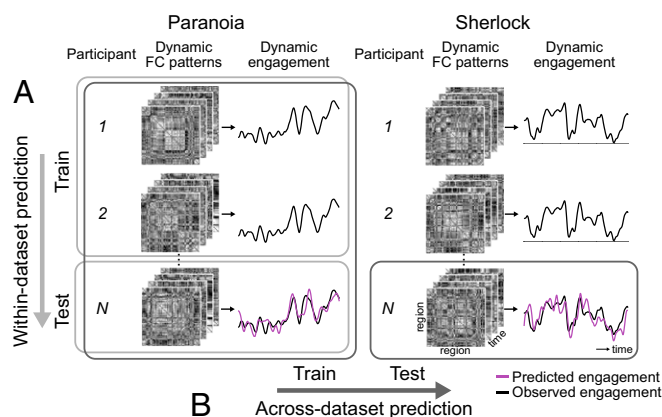
Fig. 3. Schematic of dynamic predictive modeling. (*A*) Within-dataset prediction. Internal model validation is conducted using LOO cross-validation. Multivariate patterns of brain activity (e.g., patterns of pairwise regional FC) at time *t* are aligned with a group-level behavioral score (e.g., group-average engagement) at time *t*, across all participants but one. The SVR model is trained using data from all time points and training participants, and is then applied to the held-out participant's time-varying FC patterns to predict engagement at every time step. Prediction accuracy is measured as the average of the cross-validation folds' Fisher's *z*–transformed Pearson's correlation between predicted and observed engagement time courses. (*B*) Across-dataset prediction. To externally validate predictive models, an SVR model is trained using all participants' time-varying FC and the group-average engagement ratings from a single dataset (e.g., *Paranoia*). The model is tested on fMRI data from every participant in an independent dataset (e.g., *Sherlock*). Significance was tested by training null models to predict phase-randomized behavioral time courses. The observed correlation between the predicted and observed engagement was compared with the distribution of null model correlations (one-tailed tests).

Song et al.
Neural signatures of attentional engagement during narratives and its consequences for event memory

PNAS | **5 of 12**
https://doi.org/10.1073/pnas.2021905118

Fig. 4. FC predicts group-average dynamic states of engagement. (*A*) Model performance for the within-dataset (*Left*) and across-dataset (*Right*) predictions for *Paranoia* (dark gray) and *Sherlock* (light gray). Black dots indicate Pearson's correlations between predicted and observed engagement time courses for every participant. Lines indicate prediction performance, or the mean *r* across cross-validation folds. Gray violin plots show null distributions of mean prediction performance, generated from models that predict phase-randomized engagement time courses. The significance of empirical *r* was computed based on the null distribution (one-tailed tests). (*B*) FCs selected in every cross-validation fold, grouped in predefined functional networks (59). Colors indicate the proportion of selected FCs from all possible connections of each network pair. The upper triangle matrix (red) illustrates the proportion of FCs positively correlated with engagement, and the lower triangle matrix (blue) illustrates the proportion of FCs negatively correlated with engagement. Network pairs that are selected more than expected by chance are indicated with asterisks (one-tailed tests, FDR-*P* < 0.01). (*C*) The number of FCs correlated with engagement dynamics of both the *Paranoia* and *Sherlock*, summarized in functional networks. The upper triangle indicates FCs positively correlated with engagement, and the lower triangle indicates FCs negatively correlated with engagement in both samples (FDR-*P* < 0.01). (*D*) The overlapping FCs in *C*, visualized in the brain (66). Red lines indicate FCs positively correlated with engagement, whereas blue lines indicate FCs negatively correlated with engagement. Node colors indicate different functional networks. Node sizes indicate total number of FCs involving that node.
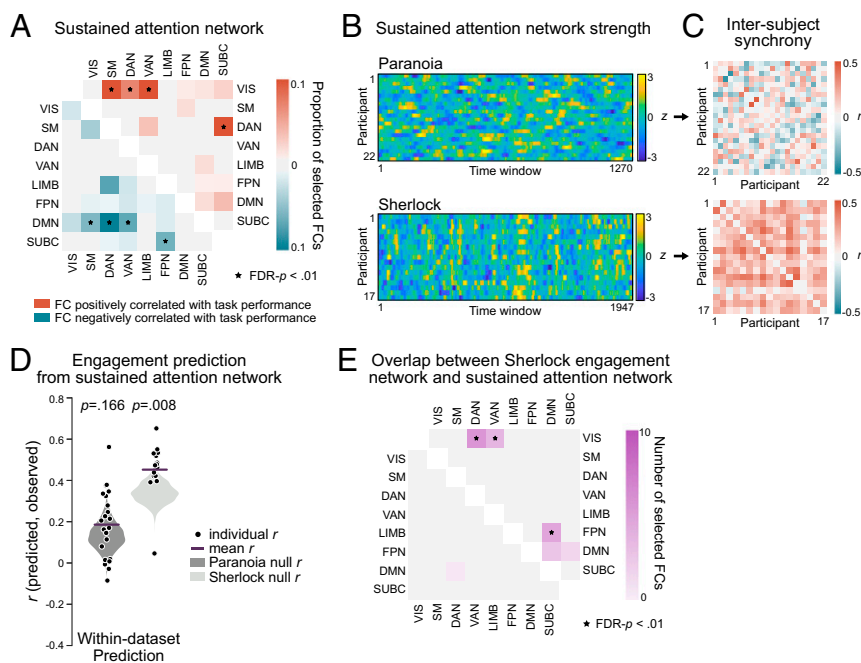
Visual Sustained Attention Network Predicts Engagement during Visual Narratives. Previous work theorized that narrative engagement involves attentional focus to story events (9, 10, 67, 68). To test this empirically, we asked whether a sustained attention network, a set of FCs that predicts trait-like differences in sustained attention abilities (43) and state-like changes in attentional task performance during a gradual-onset continuous performance task (gradCPT) (44), also predicts dynamic changes in naturalistic narrative engagement. We hypothesized that FCs that predict the ability to stay focused during a controlled experimental task may also reflect the degree to which individuals are engaged in narratives at each moment of time. To do so, we first redefined the sustained attention network using data from Rosenberg et al. (43) in Yeo atlas space in order to apply it to the *Paranoia* and *Sherlock* datasets (Fig. 5*A*; *Materials and Methods*). Results revealed that connections between the visual network (VIS) and the somatosensory-motor (SM), dorsal attention (DAN), and ventral attention networks (VAN), and connections between the DAN and subcortical network (SUBC) were positively correlated with individuals' gradCPT performance, potentially reflecting the fact that the task requires participants to exert goal-directed attention to visually presented stimuli. On the other hand, connections between the DMN and SM, DAN, and VAN, and connections between the FPN and SUBC were negatively correlated with gradCPT performance, aligning with conceptions of the DMN as task-negative network in certain contexts (34, 35).

To test whether sustained attention network dynamics were time locked to the narratives, we examined whether network strength was synchronized across individuals when comprehending

the same story. Fig. 5*B* depicts the time-resolved network strength of all participants. The strength of sustained attention network was computed as the difference between the mean FC of the positively correlated FCs, minus the mean FC of the negatively correlated FCs. For *Paranoia*, pairwise subject similarity in network strength was, on average, $r = 0.002 \pm 0.168$, with only 39.39% of pairwise participants showing positive correlations (FDR-*P* < 0.05; Fig. 5*C*). For *Sherlock*, however, participants exhibited significant degrees of synchrony in sustained attention network strength, with a mean $r = 0.191 \pm 0.117$ and 89.71% of pairwise participants showing positive correlations (FDR-*P* < 0.05; Fig. 5*C*). We examined whether the dynamics of sustained attention network strength are correlated with group-average engagement. We did not observe evidence of correlations between sustained attention network dynamics and engagement (mean of the correlations between each participant's sustained attention network strength time course and the group-average engagement time course: *Paranoia*: $r = 0.033 \pm 0.225$; *Sherlock*: $r = 0.047 \pm 0.140$), suggesting that the sustained attention network does not explicitly represent fluctuating states of narrative engagement in its overall strength alone.

We next asked if multivariate patterns of time-resolved FCs in the sustained attention network predict group-average engagement (Fig. 5*D*). When trained with LOO cross-validation, SVRs based on FC patterns of the sustained attention network predicted changes in engagement in the *Sherlock* dataset ($r = 0.458$, $P = 0.008$, MSE = 0.794, $P = 0.009$, $R^2 = 0.206$, $P = 0.009$), in which the narrative was delivered in the auditory and visual modalities but not in the *Paranoia* dataset ($r = 0.192$, $P = 0.166$, MSE = 0.998, $P = 0.206$, $R^2 = 0.002$, $P = 0.206$), which was

Fig. 5. Visual sustained attention network in relation to narrative engagement. (A) FCs significantly correlated with the gradCPT performance of 25 individuals, summarized in Yeo et al.'s (59) functional network space (one-tailed tests, FDR-P < 0.01). The upper triangle matrix (red) illustrates the proportion of FCs positively correlated with sustained attention scores, and the lower triangle matrix (blue) illustrates the proportion of FCs negatively correlated with sustained attention scores. (B) Time-resolved sustained attention network strength of every fMRI participant. Network strength is computed as the difference between the average FC time courses of the positively correlated FCs and negatively correlated FCs, z-normalized across time within participant. (C) Pairwise participants' synchrony (Pearson's correlation) in sustained attention network strength time courses of B. (D) Predictive performance of the within-dataset engagement prediction from the sustained attention network, respectively, for *Paranoia* (dark gray) and *Sherlock* (light gray). The patterns of FCs included in A were used as multivariate features in the model, where the model predicted time-varying narrative engagement. (E) The number of overlapping FCs for the engagement network during movie watching (*Sherlock*) and sustained attention network during gradCPT performance, summarized in functional networks. The upper triangle indicates FCs positively correlated with state-level differences in narrative engagement and trait-level differences in sustained attention task performance, whereas the lower triangle indicates FCs with negative correlations in both datasets (FDR-P < 0.01).

delivered only in the auditory modality. Prediction performance was robust for *Sherlock* when compared to the size-matched random networks selected from FCs outside the engagement network and sustained attention network ($P = 0.013$, $P = 0.013$, $P = 0.013$, respectively for Pearson's r, MSE, and R^2). The results provide empirical evidence that the sustained attention network—which was defined in a completely independent study using a controlled experimental paradigm with visual stimuli—predicts stimulus-related engagement during an audiovisual narrative (*Sherlock*) but not an audio-only narrative (*Paranoia*).
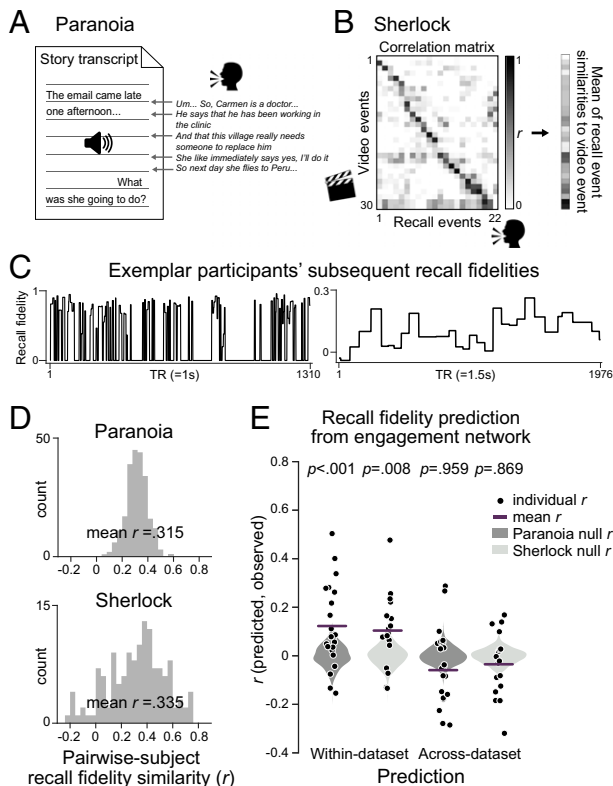
Finally, we characterized the anatomical overlap between the *Sherlock* engagement network (Fig. 4 *B*, *Right*) and the sustained attention network (Fig. 5*A*). There was a significant overlap of the FCs positively correlated with both sustained attention and *Sherlock* engagement (20 overlapping FCs; calculated using the hypergeometric cumulative distribution function, $P = 0.002$; Fig. 5*E*). Connections between the VIS and the DAN and VAN and connections between the DMN and FPN were included in both networks above chance (one-tailed test nonparametric FDR $P < 0.01$). On the other hand, there was no significant overlap of the negatively correlated FC (one overlapping FC; $P = 0.769$). The results were robust when we calculated anatomical overlap using a range of feature-selection thresholds (*SI Appendix*, Table S3). The partial overlap between the networks predicting sustained attention and engagement suggests that shared but also distinctive processes are involved in maintaining focus on a visual sustained attention task and a visual narrative. However, since these results are based on only a single narrative representing each sensory modality, the modality specificity of the overlap between the

sustained attention network and engagement network should be tested in future work.

Engagement Network during Memory Encoding Predicts Later Recall of the Events. During both a television episode and an audio-narrated story, people's engagement fluctuated over time with patterns of functional brain connectivity. What consequences do these changes in narrative engagement have for how we remember stories? To address this question, we asked whether attentional engagement facilitates encoding of events into long-term memory. To measure how well each moment of the narratives was remembered, we analyzed fMRI participants' free recall data provided by Finn et al. (18) and Chen et al. (13). Again, to test for the robustness of results with different analysis pipelines, we used two different approaches to quantify recall. For *Paranoia*, each phrase of participants' recall was manually matched to the semantically closest sentence in the original story transcript (Fig. 6*A*). Every word in the story and recall transcripts was then represented as a vector in a distributional word embedding space, GloVe (69), and was combined to create a sentence embedding vector (70). The degree of semantic similarity between the audio transcript and matched recall was computed using the cosine similarities between the sentence embedding vectors. For *Sherlock*, we used a data-driven topic modeling approach (71), similar to that implemented by Heusser et al. (72), to segment the stories into multiple events and quantified the similarities between recalled events with the video annotations (Fig. 6*B*). Fig. 6*C* illustrates example "recall fidelity" time courses.

To examine whether people tend to remember and forget similar events, we calculated the similarity of recall fidelity time

Song et al.
Neural signatures of attentional engagement during narratives and its consequences for event memory

PNAS | 7 of 12
https://doi.org/10.1073/pnas.2021905118

Fig. 6. Individual-specific story recall in relation to engagement network dynamics. To quantify the fidelity with which events in a story were recalled, (A) in *Paranoia*, we matched the phrases uttered during free recall to the story transcript and calculated semantic similarity using distributional word embeddings. (B) In *Sherlock*, we applied dynamic topic modeling fitted with video annotations to extract topic vectors of individuals' recall transcripts. The topic vector similarities between recall and video annotations were summed across recalled events (71, 72). (C) Representative participants' subsequent recall fidelity of the events for *Paranoia* (*Left*) and *Sherlock* (*Right*). (D) Histograms of the pairwise participants' recall similarity values. Recall similarity was calculated by the Pearson's correlation between pairs of participants' recall fidelity time courses. (E) Predictive performance of individual-specific recall fidelity from the engagement networks of within dataset (*Left*) and across dataset (*Right*) for *Paranoia* (dark gray) and *Sherlock* (light gray). The patterns of FCs that predicted group-average engagement, illustrated in Fig. 4B, were used as multivariate features in the model. Black dots indicate Pearson's correlations between predicted and observed recall fidelity time courses for every participant. Lines indicate prediction performance, or the mean r across cross-validation folds. Gray violin plots show null distributions of mean prediction performance, generated from models that predict shuffled recall fidelity time courses.

courses of all pairwise fMRI participants (Fig. 6D). Almost every pair of participants exhibited similar trends of recall, for both *Paranoia* (mean $r = 0.315 \pm 0.092$, with 100% of pairwise participants showing positive correlations, FDR-$P < 0.05$) and *Sherlock* (mean $r = 0.335 \pm 0.255$, with 86.77% showing positive correlations, FDR-$P < 0.05$). The results imply that particular events in a story tend to be more likely to be recalled than other events (72), replicating previous empirical findings of Meyer and McConkie (73), and revisiting theoretical explanations that the subjective importance of events within situational contexts influences memory for those events (74).

We asked whether fMRI participants' individual-specific recall fidelity time courses were significantly related with behavioral participants' group-average engagement ratings. Correlations between individual-specific recall time courses and engagement were averaged and then compared with the null distribution where

engagement ratings were phase randomized (two-tailed tests, iteration = 10,000). We observed different results for the two datasets: significant correlations between engagement and recall in *Sherlock* (mean $r = 0.197 \pm 0.145$, $P = 0.024$) but not *Paranoia* ($r = 0.014 \pm 0.103$, $P = 0.833$).

The disparate results may be due to story-specific components, differences in the behavioral analysis pipelines, or a possibility that group-averaged behavioral engagement ratings do not fully capture attention fluctuations that are consequential for memory. Therefore, we applied a connectome-based dynamic predictive modeling approach, asking whether each narrative's engagement network (Fig. 4B) predicted the same narrative's individual-specific story recall (Fig. 6C). We trained SVR models with LOO cross-validations to predict recall fidelity from patterns of FCs in the engagement network and tested whether models predicted held-out participant's recall given their engagement network (replacing group-average engagement with individual-specific recall time courses in Fig. 3). Models predicted the fidelity with which each moment of the story would be later recalled (Fig. 6E) in both the *Paranoia* ($r = 0.123$, $P < 0.001$; MSE = 1.041, $P < 0.001$; $R^2 = -0.042$, $P < 0.001$) and *Sherlock* datasets ($r = 0.104$, $P = 0.008$; MSE = 1.058, $P = 0.018$; $R^2 = -0.058$, $P = 0.018$). However, the engagement network did not show specificity in predicting subsequent recall, given that the accuracy of recall fidelity predictions from engagement network FCs did not significantly differ from the accuracy of predictions made by the size-matched random networks selected from FCs outside the engagement networks (*Paranoia*: $P = 0.092$, $P = 0.155$, $P = 0.166$; *Sherlock*: $P = 0.823$, $P = 0.851$, $P = 0.851$, respectively for Pearson's r, MSE, and R^2). Models did not generalize to predict recall across datasets (*Paranoia* predicted from a model trained with *Sherlock*: $r = -0.059$, $P = 0.959$; MSE = 1.175, $P = 0.998$; $R^2 = -0.176$, $P = 0.998$; *Sherlock* predicted from model trained with *Paranoia*: $r = -0.035$, $P = 0.869$; MSE = 1.126, $P = 0.718$; $R^2 = -0.127$, $P = 0.717$), suggesting that the engagement network model specific to each narrative predicts subsequent recall. Thus, although time-varying FC in networks related to engagement predicted subsequent recall within dataset, the lack of cross-story generalization leaves open the possibility that models are capturing stimulus-specific effects in addition to signal related to memory encoding of narrative events.

Discussion

The degree to which we are engaged in narratives in the real world fluctuates over time. Sometimes we may find ourselves immersed in a particularly engrossing story, whereas at other moments we may become bored with a plot line or struggle to follow a conversation thread. What drives these changes in narrative engagement, and how does engagement affect what we later remember about a story?

Here, using data from independent behavioral and fMRI studies, we test theoretical proposals that narrative engagement reflects states of heightened attentional focus and emotional arousal. Specifically, we analyzed two open-source fMRI datasets collected as participants watched an episode of *Sherlock* or listened to an audio-narrated story, *Paranoia*. We ran behavioral experiments in which independent participants continuously rated how engaging they found the narratives. Self-reported engagement fluctuated across time and was synchronous across individuals, demonstrating that this paradigm captures states of engagement that are shared across individuals (Fig. 1). Group-average changes in engagement were correlated with the narratives' emotional contents but not their low-level sensory salience. Furthermore, dynamic ISC revealed that activity in DMN is more synchronized across individuals during engaging moments of the narratives, suggesting that DMN activity becomes more entrained to narratives when people are more engaged (Fig. 2). Fully cross-validated models trained on time-resolved whole-brain FC (Fig. 3), but not BOLD activity, predicted moment-to-moment changes in engagement in each fMRI sample.

8 of 12 | PNAS
https://doi.org/10.1073/pnas.2021905118

Song et al.
Neural signatures of attentional engagement during narratives and its consequences for event memory

Predictive models further generalized across datasets (Fig. 4), demonstrating that FC contains information about higher-order, modality-general cognitive states of engagement. We observed that the sustained attention network, defined using data collected during a visual attention task, predicted changes in engagement as individuals watched *Sherlock*, but not as individuals listened to *Paranoia* (Fig. 5), suggesting that brain networks involved in sustained attention and attentional control may be partially involved when individuals are attentively engaged in naturalistic contexts. Lastly, we found that dataset-specific engagement networks predicted how well participants remembered story events (Fig. 6), suggesting that brain networks related to dynamic attentional engagement during encoding are also related to later memory for narratives.

Historically, research on sustained attention has studied attention to external stimuli during psychological tasks requiring top-down attentional control and cognitive effort (5, 75–77). Complementary work has described states of attention associated with less subjective effort (78), such as flow states of complete absorption in an activity (8), in-the-zone moments of optimal task performance (36, 38, 79, 80), and soft fascination to natural environments (81) or aesthetic pieces of art (82). Recent work has also characterized stimulus-independent internal attention and mind wandering (7, 83–86). However, relatively few studies have tackled attentional mechanisms during the processing of contextually rich everyday narratives to ask, for example, to what degree narrative engagement requires attentional control, how attentional fluctuations modulate moment-to-moment comprehension and memory, or how state- or trait-like differences in subjective engagement are reflected in the brain activity. Bellana and Honey* theorized that deep information processing (87) is involved when comprehending narratives, which leads to high engagement (or transportation, absorption, and immersion) during perception and rich memory representation.

Our work provides empirical support that the functional networks involved in top-down sustained attention (Fig. 5A) and stimulus-related narrative engagement (Fig. 4B) are partially overlapping and that these two networks both play a role in predicting subjective engagement and subsequent memory. Patterns of FC in the sustained attention network, defined to predict performance on a visual CPT (43), predicted narrative engagement of an audiovisual narrative, *Sherlock*, but not an audio-narrated story, *Paranoia*. This raises the possibility that the sustained attention network captures visual sustained attention specifically. However, because the current results are based on a single narrative from each sensory modality, many other differences between the *Paranoia* and *Sherlock* datasets could account for the sustained attention network's pattern of generalization (e.g., sample size, scan parameters, or preprocessing pipelines). It is also possible that the inconsistent pattern of results across datasets could have arisen due to a false-positive finding in one sample. Thus, future work can test whether networks defined to predict visual attention capture visual narrative engagement whereas networks trained to predict auditory attention capture auditory narrative engagement (88, 89). Analyses of the overlap between these hypothetical modality-specific networks could provide evidence for a hierarchical network organization involving both modality-specific and modality-general attention subnetworks.

The current results support characterizations of the DMN as a modality-general network involved in attention and narrative processing. In both the *Paranoia* and *Sherlock* datasets, stimulus-driven patterns of DMN activity varied with changes in narrative engagement, replicating previous observations that intersubject neural synchrony increases during "emotionally charged and surprising" movie moments (28). In addition to cross-subject synchrony in DMN activity, time-varying FC involving DMN regions also reflects time-varying engagement. In particular, within-network DMN connections and connections between DMN and FPN regions (Fig. 4C) predicted subjective engagement in both *Paranoia* and *Sherlock*. These results align with the findings of Zhang et al. (90) that resting-state FC within the DMN and between the DMN and visual sensory-processing network predicts trait-level differences in narrative comprehension and attention during reading. Together, our results suggest that attentional engagement may modulate the degree to which DMN integrates narrative events on a moment-to-moment basis by changing its activations and large-scale interactions with other brain regions.

Previous research has studied interactions between sustained attention and subsequent memory using controlled psychological tasks that measure participants' memory for stimuli encountered in different attentional states. For example, participants in a study from deBettencourt et al. (4) performed a CPT and then reported their incidental recognition memory of the presented images. CPT stimuli seen during engaged attentional states (indexed with response times) were better remembered. Complementary work has examined relationships between attention and recall during a naturalistic task using individual differences approaches (29, 90, 91). Jangraw et al. (91) asked participants to read transcripts of Greek history lectures during fMRI, then measured their performance on a postscan comprehension test. Individuals who showed FC signatures of stronger sustained attention during reading performed better on a comprehension test. Cohen and Parra (29) found that individuals with EEG activity time courses more similar to the rest of the group during narrative presentation scored higher on a memory test conducted 3 wk later. However, attending to and remembering events constantly fluctuates across time, heavily dependent on the situational contexts and relational structure. It is yet unknown the role of dynamic memory when constructing structured representation of narratives (92, 93) or how we modulate attention online to encode critical events and reinstate relevant memories. The implication of our results—that neural signatures of engagement predict subsequent memory—motivates future work addressing neural mechanisms of the bidirectional interactions between sustained attention and memory in naturalistic contexts.

The current study focuses on group-level states of engagement that are shared across individuals and are thus stimulus-related. Future work characterizing each person's unique pattern of attentional engagement during narratives can help disentangle intrinsic from stimulus-related attention fluctuations and elucidate their consequences for memory at the individual level. Narrative engagement can be inferred online using physiological measures such as heart rate and electrodermal activity (94, 95), facial expressions (96), or with pupil dilation (97) and blink rate (98, 99). In a recent study, van der Meer et al. (100) inferred participants' subjective immersion to the movie from concurrent recordings of heart rate and pupil diameter during fMRI. Furthermore, intermittent experience sampling using thought probes, or multidimensional experience sampling (83, 84), during fMRI can track individuals' moment-to-moment degree of attentional engagement and even thought contents during movie watching or story listening.

In sum, we show that engagement during story comprehension dynamically fluctuates across time, driven by narrative content. Our results characterize the relationship between sustained attention and narrative engagement and elucidate neural signatures that predict future event memory.

Materials and Methods

Behavioral Experiments. Twenty-one individuals from the University of Chicago or surrounding community participated in a behavioral study in which they listened to the *Paranoia* story (3 left handed, 11 females, 18 to 30 y, mean age 23.57 ± 4.04; 12 native English speakers, 2 bilingual English speakers). Seventeen individuals participated in a behavioral study in which

---

*B. Bellana, C. J. Honey, "A persistent influence of narrative transportation on subsequent thought" Presented in Context & Episodic Memory Symposium (23 August 2020).

Song et al.
Neural signatures of attentional engagement during narratives and its consequences for event memory

PNAS | 9 of 12
https://doi.org/10.1073/pnas.2021905118

PSYCHOLOGICAL AND COGNITIVE SCIENCES

they watched an episode of television series *Sherlock* (1 left-handed, 11 females, 18 to 30 y, mean age 22.06 ± 3.75; 11 native English speakers, 2 bilingual English speakers). All participants reported no history of visual or hearing impairments, provided written informed consent, and were compensated for their participation. The study was approved by the Institutional Review Board of the University of Chicago.

Stimulus presentation and response recording were controlled with PsychoPy3 (101). The experiment took place in a dimly lit room in front of a Macintosh Apple monitor. Participants either listened to a 20-min audio-narrated *Paranoia* or watched a 50-min episode of *Sherlock* while continuously rating how engaging they found the story by adjusting a scale bar from 1 ("Not engaging at all") to 9 ("Completely engaging"). The scale bar was constantly visible on the bottom of the computer monitor. Other aspects of the experiment replicated, as closely as possible, previous fMRI studies: the same visual image was shown at the center of the screen as participants listened to *Paranoia* (18), and "run breaks" occurred at the same moments for both stories. The timing of participants' button presses was recorded.

To confirm compliance, individuals who listened to *Paranoia* completed nine comprehension questions asking about the events. The average percent correct was 95.24 ± 8.29%, verifying that all participants maintained overall attention to the story. Participants who watched *Sherlock* were asked to provide a spoken recall of the story in as much detail as they remembered as soon as they finished watching the episode. These recall data were not analyzed here. All participants were qualitatively assessed to have paid attention to the story.

Behavioral Data Analysis. All participants' button-press responses were sampled to TR duration (1s for *Paranoia*; 1.5s for *Sherlock*) and z-normalized across time. Group-average engagement ratings were calculated by averaging all participants' normalized engagement ratings. The averaged engagement time courses were convolved with the HRF to be applied to separate groups of individuals who participated in the fMRI studies.

Several fMRI analyses in the study included the application of a tapered sliding window. To relate behavioral time course to fMRI time courses, we applied the same tapered sliding windows to the behavioral data time courses. A sliding window size of 40 TR (= 40 s) for *Paranoia* and 30 TR (= 45 s) for *Sherlock* datasets, with a step size of 1 TR and a Gaussian kernel σ = 3 TR were applied. For the beginning and end of the time courses (for which the tail of the Gaussian kernel was cropped), we adjusted the weights so that the cumulative Gaussian distribution summed to 1. The adjustments changed the peak weight of the Gaussian distribution from 1 to 1.031 maximum for *Paranoia* and 1.041 maximum for *Sherlock*. The same parameters were used throughout the study.

fMRI Image Acquisition and Preprocessing. The raw structural and functional images of the *Paranoia* dataset were downloaded from OpenNeuro (102). Functional images of Finn et al. (18) were acquired from 3T Siemens Trim-Trio, using T2*-weighted echo planar imaging (EPI) multiband sequence (TR = 1,000 ms, echo time = 30 ms, voxel size = 2.0 mm isotropic, flip angle = 60°, field of view = 220 × 220 mm). Structural images were bias-field corrected and spatially normalized to the Montreal Neurological Institute (MNI) space. Functional images were motion corrected using six rigid-body transformation parameters and registered to MNI-aligned T1-weighted images. White matter and cerebrospinal fluid masks were defined in MNI space and warped into the native space. Linear drift, 24-parameter motion parameters, mean signals from white matter and cerebrospinal fluid, and mean global signal were regressed from the BOLD time course. We applied a band-pass filter (0.009 Hz < f < 0.08 Hz) to remove low-frequency confounds and high-frequency physiological noise. The data were spatially smoothed with a Gaussian kernel of full width at half maximum of 4 mm.

The preprocessed images from the *Sherlock* dataset were downloaded from Princeton University's DataSpace repository (103). Functional images of Chen et al. (13) were collected from 3T Siemens Skyra, using T2*-weighted EPI sequence (TR = 1,500 ms, echo time = 28 ms, voxel size = 3.0 × 3.0 × 4.0 mm, flip angle = 64°, field of view = 192 × 192 mm). Preprocessing steps followed slice timing correction, motion correction, linear detrending, high-pass filtering (140-s cutoff), coregistration, and affine transformation to the MNI space. The functional images were resampled to 3-mm isotropic voxels. One participant's dataset was missing 51 TRs from the end of the second run. These TRs were discarded in all of the analyses.

All analyses were conducted in the volumetric space using AFNI software. The cortical surface of the MNI standard template was reconstructed using Freesurfer (104) for visualization purposes.

Whole-Brain Parcellation. Cortical regions were parcellated into 114 ROIs following Yeo et al. (58) based on a seven-network cortical parcellation

estimated from the resting-state functional data of 1,000 adults (59). Subcortical regions were parcellated into eight ROIs, corresponding to the bilateral amygdala, hippocampus, thalamus, and striatum, extracted from the subcortical nuclei masks of the Brainnetome atlas (60). The time courses of the voxels corresponding to each of the 122 ROIs were averaged to a single, representative time course. The eight functional networks include VIS and SM relevant to sensory-motor processing, DAN and VAN relevant to top-down guided attention or attentional shifts (105), limbic network (LIMB), DMN and FPN relevant to transmodal information processing (106), and the SUBC. Each ROI was labeled with one of the eight functional networks, which was provided by Yeo et al. (58, 59, 107). To examine the robustness of our results, we replicated the dynamic ISC analysis using the 268-node Shen atlas (61).

Dynamic Intersubject Correlation. A tapered sliding window was applied to calculate pairwise subjects' BOLD time course ISC (Fig. 2A). Nonparametric permutation tests were used to quantify the significance of the dynamic ISC of each ROI. Null distributions were generated by taking Pearson's r between the actual dynamic ISC and the phase-randomized engagement ratings (two-tailed test, uncorrected for multiple comparisons, iteration = 10,000).

Dynamic Predictive Modeling. For models based on FC patterns, time-resolved FC matrices were computed as the Pearson's correlations (Fisher's r-to-z–transformed) between the BOLD signal time courses of every pair of regions (122 × 122 ROIs) using a tapered sliding window. SVR models were implemented with python (sklearn.svm.SVR; rbf kernel, maximum iteration set to 1,000). The time series of every feature was z-normalized across time to retain temporal variance within a person while ruling out across-individual variance. Every time step was treated as an independent sample to the model.

In addition to measuring prediction performance represented with a mean r across results in all cross-validation folds, we tested prediction performance respectively for each cross-validation (*SI Appendix*, Figs. S6 and S7).

Figs. 4B and 5A illustrate the proportions of pairwise regions that were selected in every cross-validation fold, grouped by the ROI's predefined functional network. The proportion of selected pairs was compared with chance using nonparametric permutation test. The empirical proportion was compared with a null distribution in which the same number of features was randomly selected from all possible pairs of regions (one-tailed test FDR-P < 0.01, corrected for network pairs; iteration = 10,000). The significance in the number of overlapping FCs in Figs. 4C and 5E was calculated by comparing with the null distribution of the overlapping network pairs of the 10,000 size-matched random networks.

Sustained Attention Network. We reanalyzed raw fMRI data from Rosenberg et al. (43), which were collected from 25 participants as they completed three runs (each with four 3-min blocks) of a sustained attention task, the gradCPT (36). In the task, participants viewed a stream of scene images that gradually transitioned from one to the next every 800 ms. Participants were instructed to press a button every time they saw a frequent category image (cities; 90% of trials) but to withhold response to infrequent category images (mountains; 10% of trials). Functional MRI preprocessing steps matched those applied to *Paranoia* data (and differed from those applied to *Sherlock* data). We used Yeo et al.'s (59) 122-ROI parcellation scheme to directly compare with our results and increase the computational feasibility of dynamic FC pattern–based predictive modeling (the Yeo atlas has 122 ROIs, whereas the Shen atlas, used in Rosenberg et al. (43), has 268). The behavioral performance of each gradCPT block was assessed with sensitivity (d') or hit rate relative to false alarm rate. For each participant, overall task performance was computed by taking the average d' of all task blocks.

Replicating Rosenberg et al. (43) with Yeo atlas ROIs, we employed connectome-based predictive modeling (65) to predict individuals' d' scores. In a LOO cross-validation approach, we selected FCs that were significantly correlated with training participants' d' scores (Spearman's correlation, P < 0.01) (65). We separated the FCs that were positively or negatively correlated with d' and averaged the strength of positively correlated FCs and negatively correlated FCs. A multiple linear regression model was trained to learn the coefficients, where the dependent variable was individuals' d' scores, and the two independent variables were the average of positively and negatively correlated FCs, respectively. The model predicted the held-out participant's d' score, and the predicted scores of all cross-validations were validated by taking the Pearson's r with the observed scores. Replicating findings from Rosenberg et al. (43), FC matrices using Yeo atlas significantly predicted individuals' sustained attention scores (r = 0.801, P < 0.001; comparable to r = 0.822, P < 0.001 when using Shen atlas' 268-ROI parcellation).

The set of FCs selected in every cross-validation is termed the sustained attention network. Dynamic strength of the sustained attention network was computed by taking the difference between average FC time courses of positively correlated FCs and the negatively correlated FCs, calculated with the same sliding window analysis as above.

Event Recall. For *Paranoia*, each phrase of a participant's recall was manually matched to a sentence in the story transcript by the first author, which was annotated with timestamps of when that sentence was heard by participants during fMRI. We vectorized all words included in the transcript and recall data with the predefined, distributed word embedding, GloVe, which is trained on Wikipedia and Gigaword 5 corpus (dimension = 100) (69). Sentence embeddings were generated with the method adopted from Arora et al. (70), such that the embedding vectors of every word included in the recall phase were averaged, weighted by the smooth inverse frequency of the word over a total number of words in the annotation or participant-specific recalls. Conceptually, this gives higher weights to infrequently occurring words when creating a sentence embedding vector. The smoothing parameter was set to $\alpha = 0.0001$, and we did not subtract out singular vector (i.e., a common component) for comparison purposes. The cosine similarity between the story and recall sentence embeddings was multiplied by the binary index of whether a participant recalled the event or not, which represents "recall fidelity" of *Paranoia* participants. The output recall fidelity scores were extended in time to TR space so as to match the moments when a sentence was uttered in the narration.

For *Sherlock*, we applied dynamic topic model (71) together with hidden Markov model (11) as implemented by Heusser et al. (72). We followed the hyperparameter selections and experimental steps of Heusser et al. (72). We calculated a Pearson's *r* similarity matrix between the annotated video event embeddings (30 events detected by a hidden Markov model) and recall event embeddings (the number of events differed for each participant). The mean of every recall event's similarity (Fisher's *z*–transformed Pearson's correlation) with each of the video events was averaged to generate an indicator of "recall fidelity." This approach differs from Heusser et al.'s (72) recall precision and distinctiveness metrics, which capture whether participants' recall events precisely and specifically match the topic contents of the movie events. Specifically, recall precision was calculated by the maximum correlation values between the topic vector of every recalled event that was matched one-on-one with the movie event. In contrast, our recall fidelity metric captures whether participants' recalls were similar to the annotated contents of the narrative stimulus, irrespective of precision to specific movie events. Thus, we averaged the similarities of the entire recall event with every moment of the movie event (average of every row in Fig. 6*B*), which was expanded into a TR space to generate a recall fidelity time course that matches the total duration of stimulus presentation.

To test the significance of recall prediction, we generated surrogate data by randomly shuffling the event-specific recall fidelity scores for each individual. Then, using the same event-to-TR mapping, we expanded the event-wise recall fidelity scores to a recall fidelity time course in a TR space. This not only retains the same range of recall scores per individual but also retains the event boundaries. The HRF convolution and sliding time window were applied to surrogate recall data, which were used with actual fMRI data to train and test SVR models.

1. W. James, *The Principles of Psychology* (Henry Holt and Company, New York, 1890).
2. M. M. Chun, J. D. Golomb, N. B. Turk-Browne, A taxonomy of external and internal attention. *Annu. Rev. Psychol.* **62**, 73–101 (2011).
3. M. Esterman, D. Rothlein, Models of sustained attention. *Curr. Opin. Psychol.* **29**, 174–180 (2019).
4. M. T. deBettencourt, K. A. Norman, N. B. Turk-Browne, Forgetting from lapses of sustained attention. *Psychon. Bull. Rev.* **25**, 605–611 (2018).
5. I. H. Robertson, T. Manly, J. Andrade, B. T. Baddeley, J. Yiend, 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* **35**, 747–758 (1997).
6. M. Rosenberg, S. Noonan, J. DeGutis, M. Esterman, Sustaining visual attention in the face of distraction: A novel gradual-onset continuous performance task. *Atten. Percept. Psychophys.* **75**, 426–439 (2013).
7. A. Kucyi, M. Esterman, C. S. Riley, E. M. Valera, Spontaneous default network activity reflects behavioral variability independent of mind-wandering. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13899–13904 (2016).
8. M. Csikszentmihalyi, J. Nakamura, "Effortless attention in everyday life: A systematic phenomenology" in *Effortless Attention: A New Perspective in the Cognitive Science of Attention and Action*, B. Bruya, Ed. (MIT Press, Cambridge, MA, 2010), pp. 179–189.
9. R. Busselle, H. Bilandzic, Measuring narrative engagement. *Media Psychol.* **12**, 321–347 (2009).
10. H. Bilandzic, F. Sukalla, C. Schnell, M. R. Hastall, R. W. Busselle, The narrative engageability scale: A multidimensional trait measure for the propensity to become engaged in a story. *Int. J. Commun.* **13**, 801–832 (2019).
11. C. Baldassano *et al.*, Discovering event structure in continuous narrative perception and memory. *Neuron* **95**, 709–721.e5 (2017).
12. C. H. C. Chang, C. Lazaridi, Y. Yeshurun, K. A. Norman, U. Hasson, Relating the past with the present: Information integration and segregation during ongoing narrative processing. *J. Cogn. Neurosci.* **33**, 1–23 (2021).
13. J. Chen *et al.*, Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).
14. U. Hasson, E. Yang, I. Vallines, D. J. Heeger, N. Rubin, A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* **28**, 2539–2550 (2008).
15. C. J. Honey *et al.*, Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* **76**, 423–434 (2012).
16. Y. Lerner, C. J. Honey, L. J. Silbert, U. Hasson, Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* **31**, 2906–2915 (2011).
17. Y. Yeshurun *et al.*, Same story, different story. *Psychol. Sci.* **28**, 307–319 (2017).
18. E. S. Finn, P. R. Corlett, G. Chen, P. A. Bandettini, R. T. Constable, Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nat. Commun.* **9**, 2043 (2018).
19. L. Nummenmaa *et al.*, Emotions promote social interaction by synchronizing brain activity across individuals. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9599–9604 (2012).
20. D. C. Gruskin, M. D. Rosenberg, A. J. Holmes, Relationships between depressive symptoms and brain responses during emotional movie viewing emerge in adolescence. *Neuroimage* **216**, 116217 (2020).
21. L. J. Chang *et al.*, Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Sci. Adv.* **7**, eabf7129 (2021).
22. M. Regev *et al.*, Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. *Cereb. Cortex* **29**, 4017–4034 (2019).
23. T. Çukur, S. Nishimoto, A. G. Huth, J. L. Gallant, Attention during natural vision warps semantic representation across the human brain. *Nat. Neurosci.* **16**, 763–770 (2013).
24. J. J. Ki, S. P. Kelly, L. C. Parra, Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *J. Neurosci.* **36**, 3092–3101 (2016).
25. R. Schmälzle, F. E. K. Häcker, C. J. Honey, U. Hasson, Engaged listeners: Shared neural processing of powerful political speeches. *Soc. Cogn. Affect. Neurosci.* **10**, 1137–1143 (2015).
26. J. P. Dmochowski *et al.*, Audience preferences are predicted by temporal reliability of neural processing. *Nat. Commun.* **5**, 4567 (2014).
27. J. P. Dmochowski, P. Sajda, J. Dias, L. C. Parra, Correlated components of ongoing EEG point to emotionally laden attention—A possible marker of engagement? *Front. Hum. Neurosci.* **6**, 112 (2012).
28. U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, R. Malach, Intersubject synchronization of cortical activity during natural vision. *Science* **303**, 1634–1640 (2004).
29. S. S. Cohen, L. C. Parra, Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *eNeuro* **3**, 1–11 (2016).
30. U. Hasson, O. Furman, D. Clark, Y. Dudai, L. Davachi, Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron* **57**, 452–462 (2008).
31. M. Corbetta, G. L. Shulman, Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 201–215 (2002).
32. M. I. Posner, S. E. Petersen, The attention system of the human brain. *Annu. Rev. Neurosci.* **13**, 25–42 (1990).
33. M. D. Fox *et al.*, The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9673–9678 (2005).
34. M. E. Raichle, The brain's default mode network. *Annu. Rev. Neurosci.* **38**, 433–447 (2015).
35. R. L. Buckner, L. M. DiNicola, The brain's default network: Updated anatomy, physiology and evolving insights. *Nat. Rev. Neurosci.* **20**, 593–608 (2019).
36. M. Esterman, S. K. Noonan, M. Rosenberg, J. Degutis, In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cereb. Cortex* **23**, 2712–2723 (2013).
37. M. Esterman, M. D. Rosenberg, S. K. Noonan, Intrinsic fluctuations in sustained attention and distractor processing. *J. Neurosci.* **34**, 1724–1730 (2014).

Song et al.
Neural signatures of attentional engagement during narratives and its consequences for event memory

38. A. Yamashita, D. Rothlein, A. Kucyi, E. M. Valera, M. Esterman, Brain state-based detection of attentional fluctuations and their modulation. *Neuroimage* **236**, 118072 (2021).

39. H. Song, B. Park, H. Park, W. M. Shim, Cognitive and neural state dynamics of story comprehension. *bioRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.07.10.194647 (Accessed 18 July 2020).

40. Y. Yeshurun, M. Nguyen, U. Hasson, The default mode network: Where the idiosyncratic self meets the shared social world. *Nat. Rev. Neurosci.* **22**, 181–192 (2021).

41. T. J. Buschman, S. Kastner, From behavior to neural dynamics: An integrated theory of attention. *Neuron* **88**, 127–144 (2015).

42. M. D. Rosenberg *et al.*, Characterizing attention with predictive network models. *Trends Cogn. Sci.* **21**, 290–302 (2017).

43. M. D. Rosenberg *et al.*, A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci.* **19**, 165–171 (2016).

44. M. D. Rosenberg *et al.*, Functional connectivity predicts changes in attention observed across minutes, days, and months. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 3797–3807 (2020).

45. S. A. Nastase, V. Gazzola, U. Hasson, C. Keysers, Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 667–685 (2019).

46. H. Song, M. D. Rosenberg, Predicting attention across time and contexts with functional brain connectivity. *Curr. Opin. Behav. Sci.* **40**, 33–44 (2021).

47. M. A. Gernsbacher, K. R. Varner, M. E. Faust, Investigating differences in general comprehension skill. *J. Exp. Psychol. Learn. Mem. Cogn.* **16**, 430–445 (1990).

48. H. Song, E. S. Finn, M. D. Rosenberg, Narrative Engagement. GitHub. https://github.com/hyssong/NarrativeEngagement. Deposited 31 May 2021.

49. Y. R. Tausczik, J. W. Pennebaker, The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* **29**, 24–54 (2010).

50. K. V. Nourski *et al.*, Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* **29**, 15564–15574 (2009).

51. J. Antony *et al.*, Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *Neuron* **109**, 377–390.e7 (2020).

52. E. Simony *et al.*, Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* **7**, 12141 (2016).

53. E. S. Finn *et al.*, Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *Neuroimage* **215**, 116828 (2020).

54. E. A. Allen *et al.*, Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* **24**, 663–676 (2014).

55. R. Liégeois *et al.*, Cerebral functional connectivity periodically (de)synchronizes with anatomical constraints. *Brain Struct. Funct.* **221**, 2985–2997 (2016).

56. L. Deng, J. Sun, L. Cheng, S. Tong, Characterizing dynamic local functional connectivity in the human brain. *Sci. Rep.* **6**, 26976 (2016).

57. W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, M. D. Greicius, Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb. Cortex* **22**, 158–165 (2012).

58. B. T. T. Yeo, J. Tandi, M. W. L. Chee, Functional connectivity during rested wakefulness predicts vulnerability to sleep deprivation. *Neuroimage* **111**, 147–158 (2015).

59. B. T. T. Yeo *et al.*, The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

60. L. Fan *et al.*, The Human Brainnetome Atlas: A new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).

61. X. Shen, F. Tokoglu, X. Papademetris, R. T. Constable, Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* **82**, 403–415 (2013).

62. J. D. Haynes, A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron* **87**, 257–270 (2015).

63. M. D. Rosenberg, E. S. Finn, R. T. Constable, M. M. Chun, Predicting moment-to-moment attentional state. *Neuroimage* **114**, 249–256 (2015).

64. M. T. deBettencourt, J. D. Cohen, R. F. Lee, K. A. Norman, N. B. Turk-Browne, Closed-loop training of attention with real-time brain imaging. *Nat. Neurosci.* **18**, 470–475 (2015).

65. X. Shen *et al.*, Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* **12**, 506–518 (2017).

66. M. Xia, J. Wang, Y. He, BrainNet viewer: A network visualization tool for human brain connectomics. *PLoS One* **8**, e68910 (2013).

67. E. F. Risko, N. Anderson, A. Sarwal, M. Engelhardt, A. Kingstone, Everyday attention: Variation in mind wandering and memory in a lecture. *Appl. Cogn. Psychol.* **26**, 234–242 (2012).

68. J. Smallwood, M. McSpadden, J. W. Schooler, When attention matters: The curious incident of the wandering mind. *Mem. Cognit.* **36**, 1144–1150 (2008).

69. J. Pennington, R. Socher, C. D. Manning, "GloVe: Global vectors for word representation" in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, W. Daelemans, Eds. (Association for Computational Linguistics, Stroudsburg, PA, 2014), pp. 1532–1543.

70. S. Arora, Y. Liang, T. Ma, "A simple but tough-to-beat baseline for sentence embeddings" in *Proceedings of International Conference on Learning Representations (ICLR)* (Toulon, France, 2016), pp. 416–424.

71. D. M. Blei, J. D. Lafferty, "Dynamic topic models" in *Proceedings of the 23rd International Conference on Machine Learning (ICML)* (International Machine Learning Society, La Jolla, CA, 2006), pp. 113–120.

72. A. C. Heusser, P. C. Fitzpatrick, J. R. Manning, Geometric models reveal behavioural and neural signatures of transforming experiences into memories. *Nat. Hum. Behav.* 10.1038/s41562-021-01051-6 (2021).

73. B. J. Meyer, G. W. McConkie, What is recalled after hearing a passage? *J. Educ. Psychol.* **65**, 109–117 (1973).

74. P. W. Thorndyke, F. R. Yekovich, A critique of schema-based theories of human story memory. *Poetics* **9**, 23–49 (1980).

75. N. H. Mackworth, The breakdown of vigilance during prolonged visual search. *Q. J. Exp. Psychol.* **1**, 6–21 (1948).

76. R. Parasuraman, Memory load and event rate control sensitivity decrements in sustained attention. *Science* **205**, 924–927 (1979).

77. J. S. Warm, R. Parasuraman, G. Matthews, Vigilance requires hard mental work and is stressful. *Hum. Factors* **50**, 433–441 (2008).

78. D. Kahneman, *Attention and Effort* (University of Michigan, Prentice-Hall, 1973).

79. F. C. Fortenbaugh, D. Rothlein, R. McGlinchey, J. DeGutis, M. Esterman, Tracking behavioral and neural fluctuations during sustained attention: A robust replication and extension. *Neuroimage* **171**, 148–164 (2018).

80. A. Kucyi *et al.*, Electrophysiological dynamics of antagonistic brain networks reflect attentional fluctuations. *Nat. Commun.* **11**, 325 (2020).

81. S. Kaplan, M. G. Berman, Directed attention as a common resource for executive functioning and self-regulation. *Perspect. Psychol. Sci.* **5**, 43–57 (2010).

82. S. Marković, Components of aesthetic experience: Aesthetic fascination, aesthetic appraisal, and aesthetic emotion. *Iperception* **3**, 1–17 (2012).

83. K. Christoff, A. M. Gordon, J. Smallwood, R. Smith, J. W. Schooler, Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8719–8724 (2009).

84. J. Smallwood, J. W. Schooler, The science of mind wandering: Empirically navigating the stream of consciousness. *Annu. Rev. Psychol.* **66**, 487–518 (2015).

85. C. Braboszcz, A. Delorme, Lost in thoughts: Neural markers of low alertness during mind wandering. *Neuroimage* **54**, 3040–3047 (2011).

86. K. A. Garrison *et al.*, Real-time fMRI links subjective experience with brain activity during focused attention. *Neuroimage* **81**, 110–118 (2013).

87. F. I. M. Craik, R. S. Lockhart, Levels of processing: A framework for memory research. *J. Verbal Learn. Verbal Behav.* **11**, 671–684 (1972).

88. P. Seli, J. A. Cheyne, K. R. Barton, D. Smilek, Consistency of sustained attention across modalities: Comparing visual and auditory versions of the SART. *Can. J. Exp. Psychol.* **66**, 44–50 (2012).

89. H. Terashima, K. Kihara, J. I. Kawahara, H. M. Kondo, Common principles underlie the fluctuation of auditory and visual sustained attention. *Q J Exp Psychol (Hove)* **74**, 705–715 (2021).

90. M. Zhang, N. Savill, D. S. Margulies, J. Smallwood, E. Jefferies, Distinct individual differences in default mode network connectivity relate to off-task thought and text memory during reading. *Sci. Rep.* **9**, 16220 (2019).

91. D. C. Jangraw *et al.*, A functional connectivity-based neuromarker of sustained attention generalizes to predict recall in a reading task. *Neuroimage* **166**, 99–109 (2018).

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

93. Q. Lu, U. Hasson, K. A. Norman, Learning to use episodic memory for event prediction. *bioRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.12.15.422882 (Accessed 20 December 2020).

94. F. Sukalla, H. Bilandzic, P. D. Bolls, R. W. Busselle, Embodiment of narrative engagement: Connecting self-reported narrative engagement to psychophysiological measures. *J. Media Psychol. Theor. Methods Appl.* **28**, 175–186 (2016).

95. D. C. Richardson *et al.*, Engagement in video and audio narratives: Contrasting self-report and physiological measures. *Sci. Rep.* **10**, 11298 (2020).

96. J. H. Cheong, Z. Molani, S. Sadhukha, L. J. Chang, Synchronized affect in shared experiences strengthens social connection. *PsyArXiv* [Preprint] (2020) https://doi.org/10.31234/osf.io/bd9wn (Accessed 3 November 2020).

97. N. Unsworth, M. K. Robison, Pupillary correlates of lapses of sustained attention. *Cogn. Affect. Behav. Neurosci.* **16**, 601–615 (2016).

98. D. Smilek, J. S. A. Carriere, J. A. Cheyne, Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychol. Sci.* **21**, 786–789 (2010).

99. A. Maffei, A. Angrilli, Spontaneous blink rate as an index of attention and emotion during film clips viewing. *Physiol. Behav.* **204**, 256–263 (2019).

100. J. N. V. Meer, M. Breakspear, L. J. Chang, S. Sonkusare, L. Cocchi, Movie viewing elicits rich and reliable brain state dynamics. *Nat. Commun.* **11**, 5004 (2020).

101. J. W. Peirce, PsychoPy-Psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).

102. E. S. Finn, P. R. Corlett, G. Chen, P. A. Bandettini, R. T. Constable, Data from "ParanoiaStory." OpenNeuro database. https://openneuro.org/datasets/ds001338. Accessed 5 August 2019.

103. J. Chen *et al.*, Data from "Sherlock Movie Watching Dataset." Princeton University's DataSpace. https://dataspace.princeton.edu/jspui/handle/88435/dsp01nz8062179. Accessed 26 September 2019.

104. B. Fischl, FreeSurfer. *Neuroimage* **62**, 774–781 (2012).

105. S. Vossel, J. J. Geng, G. R. Fink, Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *Neuroscientist* **20**, 150–159 (2014).

106. D. S. Margulies *et al.*, Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12574–12579 (2016).

107. B. T. T. Yeo *et al.*, Data from "Yeo JNeurophysiol. Split Labels (MNI152)." GitHub. https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Yeo2011_fcMRI_clustering/1000subjects_reference/Yeo_JNeurophysiol11_SplitLabels/MNI152. Accessed 17 February 2018.