Behavioral/Cognitive

Cognitive and Neural State Dynamics of Narrative Comprehension

Hayoung Song,[1,2,3] Bo-yong Park,[1,4,5,6] Hyunjin Park,[1,7] and Won Mok Shim[1,2,8]

[1]Center for Neuroscience Imaging Research, IBS, Suwon, Korea, 16419, [2]Department of Biomedical Engineering, Sungkyunkwan University, Suwon, Korea, 16419, [3]Department of Psychology, University of Chicago, Chicago, Illinois, 60637, [4]Department of Electronic, Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea, 16419, [5]McConnell Brain Imaging Centre, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec Canada, H3A 2B4, [6]Department of Data Science, Inha University, Incheon, Korea, 22201, [7]School of Electronics and Electrical Engineering, Sungkyunkwan University, Suwon, Korea, 16419, and [8]Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, Korea, 16419

Narrative comprehension involves a constant interplay of the accumulation of incoming events and their integration into a coherent structure. This study characterizes cognitive states during narrative comprehension and the network-level reconfiguration occurring dynamically in the functional brain. We presented movie clips of temporally scrambled sequences to human participants (male and female), eliciting fluctuations in the subjective feeling of comprehension. Comprehension occurred when processing events that were highly causally related to the previous events, suggesting that comprehension entails the integration of narratives into a causally coherent structure. The functional neuroimaging results demonstrated that the integrated and efficient brain state emerged during the moments of narrative integration with the increased level of activation and across-modular connections in the default mode network. Underlying brain states were synchronized across individuals when comprehending novel narratives, with increased occurrences of the default mode network state, integrated with sensory processing network, during narrative integration. A model based on time-resolved functional brain connectivity predicted changing cognitive states related to comprehension that are general across narratives. Together, these results support adaptive reconfiguration and interaction of the functional brain networks on causal integration of the narratives.

Key words: **causality; cognitive neuroscience; default mode network; fMRI; functional connectivity; narrative comprehension**

Significance Statement

The human brain can integrate temporally disconnected pieces of information into coherent narratives. However, the underlying cognitive and neural mechanisms of how the brain builds a narrative representation remain largely unknown. We showed that comprehension occurs as the causally related events are integrated to form a coherent situational model. Using fMRI, we revealed that the large-scale brain states and interaction between brain regions dynamically reconfigure as comprehension evolves, with the default mode network playing a central role during moments of narrative integration. Overall, the study demonstrates that narrative comprehension occurs through a dynamic process of information accumulation and causal integration, supported by the time-varying reconfiguration and brain network interaction.

Received Jan. 7, 2021; revised Sep. 3, 2021; accepted Sep. 7, 2021.

Author contributions: H.S. and W.M.S. designed research; H.S. and W.M.S. performed research; H.S., B.-y. P., H.P., and W.M.S. contributed unpublished reagents/analytic tools; H.S. analyzed data; H.S. wrote the first draft of the paper; H.S., B.-y.P., H.P., and W.M.S. edited the paper; H.S., and W.M.S. wrote the paper.

This work was supported by Institute for Basic Science Grant IBS R015-D1 and Korean government National Research Foundation of Korea Grants NRF-2019M3E5D2A01060299 and NRF-2019R1A2C1085566. Behavioral experiment data are available at https://github.com/hyssong/comprehension and fMRI data are available at https://doi.org/10.5281/zenodo.5108941. We thank Youngmin Jeon for assistance with behavioral data collection; Boohee Choi for technical support in fMRI data collection; Jeongjun Park, Monica D. Rosenberg, Hun Seok Choi, and Woochul Choi for constructive feedback on the manuscript; Oliver James, Choong-wan Woo, and Janice Chen for feedback on data analyses; and two anonymous reviewers for providing helpful comments on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Hayoung Song at hyssong@uchicago.edu or Won Mok Shim at wonmokshim@skku.edu.

https://doi.org/10.1523/JNEUROSCI.0037-21.2021

Introduction

We make sense of our memory and others' behavior by constantly constructing narratives from an information stream that unfolds over time. Comprehending a narrative is a process of accumulating ongoing information, storing it in memory as a situational model, and simultaneously integrating it to construct a coherent representation (Zwaan et al., 1995; Langston and Trabasso, 1999; Polyn et al., 2009; Ranganath and Ritchey, 2012). Forming a coherent representation of a narrative involves comprehending the causal structure of the events, including the causal flow that links consecutive events or even a long-range causal connection that exists between temporally discontiguous events. Past research theorized that narrative comprehension requires reinstating causally related past events and integrating

them into a structured representation (Trabasso and Sperry, 1985; Graesser et al., 1994; Chang et al., 2021). However, empirical evidence regarding the integration of causal relations related to the ongoing process of comprehension is lacking.

Recent neuroscientific literature suggests that narratives are represented in activation patterns (Baldassano et al., 2017; Chen et al., 2017) and functional connectivity (FC) (Simony et al., 2016; Aly et al., 2018; Ritchey and Cooper, 2020) of the distributed regions in the default mode network (DMN), based on their capacity to integrate information over prolonged periods (Hasson et al., 2008; Lerner et al., 2011; Honey et al., 2012). However, traditional cognitive models theorize that the representation of narratives is updated by the interaction of the broader networks of the whole brain, including regions involved in sensory processing, memory, and cognitive control (Mar, 2004). Prior research has indicated that large-scale brain networks alternate between functionally segregated and integrated states (Tononi et al., 1994; Shine et al., 2016), depending on the information processing that is adaptively recruited at the moment (Bullmore and Sporns, 2012; Zalesky et al., 2014). Studies reported that brain activity, FC, and the occurrences and transitions of the large-scale brain states were synchronized as participants watched the same movies, which were reliably coupled to the narratives (Simony et al., 2016; Nastase et al., 2019; Betzel et al., 2020; van der Meer et al., 2020). However, how these synchronized reconfigurations in large-scale brain networks are related to the ongoing process of narrative comprehension remain unclear.

Previous work theorized external and internal modes of information processing in the brain, illustrating how the brain undergoes adaptive state changes between the accumulation of information from the external world, and its integration into internal thoughts (Dixon et al., 2014; Honey et al., 2018). Through the adoption of this framework, this study characterizes narrative comprehension as an exemplary naturalistic cognitive process that entails transitions between dual modes of information processing (external and internal modes) that are accompanied by corresponding state changes in large-scale functional networks (segregated and integrated states). We hypothesize that the relative proportion of the two distinct processing modes would vary depending on an individual's degree of comprehension over time. Specifically, we assume that, when a person experiences a high degree of comprehension (i.e., when narratives are being internally integrated to form a coherent situational model), an integrated state of functional networks would emerge, with tight connections between higher-order, transmodal brain networks and sensory-specific, unimodal brain networks (Mesulam, 1998; Margulies et al., 2016; Murphy et al., 2018). In contrast, when individuals have a lesser degree of comprehension and thus focusing on processing external inputs, we predict that the functional brain network would be biased toward a segregated state, where each module operates independently.

Here, we characterized the cognitive processes involved in narrative comprehension and examined the dynamic reconfiguration of large-scale functional networks during comprehension. To track cognitive state changes, we presented three movie clips of temporally scrambled sequences and collected behavioral responses when individuals experienced the subjective feeling of comprehension. Depending on participants' responses on moments of comprehension, we characterized a group-aggregate behavior measure that represents fluctuating states of comprehension during scrambled movie watching. In a separate behavioral study, we measured causal relationships between pairwise moments of these scrambled movies and observed that high-comprehension moments correspond to the moments when past events are causally connected to the present event, suggesting that comprehension entails the integration of narratives into a causally coherent structure. In a functional neuroimaging study, using the group measure of comprehension, we inferred the cognitive state dynamics of independent participants as they watched the same scrambled movies inside the scanner. We observed an increased level of BOLD activity in the DMN regions during narrative integration, whereas the dorsal attention network (DAN) increased its BOLD responses when comprehension was low. On a larger network scale, the functional brain entered an integrated state during narrative integration, which was modulated by the across-modular functional connections of the DMN and frontoparietal control network (FPN). Using hidden Markov modeling (HMM) to characterize latent neural states (Rabiner and Juang, 1986), we identified synchronized neural state dynamics across individuals during novel movie watching, with the DMN, integrated with the sensory processing network, being the dominant state during high-comprehension moments. We further demonstrated that evolving cognitive states of comprehension that are robust across narratives can be predicted from time-varying functional connections between brain regions, but not from patterns of regional BOLD activity, suggesting that the functional interaction of the distributed brain regions is involved in narrative integration and comprehension. By characterizing narrative comprehension as a dynamic process of causal integration and relating it to changes in large-scale brain activity and connectivity, our study provides the basis of cognitive and neural states that underlie real-world information processing.

Materials and Methods

Participants. An independent group of individuals participated in two behavioral and one fMRI experiment (Behavioral Experiment 1: 20 participants per movie, with a total of 27 participants; 5 women, mean age 22.6 ± 2.1 years; Behavioral Experiment 2: 12 participants per movie, with a total of 29 participants; 10 women, mean age 21.6 ± 2.2 years; fMRI experiment: 24 participants for *Cops*, 23 participants for *The Kid*, and 20 participants for *Mr. Bean*, with a total of 30 participants; 10 women, mean age 24 ± 2.1 years). The number of participants was determined based on previous fMRI studies that used similar naturalistic task paradigms (Chen et al., 2017; Aly et al., 2018; Baldassano et al., 2018; Finn et al., 2018). A number of individuals participated in the experiment multiple times watching different movie stimuli. None of the participants had watched the movies before the experiment. All but one participant were native Korean. All participants in the fMRI study were right-handed, except one. Participants reported no history of visual, hearing, or any form of neurologic impairment. The participants provided informed consent before taking part in the study and were monetarily compensated. The study was approved by the Institutional Review Board of Sungkyunkwan University.

Movie stimuli. Three movie clips: *Cops* (1922, Keaton & Cline), *The Kid* (1921, Chaplin), and *Mr. Bean: The Animated Series, Art Thief* (season 2, episode 13; 2003, Fehrenbach), were used in both behavioral and fMRI experiments. The three movies were selected as they comprised rich narratives within ~10 min duration and did not contain any form of verbal conversations or narration (with an exception of background music). Since narratives were delivered in the visual modality, the viewers had to actively infer the narratives through characters' facial expressions, body movements, and changes in the backgrounds. The movies *Cops* and *The Kid*, which were black-and-white silent pictures from the early 1900s, contained six and five clips, respectively, that projected a dialogue on a full screen. These dialogues were translated into Korean and projected in the same manner. All three movies contained

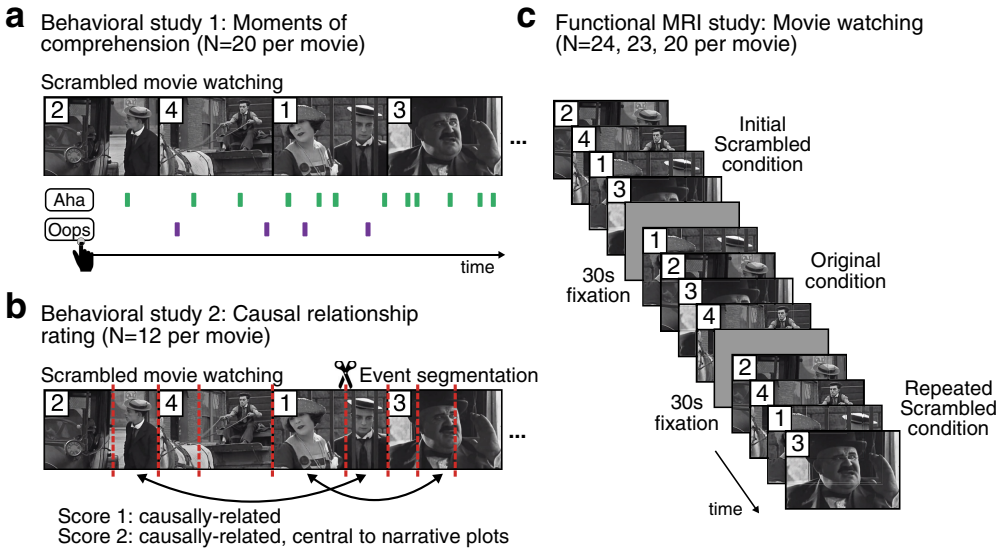

Figure 1. Experiment design. ***a***, Behavioral study 1: Reports on the subjective moments of comprehension of the narratives. As participants watched a temporally scrambled movie (three movie stimuli; $N = 20$ per movie), they were instructed to press the "Aha" button whenever they experienced subjective comprehension of the plot (green), and the "Oops" button whenever they realized that their previous understanding was incorrect (purple). The duration of each audiovisual movie was 10 min, and the duration of each scene was 36 $\pm$ 4 s. ***b***, Behavioral study 2: Event segmentation and causal relationship rating between events. In the first part of the study, an independent group of participants ($N = 12$ per movie) watched the movie in a temporally scrambled sequence, followed by the original sequence. In the second part, they were instructed to mark perceived event boundaries to the scrambled movie (red dashed lines) and to annotate each event with a short description. In the third part, they were instructed to rate the degree of a causal relationship between pairwise events they segmented themselves (bidirectional arrows). A score of 1 was given if pairwise events were thought to be causally related, a score of 2 when pairwise events held critical causal importance within the narrative plots, and a score of 0 was given otherwise. ***c***, fMRI study. Another independent group of participants ($N = 24$, 23, and 20 per movie) participated in the fMRI experiment, where they watched the same movie in an Initial Scrambled, Original, and Repeated Scrambled conditions in a single scan run. The conditions were separated by a 30 s fixated rest. No behavioral response was collected during the scan.

background music, which sets the emotional tone of the narratives and characters. The movies were edited to an exact 10 min version where a coherent narrative was complete within the given time. Each movie was segmented into 16 (*The Kid*) or 17 (*Cops* and *Mr. Bean*) scenes and shuffled in pseudorandom order, such that the fluctuations in the subjective feeling of comprehension could be maximally induced. The scenes were segmented mostly following the director's cut (e.g., changes in background or camera angle), and the scene duration ranged between 32 and 40 s. To match the scene duration to the sampling rate of the fMRI sequence (TR = 1 s), we minimally adjusted the speed of the scenes by rounding the duration to the nearest second. None of the participants reported perceived differences in the speed of the scenes.

Behavioral Experiment 1: reports on the subjective moments of comprehension. We collected behavioral responses while participants were watching a 10 min movie in a scrambled sequence (Scrambled movie watching) (Fig. 1*a*). The stimuli were presented by the GStreamer library (Open-Source, 2014), and the responses were recorded using MATLAB (The MathWorks) and Psychtoolbox (Brainard, 1997; Pelli, 1997). The experiment was conducted in a dimly lit room where the movies were presented on a CRT monitor. Before the experiment, participants participated in a practice session with a different movie clip, *Oggy and the Cockroaches: The Animated Series, Panic Room* (season 4, episode 8; 2013, Jean-Marie). Participants pressed an "Aha" button when they thought that they had comprehended the narrative, specifically when they comprehended the temporal sequence or causal relationship of the original narrative or when an interim comprehension of previously presented events occurred. On the other hand, they were instructed to press an "Oops" button when they realized their prior comprehension was incorrect. Participants were told that their reports of comprehension did not necessarily have to be correct. Rather, they were instructed to report whenever they experienced subjective feelings of "Aha" or "Oops," at moments of sudden insight or when their comprehension had changed. As a *post hoc* verification of comprehension, participants completed a comprehension quiz about the plots and contents of the narrative. The data of 2 participants who scored exceptionally low were excluded from the analyses.

As both "Aha" and "Oops" characterize moments when participants experienced subjective feelings of comprehension, no distinction was made between the two response types in the analyses. The moments of a button press were resampled to a 1 s interval. The intersubject similarity of the resampled button-press moments was calculated by averaging Dice coefficients for all pairwise participants. A window of 4 s centered around the time of a button press was considered as a button-press moment (Baldassano et al., 2017). The number of overlapping button-press moments between pairwise participants was divided by the mean of the total button presses of the pairwise participants. Nonparametric permutation tests were conducted on the Dice coefficients computed from the same number of randomly shuffled button presses of every participant, also using a 4 s window (one-tailed test, 10,000 iterations).

All participants' button responses (sampled at 1 s) were convolved with a canonical HRF to relate to the independent group of participants' fMRI data. To represent a gradual change in cognitive states related to narrative comprehension that is shared across individuals, we applied a sliding-window analysis with a window size of 36 s and a step size of 1 s to aggregate (i.e., summation) the convolved behavioral responses of all participants. Since this results in a time duration that is 36 s deduced from the initial duration (movie duration minus the sliding-window size), we padded 18 s (half the size of a sliding window) of zeros at the beginning and end and applied the same sliding-window analysis to generate a behavioral index that matches the movie duration. We initially chose a window size of 36 s to match the average duration of each scene segment in the scrambled movies; however, we replicated our results with the window sizes of 24, 30, and 42 s. The HRF-convolved, sliding-window-applied output time course represented group-aggregate continuous behavioral measures of comprehension for each narrative stimulus.

We also generated binary behavioral measures by categorizing each time step of the movie into moments when participants experienced a generally high or low degree of comprehension. The top one-third of the moments with the high number of aggregate button responses were labeled as the moments of "high comprehension" and the bottom one-third were labeled as "low comprehension." We discarded the middle

one-third because of its susceptibility to individual variances in cognitive states. With an assumption that cognitive states of comprehension are not instantaneous but prolonged, the labeled moments were discarded if they did not persist for at least 10 consecutive time points (10 s). The number of discarded time points was small, which amounted to 6.02 ± 5.72% of the one-third splits of the total time points.

For a nonparametric permutation test, we conducted phase-randomization of the behavioral time course after HRF convolution. A sliding window with the same parameters and zero paddings was applied to the phase-randomized time courses.

Behavioral Experiment 2: rating causal relationship between narrative events. We collected reports of a causal relationship between pairwise events of the scrambled movies. The stimuli presentation and response recording were controlled with Adobe Premiere Pro CC (Adobe Systems). Participants initially watched the movie in scrambled and original orders. Then, they were asked to segment the scrambled movie in terms of the events' narrative contents, by marking perceived event boundaries without limit on the total number (Zacks and Swallow, 2007; Baldassano et al., 2017) (Fig. 1*b*). The experiment was conducted using the scrambled movie only. Participants freely swiped through the movie on the video editor during the event segmentation, marking exact moments in time at which they perceived event boundaries. The scene boundaries created from temporal scrambling were marked as physical event boundaries in the video editor before the experiment. Thus, the scene segments from experimental manipulation were shared across all participants, while additional perceived event segments differed across participants. Next, participants were instructed to write a short description of each segmented event. Finally, using the descriptions of the events, participants were asked to rate the degree of a causal relationship between all possible pairwise events, on a scale of 0-2 (Fig. 1*b*). A pair of events was rated with a score of 1 if one event was causally attributed to, or explained by, the happening of another event. A pair was rated with a score of 2 if a causal relationship between pairwise events played a major role in developing the plot of the narratives. The pairwise events that were not rated by the participants were automatically scored 0. Participants performed the task at their own pace without a time limit.

The timing of the perceived event boundaries was rounded to a 1 s sampling rate. The causal relationship score originally given to the pairwise events was assigned to the corresponding pairwise moments of the narratives, considering the duration of participant-specific perceived event. By summing all participants' ratings, a causal relationship matrix was constructed for each scrambled movie, which specified the degree of a causal relationship between all pairwise moments of the movie. We unscrambled the event sequence back to the original order to further visualize causal relationships in an original event sequence. Importantly, the causal relationship score of each moment was computed by averaging the past moments' causal relationship scores with respect to the present moment. A causal relationship score of the nearby past moments that corresponded to the same scene (i.e., a scene of 36 ±4 s duration that was segmented in temporal scrambling) was not included in the analysis. A high causal relationship score represents that an event is highly causally related to the events that occurred in the past, suggesting that the corresponding moment is important in the narrative context and that the past events are more likely to be reinstated in memory and subsequently integrated into ongoing narratives while processing the event.

To relate the causal relationship to the group measure of narrative comprehension, the same analysis steps were applied to the causal relationship time course (HRF convolution and sliding-window analysis).

Control analysis: separating the effect of semantic relationship from the causal relationship of the narrative events. As a control analysis of the causal relationship experiment, we measured the degree of semantic relationship between all pairwise events of the scrambled movies. Written annotations of the narrative contents were generated for every 2 s of the scrambled movies by four independent annotators (4 women, mean age 24.5 ± 1.3 years) with native-level English proficiency, including the first author. The annotators had never watched the movies before the annotations, except the first author. The example annotations from previous work (Nishida and Nishimoto, 2018) were used to instruct the annotators. Specifically, the annotators were instructed to make detailed descriptions every 2 s of the movie, including what is happening at the moment, by whom, where, when, how, and why. *Cops* and *The Kid* were annotated by four annotators and *Mr. Bean* was annotated by three annotators.

We used latent semantic analysis (LSA) to represent annotated sentences from the co-occurrence statistics of the words in the sentence, given the total words that appeared in the movie annotation. A pair of sentences with similar word distributions resulted in similar embedding vectors. The method uses singular value decomposition (100-dimensional embedding vector; sklearn.decomposition.TruncatedSVD) (Landauer and Dumais, 1997) on the one-hot word count matrix (number of sentences in annotation × total number of unique words in the annotation), to characterize the semantics of the sentences in the movie-specific embedding space. To further examine the validity of our semantic quantifications and test the robustness of our findings, we quantified semantic embedding vectors with an alternative method, a universal sentence encoder (USE) (Cer et al., 2018). USE is a pretrained deep averaging network encoder in which each annotated sentence is used as an input to the pretrained model that is publicly available in Google's Tensorflow-hub (512-dimensional embedding vector). USE is different from LSA in that it quantifies annotated sentences in a fixed, pretrained embedding space.

The semantic relationship between the pairwise moments (2 s) of the movies was calculated by the cosine similarities between the sentence embedding vectors. The semantic relationship score of each moment was computed by averaging the past moments' (of the different scenes) semantic relationship to the present moment. Similarly, to relate the semantic relationship of each moment to the group measure of narrative comprehension, we convolved HRF and applied the same sliding-window analysis.

Control analysis: stimulus saliency. The visual salience was measured for all video frames at a sampling rate of 1 s. The pixelwise intensity of the frame was measured using SaliencyToolbox (Walther and Koch, 2006), and the intensities of every location were averaged to represent framewise salience. The saliency measures of the frames that corresponded to the high- and low-comprehension moments were compared using a paired Wilcoxon signed-rank test.

fMRI experiment. Participants were scanned using a 3T scanner (Magnetom Prisma; Siemens Healthineers) with a 64-channel head coil. A session consisted of one anatomic run and one task-based functional run. The anatomic images were acquired using a T1-weighted MPRAGE pulse sequence (TR = 2200 ms, TE = 2.44 ms, FOV = 256 mm × 256 mm, and 1 mm isotropic voxels). Functional images were acquired using a T2*-weighted EPI sequence (TR = 1000 ms, TE = 30 ms, multiband factor = 3, FOV = 240 mm × 240 mm, and 3 mm isotropic voxels, with 48 slices covering the whole brain). A single EPI run lasted for 31 min 20 s, which included 30 s of blank fixation periods in between the three movie watching conditions (Initial Scrambled, Original, and Repeated Scrambled), and 10 s of additional fixations at the start and end of the run (Fig. 1*c*). Only one movie stimulus was tested in a single session of fMRI.

Participants first watched the same scrambled movie as in the behavioral studies (Initial Scrambled condition), and watched the movie in an original sequence (Original condition), then watched the same scrambled movie again presented in the same order (Repeated Scrambled condition). We compared the Initial Scrambled condition to the Repeated Scrambled condition in which participants were viewing the same stimulus but were assumed to be engaged in a different cognitive state. During the Initial Scrambled condition, we assumed that the participants would be actively engaged in comprehending the scrambled narrative, such that the cognitive states can be inferred from the group measure of comprehension that was estimated from a behavioral study where independent participants watched the same scrambled movies for the first time. In contrast, no comparable fluctuation in comprehension was expected to occur during the Repeated Scrambled condition as this was after participants had watched the same movie in an original sequence (Original condition). Therefore, we hypothesized that, if the different neural activity is observed depending on the changes in comprehension in the Initial but not in the Repeated Scrambled

condition, they were likely to be attributed to different cognitive states driven by narrative comprehension, not by other stimulus-driven effects. No explicit task was given to the participants to exclude possible task-induced effects. Participants were instructed to attend to the movie at all times and attempt to comprehend and infer the original temporal and causal structures of the scrambled movie.

The visual stimulus was projected from a Propixx projector (VPixx Technologies), with a resolution of 1920 × 1080 pixels and a refresh rate of 60 Hz. The movies were projected onto the center of the screen, with a 22.6° × 15.1° FOV. The background music was delivered by MRI-compatible in-ear headphones (MR Confon; Cambridge Research Systems).

Image preprocessing. Structural images were bias field-corrected and spatially normalized to the MNI space. The first 10 images of the functional data were discarded to allow the MR signal to achieve T1 equilibration. Functional images were motion-corrected using the six rigid-body transformation parameters. After motion correction, there was no difference in the framewise displacement (FD) (Power et al., 2012, 2014) between the binary moments of high and low comprehension, in both the Initial (high: FD = 0.038 ± 0.008, low: FD = 0.039 ± 0.007; paired Wilcoxon signed-rank tests, $z_{(66)} = 0.79$, $p = 0.431$, Cohen's $d = 0.06$) and Repeated (high: FD = 0.041 ± 0.006, low: FD = 0.041 ± 0.006; $z_{(66)} = 0.67$, $p = 0.504$, Cohen's $d = 0.06$) Scrambled conditions. The functional images were slice timing-corrected, intensity-normalized, and registered to MNI-aligned T1-weighted images. We applied the fMRIB's independent component analysis (ICA)-based X-noiseifier (FIX) to automatically identify and remove noise components (Griffanti et al., 2014, 2017; Salimi-Khorshidi et al., 2014). The BOLD time series were linearly detrended and band pass filtered ($0.009\,\text{Hz} < f < 0.125\,\text{Hz}$) to remove low-frequency confounds and high-frequency physiological noise. The data were spatially smoothed with a Gaussian kernel of FWHM of 5 mm. All analyses were conducted in the volumetric space, and the cortical surface of the MNI standard template was reconstructed using Freesurfer (Fischl, 2012) for visualization purposes.

GLM analysis. To ask whether there exists a systematic modulation of BOLD activity in brain regions, we applied a GLM regression, using AFNI. Preprocessed functional brain images ($N = 67$) were used as dependent variables, which included scans with three movie stimuli in the Initial Scrambled, Original, and Repeated Scrambled conditions. The group-aggregate behavioral measures of comprehension were used as regressors in the model, applied to the moments during the Initial Scrambled and Repeated Scrambled conditions. We also replicated the analyses with the binary indices of high and low comprehension. The block timing of the Initial and Repeated Scrambled conditions, and a linear drift were included as nuisance regressors in the model. In a group-level analysis, we selected clusters of voxels (cluster size = 40) that were significantly (false discovery rate [FDR]-corrected, $q < 0.01$) correlated with the changes in comprehension during the Initial and Repeated Scrambled conditions respectively.

Whole-brain parcellation. For FC analysis, we parcellated the cortical and subcortical regions of the brain to extract BOLD time series from the parcellated brain regions. Cortical regions were parcellated into 114 ROIs (Yeo et al., 2015) based on a seven-network cortical parcellation estimated from the resting-state functional data of 1000 adults (Yeo et al., 2011). The seven canonical functional networks included visual (VIS), somatosensory-motor (SM), DAN, ventral attention network (VAN), limbic network, DMN, and FPN. Subcortical regions were parcellated into eight ROIs, corresponding to the bilateral amygdala, hippocampus, thalamus, and striatum, extracted from the Freesurfer segmentation of the FSL MNI152 template brain (Yeo et al., 2015). The subcortical ROIs were combined as a single, subcortical network in the functional network analysis. The time series of the voxels within each ROI were averaged, resulting in a time series matrix of functional scan duration (1870 s) × region (122 ROIs). To replicate the results using different atlases, we used the Brainnetome atlas that parcellates the whole brain into 246 ROIs (Fan et al., 2016). For a comparison between the two parcellation schemes, we calculated the topological overlap between each Brainnetome atlas ROI and the eight predefined functional networks (Yeo et al., 2011), which was regarded as the probability of a specific Brainnetome atlas ROI being identified as part of each of the eight functional networks. The network label with the highest probability was assigned to each Brainnetome atlas ROI.

Time-resolved FC. We hypothesized that the whole-brain FC patterns would be systematically modulated depending on changes in cognitive states related to narrative comprehension. To this end, we extracted time-resolved FC to estimate dynamically changing interactions of the pairwise functional brain regions of the 122 ROIs (Sakoğlu et al., 2010; Handwerker et al., 2012; Hutchison et al., 2013; Allen et al., 2014; Leonardi and Van De Ville, 2015). We applied a tapered sliding-window analysis, matching the hyperparameter selection of the behavioral data analysis. The chosen window size fell within the range of optimal window size suggested by previous research (Shirer et al., 2012; Deng et al., 2016; Liégeois et al., 2016). A tapered window was convolved with a Gaussian kernel of $\sigma = 3$ s to give higher weights to the center of the window (Allen et al., 2014; Barttfeld et al., 2015; Preti et al., 2017). An L1 penalty was added to increase the sparsity of the resulting correlation matrices, using the Graphical Lasso (Friedman et al., 2008). The regularization parameter was fixed to $\lambda = 0.01$ for the ROIs selected from the Yeo et al. (2011) atlas and to $\lambda = 0.1$ for the Brainnetome atlas ROI. The regularized correlation matrices were Fisher's *r*- to *z*-transformed.

Graph theoretical network analysis. Using sparse, weighted, and undirected FC matrices, we conducted graph theoretical network analyses, using the Brain Connectivity Toolbox (https://sites.google.com/site/bctnet/) (Rubinov and Sporns, 2010). As a global network measure, we calculated modularity by iteratively maximizing the modular structures using the Louvain algorithm (Newman, 2004, 2006; Blondel et al., 2008; Fortunato, 2010) with a resolution parameter $\gamma = 1$. Both the positive and negative edges were included, but a reduced weight was given to the negative edges (Rubinov and Sporns, 2011; Shine et al., 2016) as follows:

$$Q_T = \frac{1}{v^+}\Sigma_{ij}(w_{ij}^+ - e_{ij}^+)\delta_{M_iM_j} - \frac{1}{v^+ + v^-}\Sigma_{ij}(w_{ij}^- - e_{ij}^-)\delta_{M_iM_j} \quad (1)$$

Equation 1 indicates a time-resolved Louvain modularity algorithm (Q_T). w_{ij}^+ indicates the weights of the positive functional connections between regions i and j within the range (0, 1), and w_{ij}^- indicates the weights of the negative functional connections between regions i and j within the range (0, 1). $v^\pm$ indicates the sum of all positive or negative connection weights within the graph, where $v^\pm$ equals $\Sigma_{ij}w_{ij}^\pm$. $\delta_{M_iM_j}$ indicates the module partitions between regions i and j, where $\delta_{M_iM_j} = 1$ identifies that i and j lie within the same module, and $\delta_{M_iM_j} = 0$ identifies that i and j lie in different modules. $e_{ij}^\pm$ indicates the strength of a connection divided by the total weight of the graph, where $e_{ij}^\pm = \frac{\Sigma_j w_{ij}^\pm \Sigma_i w_{ij}^\pm}{v^\pm}$.

Further, we quantified global efficiency, after thresholding the matrices by leaving only the positive edges (Rubinov and Sporns, 2010). The global efficiency was measured as the average inverse shortest path length between all pairs of regions in the network (Latora and Marchiori, 2001) as follows:

$$E_{gT} = \frac{\Sigma_{i\neq j\in G}(d_{ij}^w)^{-1}}{\mathrm{N}(\mathrm{N}-1)} = \frac{1}{\mathrm{N}(\mathrm{N}-1)}\sum_{i\neq j\in G}\frac{1}{d_{ij}^w} \quad (2)$$

Equation 2 is the measure of global efficiency (E_{gT}), where d_{ij}^w indicates the shortest path length between the regions i and j, and N indicates the total number of regions in the graph.

As regional graph theoretical measures of the across- and within-modular connections, we calculated the participation coefficient and within-module degree z score, based on the time-resolved community structure derived from the Louvain modularity algorithm (Guimerà and Nunes Amaral, 2005; Shine et al., 2016) as follows:

$$PC_{iT} = 1 - \sum_{s=1}^{N_{MT}} \left(\frac{\kappa_{isT}}{\kappa_{iT}}\right)^2 \quad (3)$$

$$WMDZ_{iT} = \frac{\kappa_{iT} - \kappa'_{SiT}}{\sigma_{\kappa_{SiT}}} \quad (4)$$

Equations 3 and 4 show the time-resolved measure of participation coefficient (PC_{iT}) and within-module degree z score ($WMDZ_{iT}$). In Equation 3, PC_{iT} ranges between 0 and 1. κ_{isT} indicates the strength of positive functional connections between region i and all other regions at module S_i at time T, and κ_{iT} indicates the strength of positive functional connections of region i to all other regions, regardless of module assignment. N_{MT} indicates the number of modules at time T, where the modules are defined using the Louvain algorithm. The time-resolved participation coefficient of a region approximates to 1 if the connections are made with the regions of other modules. In Equation 4, κ_{iT} indicates the strength of connections of region i to other regions that lie within the same module S_i at time T, and κ'_{SiT} indicates the average of κ over all regions in the module S_i at time T. $\sigma_{\kappa_{SiT}}$ indicates the SD of κ in module S_i at time T.

The time-resolved network measures were related to the binary indices of group-aggregate measures of comprehension. For each participant's ROI, the time-resolved network measures were averaged to produce a single summary value representing high- and low-comprehension moments, respectively. With every ROI being assigned to one of eight functional networks (Yeo et al., 2011), we averaged the summary values of ROIs that corresponded to each functional network. The results from all participants across the three movie stimuli were combined ($N = 67$), and paired Wilcoxon signed-rank tests were performed between the summary network measures that corresponded to high- and low-comprehension moments. The effect size was estimated using a Cohen's d. Statistical values from the regional network analysis were FDR-corrected for multiple comparisons across different functional networks (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001), and the interaction effect was tested with a repeated-measures ANOVA.

FC between pairwise subregions of the DMN. We compared the FC strengths between the pairwise regions of the DMN during high- and low-comprehension moments. The ROIs indicated as part of the predefined DMN (Yeo et al., 2011) were grouped into five regions: the medial prefrontal cortex (mPFC), middle frontal gyrus (MFG), middle temporal gyrus (MTG), angular gyrus (Ang), and precuneus together with the posterior cingulate cortex (PreCu/PCC), based on their anatomic separations (Simony et al., 2016). The BOLD time series of the voxels corresponding to each subregion of the DMN were averaged, and the time-varying functional connections between the pairwise DMN regions were computed by applying the same tapered sliding-window analysis to the Fisher's r- to z-transformed correlation matrices, without regularization. Likewise, the average FC strengths during high- and low-comprehension moments were computed per individual and compared at the group level using paired Wilcoxon signed-rank tests ($N = 67$), respectively, for the Initial and Repeated Scrambled condition, and a repeated-measures ANOVA was used to test for an interaction effect.

HMM latent state analysis. To characterize the dynamics of low-dimensional, latent neural states that represent functional brain activity during movie watching, an HMM was used, which probabilistically infers latent states of the time series. We defined ROIs based on group-level ICA (Beckmann et al., 2005), using FSL-MELODIC (http://www.fmrib.ox.ac.uk/fsl/melodic/index.html). The fMRI data of all participants in the three movie-watching conditions were concatenated. The independent components (ICs) were automatically extracted; then the authors qualitatively assessed for inclusion of noise ICs. If an IC (1) spatially overlapped with white matter or CSF, (2) was derived from motion artifacts, or (3) had a temporal frequency that lied outside of a signal range ($f > 0.125$ Hz), it was discarded as noise. The signal components were further qualitatively validated from their spatial overlaps with the well-known functional networks identified from the resting-state fMRI data (Smith et al., 2009). This resulted in a total of 30 signal ICs.

To characterize canonical brain states representative of movie watching, and to infer latent states from different moments of the fMRI scan that do not overlap with our analysis of interest (i.e., Initial and Repeated Scrambled conditions), the HMM was trained using the BOLD time series extracted from the 30 ICs of the concatenated time series of all participants' Original conditions of three different movie datasets (hmmlearn.hmm.GaussianHMM). We iteratively searched for the optimal number of states (K), within a range of 2-8. K was determined based on the model's consistency (Vidaurre et al., 2018) and clustering performance (Calinski and Harabasz, 1974; Gao et al., 2021). We first tested the model consistency across iterations, where the same HMM training and inference procedures were repeated 5 times using the same hyperparameters. The mean of the pairwise iterations' similarity in latent sequence (the proportion of same state occurrence over the entire time steps) was calculated to assess the consistency of the model. In addition, we tested the model's clustering performance to examine whether the inferred states distinguish the latent clusters of the observed IC time course. Specifically, the Calinski-Harabasz score, or the variance ratio criterion (Calinski and Harabasz, 1974), calculates the ratio between the within-cluster dispersion (cohesion of the IC activation maps that were inferred as the same underlying brain state) and the across-cluster dispersion (separation of the IC activation maps that were assigned to the different brain states). The Calinski-Harabasz score was measured per each participant's fMRI time course, which was averaged to represent the HMM clustering performance per K. After K was determined, among five iterations, we chose the one that had the highest log probability of the inferred state sequence, given the observed IC time series.

To overcome the problem of local minima during the initialization of the HMM inference, we initialized the HMM parameters using the output of k-means clustering with the same number of states (K) as in the HMM analysis (sklearn.cluster.KMeans). Expectation-maximization (Dempster et al., 1977) of the forward-backward algorithm was used to estimate the optimal model parameters: transition probability and emission probability. The log-likelihood of the observation was iteratively estimated, conditioned on the model. The number of iterations with different centroid seeds was set to 500. We decided that the forward-backward algorithm approached an asymptote when the gain in log-likelihood reached 0.001 during the re-estimation process. No restraint was given to the transition probability matrix so that the transitions could occur to all possible states. We modeled the emission probabilities using a mixture Gaussian density function, where the mean vector and covariance matrix were produced from a mixture of 30 ICs for each state. The mean activation vector was characterized as the weights given to the activation of the 30 ICs, and the covariance matrix was characterized as the functional covariance between the pairwise ICs (Vidaurre et al., 2017, 2018). We defined each inferred neural state as the weighted sum of the extracted ICs with the mean activation vectors. To label each neural state as a known functional network, we masked the whole brain with eight predefined functional networks and compared the levels of activation corresponding to each network. The latent state was labeled using a functional network that showed the highest level of activation. If two functional networks had comparable activation profiles, the state was labeled using both networks (e.g., SM+VIS). The covariance matrix of each state consisted of the pairwise temporal covariance of the 30 ICs during the emergence of a latent state within the fitted sequence.

To characterize the modular structure of the functional covariance patterns of the inferred latent brain states, we applied the Louvain modularity algorithm to the latent states' covariance matrices. The output modules were largely grouped as (1) the DMN+FPN, (2) VIS, and (3) SM, which were identified from the module's probabilistic spatial correspondence to the resting-state functional networks defined in previous work (Smith et al., 2009).

The estimated transition and emission probabilities were applied to decode the most probable sequence of the concatenated time series of all participants during the Initial and Repeated Scrambled conditions, using the Viterbi algorithm (Rezek and Roberts, 2005). The outcome of the Viterbi algorithm is the probability of each latent state being the most dominant state at a specific time point. We chose the state with the

highest probability to be a latent state at a specific moment, thus discretizing the latent sequence. Importantly, to ask whether a neural state was dominantly associated with a certain cognitive state, we related the estimated neural states to the binary indices of the group measure of comprehension. The fractional occupancy of each state was calculated during the high- and low-comprehension moments per participant. We conducted paired Wilcoxon signed-rank tests to compare fractional occupancies during high and low comprehension per state, and a repeated-measures ANOVA to test the interaction between Scrambled condition and comprehension states. p values were FDR-corrected for multiple comparisons across the number of states.

Across-participant neural synchrony in latent state dynamics. Furthermore, we asked whether the inferred neural state dynamics were synchronized across participants as they comprehended the same scrambled movies. Specifically, we asked whether the degree of synchrony differed between the Initial and Repeated Scrambled conditions. To estimate the pairwise-participant similarity of the inferred state sequence during the Initial and Repeated Scrambled conditions, we calculated the proportion of the paired time points identified with an identical latent state over the entire scan duration, per pairwise participants. The neural state similarity was compared between the Initial and Repeated Scrambled conditions using a paired Wilcoxon signed-rank test, and the effect size was estimated using a Cohen's d.

To replicate the findings of across-participant neural synchrony in the Initial and Repeated Scrambled conditions, we conducted an intersubject correlation (ISC) analysis (Hasson et al., 2004; Nastase et al., 2019) using 122 ROI parcellation (Yeo et al., 2011). An ISC was computed per ROI, by iteratively leaving out a single participant and estimating a Pearson's correlation between its regional time course and the average regional time course of the rest of the participants who watched the same movie stimulus. The ISC values of 122 ROIs were summarized by functional networks, by averaging the ISCs of the brain regions that corresponded to each of the predefined functional network. The ISCs of the Initial and Repeated Scrambled conditions (results from all three movie stimuli, $N = 67$) were compared for each functional network using paired Wilcoxon signed-rank tests, and FDR-corrected for multiple comparisons.

Dynamic predictive modeling. Dynamic predictive modeling was conducted to predict a moment-to-moment group measure of comprehension using patterns of FC and patterns of regional BOLD responses. We used a cross-validated, linear support vector regression (SVR) model to predict the degree of comprehension at each time step, from a multivariate pattern of functional brain activity or connectivity at the corresponding time step. The model was validated with a linear support vector machine (SVM), which predicted the binary indices of high and low comprehension.

The model was cross-validated across participants and movie stimuli so that the model would not learn participant- or narrative-specific regularities. Specifically, the model was trained on every time step of all but one participant's fMRI data who watched two of the three movie stimuli and was tested on each time step of the held-out participant's fMRI data from a held-out movie stimulus. The mapping between multivariate brain features and behavioral score at each time step was input to the model as an independent instance. For the FC pattern-based prediction, the multivariate features to the model were Fisher's r- to z-transformed correlation matrices calculated from pairwise regions of the 122 ROIs using tapered sliding windows (window size = 36 s, step size = 1 s, σ = 3 s, without regularization). For the activation pattern-based prediction, the time courses of BOLD responses from the 122 ROIs were used as features. The time course of each feature was z-normalized per participant, to maintain within-feature temporal variance while removing across-participant and across-feature variances. Feature selection was used to select functional connections between ROIs or responses from an individual ROI that were consistently correlated with the comprehension time courses. For the training sample in each cross-validation fold, the time course of each neural feature was correlated with the comprehension measure (Pearson's correlation). A feature was selected if the distribution of correlation coefficients of the training sample was significantly different from zero (one-sample t test, $p < 0.01$) (Shen et al., 2017), regardless of whether the average correlation was positive or negative.

The predicted behavioral time course was the group measure of comprehension, estimated from the behavioral experiment. Because we had a single group measure of comprehension per movie stimulus, the same behavioral measure was predicted by multiple individuals' fMRI data who watched the same movie. The behavioral measures were zero-padded for activation pattern-based prediction, whereas non–zero-padded measures were used for FC pattern-based prediction to match the duration of the time-resolved FC matrices (scan duration minus the sliding-window size).

The prediction performance of the SVR was calculated by the Pearson's correlation between the predicted and observed behavioral time course, averaged across 67 cross-validation folds (Fisher's r- to z-transformed). The prediction performance of the SVM was the proportion of the predicted binary category (high vs low comprehension) being correct. Again, performance was calculated per cross-validation fold and averaged to represent mean prediction performance. The significance was calculated using nonparametric permutation tests, where the same model predicted phase-randomized comprehension measures (iteration = 1000). The actual prediction performance was compared with the null distribution using a one-tailed test, p = (1 + number of null prediction performance $\geq$ actual performance)/(1 + number of permutations).

We asked which of the functional network pairs were selected above the chance to be correlated with the group measure of narrative comprehension. For the Initial and Repeated Scrambled condition, respectively, we extracted a set of functional connections that were consistently selected to be correlated with the comprehension measures in every cross-validation fold (one-sample t test, $p < 0.01$) (Shen et al., 2017). To characterize the selected functional connections in the predefined functional network space, we computed the proportion of the number of selected functional connections among the total possible number of ROI connections of the pairwise functional networks. Then, we generated the size-matched random networks (iteration = 10,000) and estimated the proportion of the selected connections in the pairwise functional network matrix in the same fashion. The significance was tested per functional network pair using a one-tailed, nonparametric permutation test, and FDR-corrected for the number of pairwise functional networks.

Results

Moments of comprehension during scrambled movie watching are synchronized across individuals

To maximally induce fluctuations in comprehension during narrative movie watching, we used three 10 min silent movies that were segmented into multiple scenes (36 ± 4 s per scene), then were scrambled in their temporal order. The scrambled order was the same for all participants per movie. To quantify changes in comprehension as participants attempted to understand the scrambled movies ($N = 20$ per movie), they were asked to press a button when they thought they had understood the narrative ("Aha"), or when their previous feeling of comprehension turned out to be incorrect ("Oops"). As the "Oops" responses incorporate the psychological notion of "Aha" (Danek and Wiley, 2017), no distinction was made between the two response types and were summed in the analysis. The moments of button presses were largely consistent across individuals who watched the same movies (mean Dice coefficients = 0.256, range of null distribution = [0.169, 0.213]; 0.243, [0.157, 0.201]; 0.249, [0.162, 0.207] for three movie stimuli, nonparametric permutation test with one-tailed test, all p values < 0.001; Fig. 2*a*). The results indicate that participants experienced the subjective feeling of comprehension at similar moments.

Because the cognitive states related to narrative comprehension were synchronous across individuals, we generated a group-aggregate measure of narrative comprehension to temporally

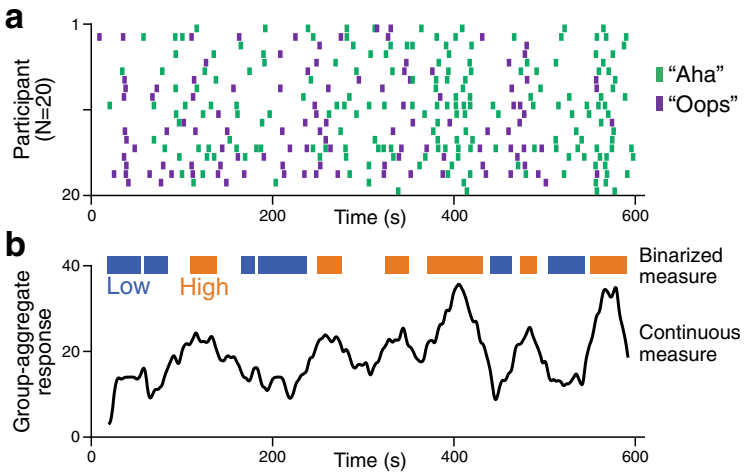

Figure 2. Changes in comprehension during temporally scrambled movie watching. ***a***, Behavioral responses of comprehending moments while watching an exemplar temporally scrambled movie ($N = 20$). Participants pressed "Aha" when they experienced subjective feelings of comprehension (green) and "Oops" when they realized their previous comprehension was incorrect (purple). ***b***, A continuous and binary group-aggregate behavioral measure of narrative comprehension. All participants' responses in ***a*** were HRF-convolved and aggregated by applying a sliding window. The top one-third of the moments with frequent responses were defined as the moments of high comprehension, whereas the bottom one-third were defined as the moments of low comprehension. Behavioral results using two other movie stimuli are shown in https://github.com/hyssong/comprehension.

relate to fMRI data collected from an independent pool of participants as they comprehended the same movie stimuli (Fig. 2*b*). The moments of button responses were convolved with a canonical HRF and were aggregated using a sliding window of 36 s, in steps of 1 s, in which the number of responses of all participants was summed within each time window. A window size of 36 s was chosen to match the average scene duration of the scrambled movies. We binarized the moments based on the aggregated response frequency into the moments of high or low comprehension. The top one-third of the moments were categorized as moments of high comprehension, and the bottom one-third were categorized as moments of low comprehension (Fig. 2*b*). We discarded the middle one-third because cognitive states during those moments were subject to higher variability across participants than the top and bottom thirds.

The group measure of comprehension in Figure 2*b* represents the degree to which a person is likely to experience a subjective feeling of comprehension while watching a scrambled movie. In the behavioral study, we assumed that the cognitive states involved in narrative comprehension would comprise a state of a sudden insight (the experience of "Aha"), change in comprehension (that is represented with, but not restricted to, "Oops" responses), or a gradual increase in comprehension. To test our assumption that participants' general comprehension would increase as the movie progresses, we conducted a linear trend analysis on the group measure of comprehension. We observed significant positive linear trends in all three movie stimuli (linear regression model fit, $t_{(572)} = 11.02$, $p < 0.001$, $r^2 = 0.175$; $t_{(572)} = 8.36$, $p < 0.001$, $r^2 = 0.109$; $t_{(572)} = 5.64$, $p < 0.001$, $r^2 = 0.053$), supporting our assumption that participants' comprehension gradually increased during movie watching. Collectively, the results indicate that the participants experience comprehension at similar moments of the narratives, and our group-level behavioral measures represented fluctuating cognitive states involved with comprehension.

Comprehension occurs at causally important moments of the narratives

Theories proposed that comprehension occurs by integrating relevant features of the narratives and creating a coherent situational model (Graesser et al., 1994; Wolfe et al., 2005). How does narrative integration occur, and what constitutes a representation of narratives? We hypothesize that comprehension occurs when the incoming information is integrated with the causally related past events, thereby constructing a causally coherent representation of narratives. To test this account, we conducted an additional behavioral experiment using the same scrambled movies ($N = 12$ per movie) that estimates the degree of causal importance of every moment of the narratives. In the experiment, participants first segmented the events by marking the perceived event boundaries of the scrambled movie, then rated the causal relationship between every possible pairwise event on a scale from 0 to 2: 0 (no causal relationship), 1 (shares a causal relationship), and 2 (shares a causal relationship that is critical in developing the narrative). The responses of all participants were summed to create moment-to-moment, causal relationship matrices that indicate the degrees of a causal relationship between the narrative events of every pairwise moment in time (Fig. 3*a*). The causal relationship matrix, unscrambled into the original order (Fig. 3*b*), indicates that not only the causal relationship between temporally consecutive events but long-range causal chains between temporally discontiguous events also exist. We calculated the causal relationship that each moment has with the past moments by averaging the causal ratings of all preceding time points that did not belong to the same scene (a "scene" indicates a 36 $\pm$ 4 s block of the movie segmented for temporal scrambling), creating a causal relationship time course per scrambled movie stimulus (Fig. 3*c*). We predicted that the moments of high comprehension would correspond to moments that are strongly causally related to past events. The group measure of comprehension was correlated with the causal relationship time courses of all three movies (Pearson's $r = 0.722$, $r = 0.440$, $r = 0.320$, compared with the null distribution in which causal relationship was correlated with 1000 phase-randomized comprehension measures, nonparametric one-tailed $p < 0.001$, $p = 0.024$, and $p = 0.065$, respectively). Additionally, when comparing the causal relationship measured at the binary moments of high and low comprehension, we observed a significantly higher causal relationship with past events during the high- compared with low-comprehension moments (paired Wilcoxon signed-rank test; $z_{(187)} = 11.40$, Cohen's $d = 2.24$; $z_{(168)} = 7.44$, Cohen's $d = 0.96$; $z_{(163)} = 4.76$, Cohen's $d = 0.55$; all p values < 0.001 for the three movies). The results suggest that individuals experience comprehension when perceiving events that are strongly causally linked with the previous events; that is, when an incoming event takes on an important role within the narrative's causal structure (Graesser et al., 1994; H. Lee and Chen, 2021). This supports the hypothesis that comprehension entails the integration of incoming events with the memory of the causally related past events, to formulate a causally coherent representation of the narrative.

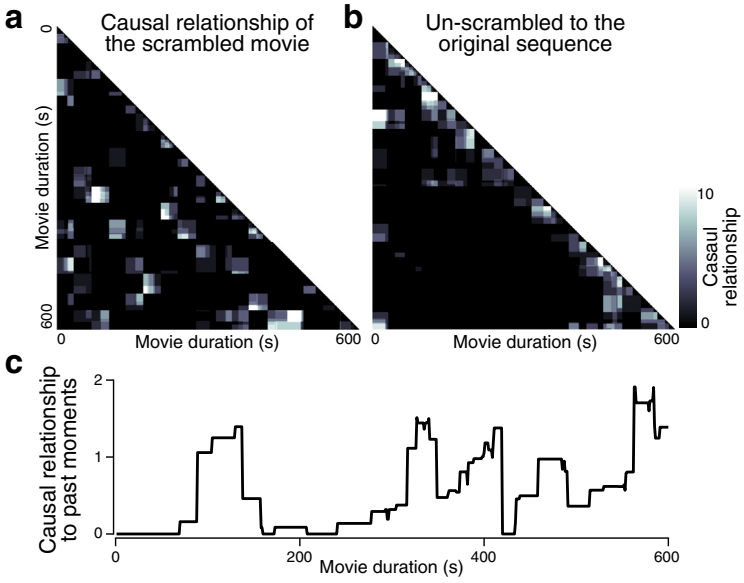

Figure 3. The causal relationship between narrative events. ***a***, Causal relationship matrix, indicating the degree of a causal relationship between all pairwise moments of an exemplar scrambled movie ($N = 12$). Participants rated the causal relationship of the pairwise perceived event segments of the scrambled movie on a scale from 0 to 2. All participants' responses were summed to generate a single causal relationship matrix. ***b***, Causal relationship matrix, unscrambled according to the original movie sequence. Strong clustering around the diagonal indicates that the temporally contiguous moments in the original sequence tend to be causally linked. Pairwise events that are temporally distant but highly causally related also existed, indicating the presence of key pairs of events that are critical in developing narratives. ***c***, The time course of causal relationship to past events, which represents causal importance. For each time point in ***a***, we averaged its causal relationship with every past moment of the different scenes.

However, there is a possibility that the semantic relationship between events, rather than the causal relationship, may have derived positive correlations with the comprehension measures. To test this alternative account, we measured the semantic relationship between pairwise moments of the movies using the pairwise similarities of the sentence embedding vectors of movie annotations. We used the written annotations generated by four native-level English speakers, which gave detailed descriptions of every moment (2 s) in the movies, including what was happening at that moment, by whom, where, when, how, and why. We used LSA to quantify semantics of narrative annotations (Landauer and Dumais, 1997), which characterizes movie-specific sentence embedding vectors based on the relative similarities of the word occurrence frequencies of the annotated sentence (for details, see Materials and Methods). A semantic relationship matrix was created for each movie by calculating the mean cosine similarities between the pairwise sentence embedding vectors that were generated by multiple annotators. We also validated our semantic quantification using USE. Semantic relationship matrices generated using LSA and USE were highly comparable for all three movies (Pearson's $r = 0.762$, $r = 0.794$, $r = 0.691$ per movie, all p values $<$ 0.0001), suggesting that the quantified semantic relationship was robust to the choice of analysis method.

A causal relationship matrix was positively correlated with a semantic relationship matrix above the chance (Pearson's $r = 0.156$, $r = 0.104$, $r = 0.105$, all p values < 0.001, compared with the null distribution where semantic relationship matrices were randomly shuffled, one-tailed test, 1000 iterations). The semantic relationship time course was computed similar to the causal relationship time course, such that the degree of semantic relationship of the past time points (except the time points that belong to the same scene) to the present moment was averaged. Changes in group-aggregate comprehension were positively correlated with the semantic relationship time courses for two among three movies (Pearson's $r = 0.468$, $p = 0.026$, $r = 0.105$, $p = 0.282$, $r = 0.498$, $p = 0.004$, compared with the null distribution in which the semantic relationship was correlated with 1000 phase-randomized comprehension time courses, one-tailed test). When comparing the semantic relationship between high- and low-comprehension moments, we observed that high-comprehension moments had a significantly higher semantic relationship with the past compared with the low-comprehension moments, again for the two among three movie stimuli ($z_{(187)} = 8.71$, $p < 0.001$, Cohen's $d = 1.16$; $z_{(168)} = 0.77$, $p = 0.443$, Cohen's $d = 0.16$; $z_{(163)} = 7.08$, $p < 0.001$, Cohen's $d = 0.96$).

However, critically, for all three movies, the causal relationship showed a significant correlation with the comprehension measures after the effect of the semantic relationship was controlled for (partial $r = 0.623$, $r = 0.436$, $r = 0.182$, all p values $<$ 0.001), whereas the semantic relationship could not consistently explain the comprehension measures when the effect of the causal relationship was controlled for (partial $r = -0.036$, $p = 0.391$, $r = 0.082$, $p = 0.050$, $r = 0.435$, $p < 0.001$). The results suggest that connecting the causal relationship between events plays a critical role in comprehending narratives, and this is not merely because of the semantic similarities between events. These findings were replicated using USE; we observed significant correlation between the causal relationship and comprehension measures while controlling for the semantic relationship (partial $r = 0.686$, $r = 0.435$, $r = 0.217$, all p values $<$ 0.001), whereas the result was not consistent across the movies when we correlated the semantic relationship and comprehension measures while controlling for the causal relationship (partial $r = 0.041$, $p = 0.322$, $r = 0.126$, $p = 0.002$, $r = 0.326$, $p < 0.001$). Overall, the results imply that narrative integration occurs based on the causal connections between events even after the semantic connections are accounted for and that the events' relative causal importance influences moments in time when narratives are integrated.

fMRI study during scrambled movie watching

A separate, independent group of participants underwent an fMRI experiment, where they watched the same set of scrambled movies inside a scanner ($N = 24$, 23, 20 for three movies). In addition to scrambled movie watching (Initial Scrambled condition), participants watched the same movie in an original order (Original condition), then watched the scrambled movie again in the same order of presentation (Repeated Scrambled condition). We compared the Initial Scrambled to the Repeated Scrambled condition in which participants viewed the same stimuli but in different cognitive states. During the Initial Scrambled condition, we assumed that participants were actively engaged in comprehending the scrambled movies, which led to dynamic fluctuations in comprehension. In contrast, no comparable

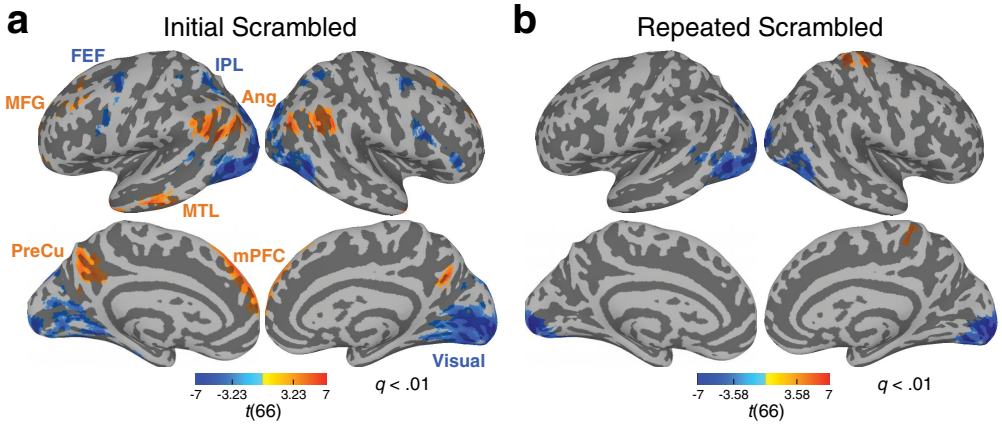

Figure 4. Modulation of BOLD activity during changes in narrative comprehension. Results of the GLM analysis using a continuous behavioral index of comprehension (results from all three movie stimuli, $N = 67$). The voxelwise group-level statistics (t test) were thresholded with cluster size > 40 and $q < 0.01$, with the thresholded t values indicated at the color bar. ***a***, Regions that show positive (orange) and negative (blue) correlations with comprehension time courses in the Initial Scrambled condition. When comprehension was high, responses in the DMN increased, whereas responses in the DAN and visual sensory network decreased. ***b***, Regions that show positive (orange) and negative (blue) correlations with comprehension time courses in the Repeated Scrambled condition. FEF, Frontal eye fields; IPL, inferior parietal lobule; MTL, middle temporal lobule; Visual, visual cortex.

comprehension was expected in the Repeated Scrambled condition as it was after participants had already watched the original narrative. Thus, if the neural activity differed depending on changes in comprehension in the Initial but not in the Repeated Scrambled condition, they were more likely to be driven by the cognitive state differences related to narrative comprehension than by the stimulus-related attributes. Furthermore, to exclude possible task-induced effects, no explicit task was given during fMRI, but participants were instructed to attend to the stimulus at all times and try to infer the temporal and causal structures of the original story.

Modulation of activity in the DMN and DAN during changes in narrative comprehension

We first examined whether any of the regions' BOLD activity was modulated by the changes in comprehension. We used the group-aggregate measures of comprehension (estimated in an independent behavioral study) as regressors in the GLM to fit voxelwise BOLD activity time series. Data of all participants in the three movie scans were included for the group-level statistical analysis to exclude stimulus-specific effects. Cognitive and neural states related to narrative comprehension were assumed to be common across different narratives and robust to the particular choice of stimuli.

In the Initial Scrambled condition, BOLD responses in the Ang, PreCu, mPFC, middle temporal lobule, and MFG, which together comprise the DMN, showed a higher level of activity when participants were likely to experience feelings of comprehension. In contrast, when comprehension was low, the frontal eye fields and inferior parietal lobule, regions of the DAN, and the visual sensory network, including the early and high-level visual areas, showed increased BOLD responses (Fig. 4*a*). These results suggest that the regions in the DMN are involved when integrating narratives to form an internal causal representation. In contrast, the regions in the DAN are involved when comprehension is low, which may suggest that the DAN is involved in attending to incoming events when trying to collect information that may later piece together as a coherent representation. Critically, in the Repeated Scrambled condition, these functional networks did not show systematic modulation of BOLD responses, except for the early visual areas, which was also found in the Initial Scrambled condition to exhibit greater activation during low-comprehension moments, and the right postcentral gyrus (Fig. 4*b*). The results indicate that the significant modulation of BOLD activity during the Initial Scrambled condition is not driven by the intrinsic properties of the stimuli, but derived from active cognitive state changes that occur during narrative comprehension. These results were replicated when the third-median-split binary indices of comprehension (high vs low) were used as regressors. Additionally, to examine whether greater BOLD activity in the visual areas during low comprehension (that appeared both in the Initial and Repeated Scrambled conditions) was a stimulus-related effect, we assessed the physical salience of the movie frames by calculating the pixelwise stimulus intensities at every frame (1 s) of the movie. For all three movies, salience was higher during low-comprehension moments compared with high (paired Wilcoxon signed-rank tests, $z_{(187)} = 2.33$, $p = 0.020$, Cohen's $d = 0.37$; $z_{(168)} = 4.23$, $p < 0.001$, Cohen's $d = 0.51$; $z_{(163)} = 7.40$, $p < 0.001$, Cohen's $d = 1.01$ for each movie), suggesting that activity differences observed in the visual areas may be because of the coincidentally stronger stimulus intensities during moments of low comprehension.

Reconfiguration of functional brain network into an integrated and efficient state during moments of narrative integration

We then examined whether the large-scale functional brain network reconfigures its interaction and information processing state as comprehension evolves. When comprehension is low, we expected a segregated brain state, where each functional network is engaged in its specialized function. However, when comprehension is high (i.e., when narratives are actively being integrated into a causally coherent structure), we anticipated a tightly integrated state that enables efficient communication across distinctive functional systems (Sadaghiani et al., 2015; Shine et al., 2016). For network analysis, we parcellated the brain into 122 ROIs and grouped them into eight predefined functional networks (Yeo et al., 2011, 2015). To account for the dynamic changes in FC in relation to cognitive state dynamics, we extracted the BOLD time series from each ROI and computed the time-resolved FC between pairwise regions during the Initial Scrambled, Original, and Repeated Scrambled conditions,

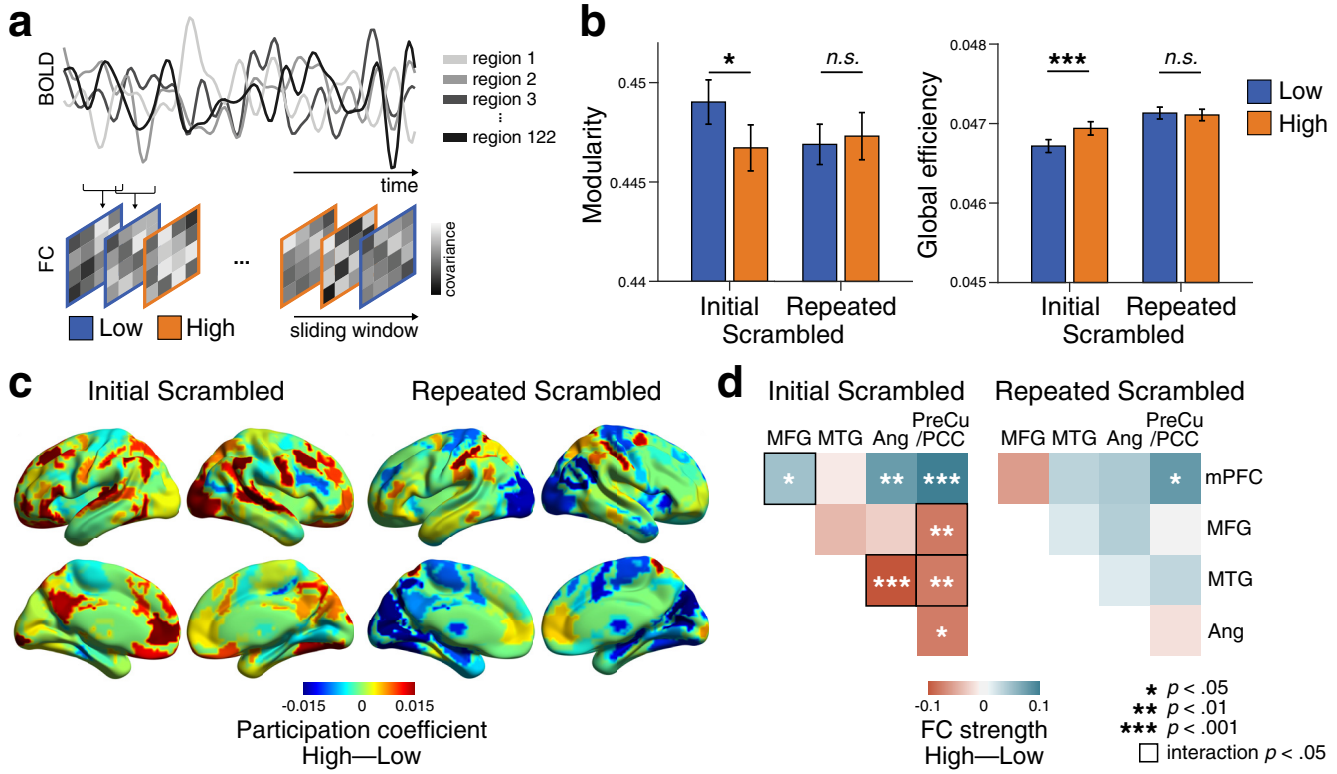

Figure 5. Dynamic reconfiguration of large-scale functional networks at moments of high and low comprehension. ***a***, Schematic overview of the time-resolved FC analysis using a sliding window. The BOLD time series was extracted from 122 ROIs (Yeo et al., 2011, 2015). Time-resolved FC matrices were constructed for each window across movie duration, and graph-theoretical network measures were computed from each FC matrix. Network measures were categorized by their correspondence to the cognitive states of either high or low comprehension, averaged within a participant, and were compared at group level. The analyses were conducted on all participants using three movie stimuli ($N = 67$). ***b***, Global network reconfiguration corresponding to cognitive state differences. Modularity and global efficiency representing high- and low-comprehension moments were compared, for both the Initial and Repeated Scrambled conditions. Error bars indicate ±1 SEM. ***c***, Differences in participation coefficients (a regional network measure of across-modular connections) between high- and low-comprehension moments, for the Initial (left) and Repeated (right) Scrambled conditions. The difference in participation coefficients was calculated per ROI and averaged across participants. Positive values (red) indicate that the regions exhibited higher participation coefficients during high-comprehension moments, whereas negative values (blue) indicate that the regions exhibited higher participation coefficients during low comprehension overall. The figure is visualized using BrainNet Viewer (Xia et al., 2013). ***d***, Difference in the FC strengths of the pairwise subregions of the DMN between high- and low-comprehension moments, for the Initial (left) and Repeated (right) Scrambled conditions. Colors represent FC strength differences, averaged across participants. The FC strength of high- compared with low-comprehension moments was compared per regional pairs. Square contour represents pairwise DMN subregions that showed significant interaction effects between Scrambled conditions and comprehension states. The significance was FDR-corrected for the number of regional pairs.

respectively (Fig. 5*a*). Graph theoretical measures were computed in a time-resolved manner to capture changes in large-scale functional network structures (Rubinov and Sporns, 2010). To examine the degree of functional segregation, we measured modularity, which captures the degree to which functionally specialized regions of the brain are clustered in a modular structure (Bassett et al., 2013). As an indicator for functional integration, we measured global efficiency that represents integrative information processing across remote regions of the brain (Achard and Bullmore, 2007; Bullmore and Sporns, 2009, 2012). With the binary group measure of high and low comprehension, we were able to extract participant-specific summary measures of modularity and global efficiency during the two distinct cognitive states. The graph-theoretical measures were compared between high- and low-comprehension moments, respectively for the Initial and Repeated Scrambled conditions.

In the Initial Scrambled condition, modularity decreased when comprehension was high (paired Wilcoxon signed-rank test, $z_{(66)} = 2.32$, $p = 0.020$, Cohen's $d = 0.25$), suggesting that tight interaction across functional modules arises when information is being integrated into coherent narratives (Fig. 5*b*). In contrast, there was no difference in modularity between high- and low-comprehension moments during the Repeated Scrambled condition ($z_{(66)} = 0.58$, $p = 0.561$, Cohen's $d = 0.06$). A significant interaction was found between the Scrambled conditions (Initial and Repeated) and comprehension states (high and low; $F_{(1,66)} = 6.98$, $p = 0.010$), although no main effect was observed (main effect of the Scrambled conditions, $p = 0.334$, comprehension states, $p = 0.146$). Similarly, global efficiency was higher during moments of high comprehension compared with low ($z_{(66)} = 3.67$, $p < 0.001$, Cohen's $d = 0.33$). There was no difference in global efficiency during the Repeated Scrambled condition ($z_{(66)} = 0.57$, $p = 0.566$, Cohen's $d = 0.04$), and the interaction was significant ($F_{(1,66)} = 10.62$, $p = 0.002$). Notably, a significant main effect of the Scrambled conditions was found ($F_{(1,66)} = 26.27$, $p < 0.001$), with a higher efficiency when the same scrambled movie was watched repeatedly. These results suggest that the efficiency of information processing increases when a coherent situational model is already represented in the brain. The results were reproduced when different sliding-window sizes or a different cortical parcellation scheme was used (for replication results, see https://github.com/hyssong/comprehension). Overall, these results indicate that the brain enters a functionally integrated and efficient state that ensures more efficient information transfer

across functional modules when integrating narratives to a coherent representation.

Changes in across- and within-network functional connections during narrative comprehension

Along with the global reconfiguration, the time-resolved regional network measures between the high- and low-comprehension moments were compared. For all ROIs, the participation coefficients and within-modular degree z scores were measured, which represent the degree of the across-modular and within-modular connections (Guimerà and Nunes Amaral, 2005; Shine et al., 2016). A higher participation coefficient indicates that a region is functionally connected to the regions of other functional networks in a distributed manner, whereas a higher within-modular degree z score indicates that a region is mainly associated with the regions that lie within the same module (Guimerà and Nunes Amaral, 2005). The network measures were computed per ROI and were averaged within a respective predefined functional network (Yeo et al., 2011). During the Initial Scrambled condition, all functional networks showed higher participation coefficients when comprehension was high compared with low. In particular, the FPN ($z_{(66)} = 4.24$, FDR $p < 0.001$, Cohen's $d = 0.43$, corrected for multiple comparisons across functional networks) and the DMN ($z_{(66)} = 2.87$, FDR $p = 0.017$, Cohen's $d = 0.31$) showed significantly higher participation coefficients when comprehension was high compared with low (Fig. 5*c*). There was a significant interaction between the Scrambled conditions and comprehension states for both networks (FPN: $F_{(1,66)} = 11.93$, DMN: $F_{(1,66)} = 12.29$; both p values < 0.001; for the full results of all functional networks, see https://github.com/hyssong/comprehension). During the Repeated Scrambled condition, no functional network showed similar patterns of modulation in their across-modular FC, except for a reversed pattern of higher participation coefficients during low comprehension in the visual and subcortical networks ($z_{(66)} = 3.35$, FDR $p = 0.005$, Cohen's $d = 0.27$, and $z_{(66)} = 3.21$, FDR $p = 0.005$, Cohen's $d = 0.29$, respectively). The within-modular connections, quantified by within-modular degree z scores, did not differ across the high- and low-comprehension moments for any of the functional networks, during both the Initial and Repeated Scrambled conditions (all FDR p values > 0.6). These results suggest that the global reconfiguration is largely driven by the increased across-modular functional connections of the FPN and DMN, but less so by the connections within functional modules.

The null results with the within-modular degree z scores, particularly for the regions of the DMN, were not consistent with prior work that suggested that dynamic changes in connections between DMN subregions reflect higher-order information processing, such as narrative integration (Simony et al., 2016; Ritchey and Cooper, 2020). With the recent studies reporting disparate functional roles of the subregions of the DMN (Andrews-Hanna et al., 2014; Braga and Buckner, 2017; Gordon et al., 2020; Ritchey and Cooper, 2020), we hypothesized that the within-network connections of the DMN may have been modulated differently for each subregional connection, which is not evident when computing within-modular degree z scores of the entire network. To examine this question, we computed time-resolved FC matrices between the pairwise BOLD time series of the five canonical subregions of the DMN (i.e., mPFC, MFG, MTG, Ang, and PreCu/PCC) and compared the mean FC strength representing the high- and low-comprehension moments (Fig. 5*d*). During the Initial Scrambled condition, we observed increased FC between Ang and MTG ($z_{(66)} = 4.09$, $p < 0.001$, Cohen's $d = 0.39$; interaction, $F_{(1,66)} = 11.81$, FDR $p = 0.007$), PreCu/PCC and MFG ($z_{(66)} = 3.30$, FDR $p = 0.003$, Cohen's $d = 0.45$; $F_{(1,66)} = 6.57$, FDR $p = 0.033$), and PreCu/PCC and MTG ($z_{(66)} = 3.16$, FDR $p = 0.004$, Cohen's $d = 0.44$; $F_{(1,66)} = 10.75$, FDR $p = 0.007$). In contrast, decreased FC was observed between MFG and mPFC ($z_{(66)} = 2.12$, FDR $p = 0.048$, Cohen's $d = 0.24$; $F_{(1,66)} = 10.22$, FDR $p = 0.007$), but not during the Repeated Scrambled condition. Other pairwise DMN subregions exhibited no significant interaction effect between Scrambled conditions and comprehension states (all FDR p values > 0.05). The coexistence of both significant increase and decrease in FC within the DMN supports our hypothesis that the subregional connections are modulated differently during narrative comprehension. The results suggest that the subregional connections of the DMN may take on different functional roles during narrative comprehension, which remains to be further studied in future research.

Brain state characterized by the integrated DMN and sensory processing network occurs at moments of narrative integration

We investigated whether large-scale neural state dynamics track changes in cognitive states involved with narrative comprehension. To infer the dynamics of low-dimensional latent states in an unsupervised data-driven manner, we applied the HMM, which assumes that the observed sequence of brain activity is probabilistically conditioned on the sequence of discrete latent states (Baker et al., 2014; Vidaurre et al., 2017, 2018; Quinn et al., 2018). To characterize the observed sequence of neural activity, we first conducted a group-level ICA (Beckmann et al., 2005) from all participants' concatenated fMRI responses of all three conditions across three movies. Thirty ICs were identified to be signal components that were involved during movie watching. The discrete latent neural states were derived from patterns of activation and functional covariance of the 30 ICs, as we trained the HMM on the data from the Original condition. When we set the number of latent states to 4, based on the model's consistency and clustering performance measures (for details, see Materials and Methods; Fig. 6*a*), the extracted states were: SM+VIS, DAN, Integrated DMN+VIS, and Segregated DMN+VIS (Fig. 6*c*). Each state was labeled as one or the combination of eight functional networks by its inferred regional activation patterns. Notably, the two DMN+VIS states (i.e., Integrated and Segregated DMN+VIS) showed similar activation patterns, yet their functional covariance significantly differed such that one had higher across-modular FC (one-sample Wilcoxon signed-rank test on the differences in functional covariance of all edges corresponding to the module pairs, FDR-corrected for module pairs; across DMN+FPN and VIS modules: $z_{(109)} = 4.78$, FDR $p < 0.001$; across DMN+FPN and SM modules: $z_{(89)} = 7.03$, FDR $p < 0.001$; however, across VIS and SM modules showed opposite pattern: $z_{(98)} = 2.14$, FDR $p = 0.033$) and lower within-modular FC than the other (within DMN+FPN module: $z_{(99)} = 2.81$, FDR $p = 0.006$; within VIS module: $z_{(120)} = 7.03$, FDR $p < 0.001$; within SM module: $z_{(80)} = 7.73$, FDR $p < 0.001$). The one with higher across-modular but lower within-modular FC was termed the "Integrated" DMN+VIS state, and the other the "Segregated" DMN+VIS state. We applied the derived states from the Original movie watching condition to infer the latent state dynamics in the Initial and Repeated Scrambled conditions. We verified that the inferred latent states were dynamic in nature. The maximal fractional occupancy, the highest proportion

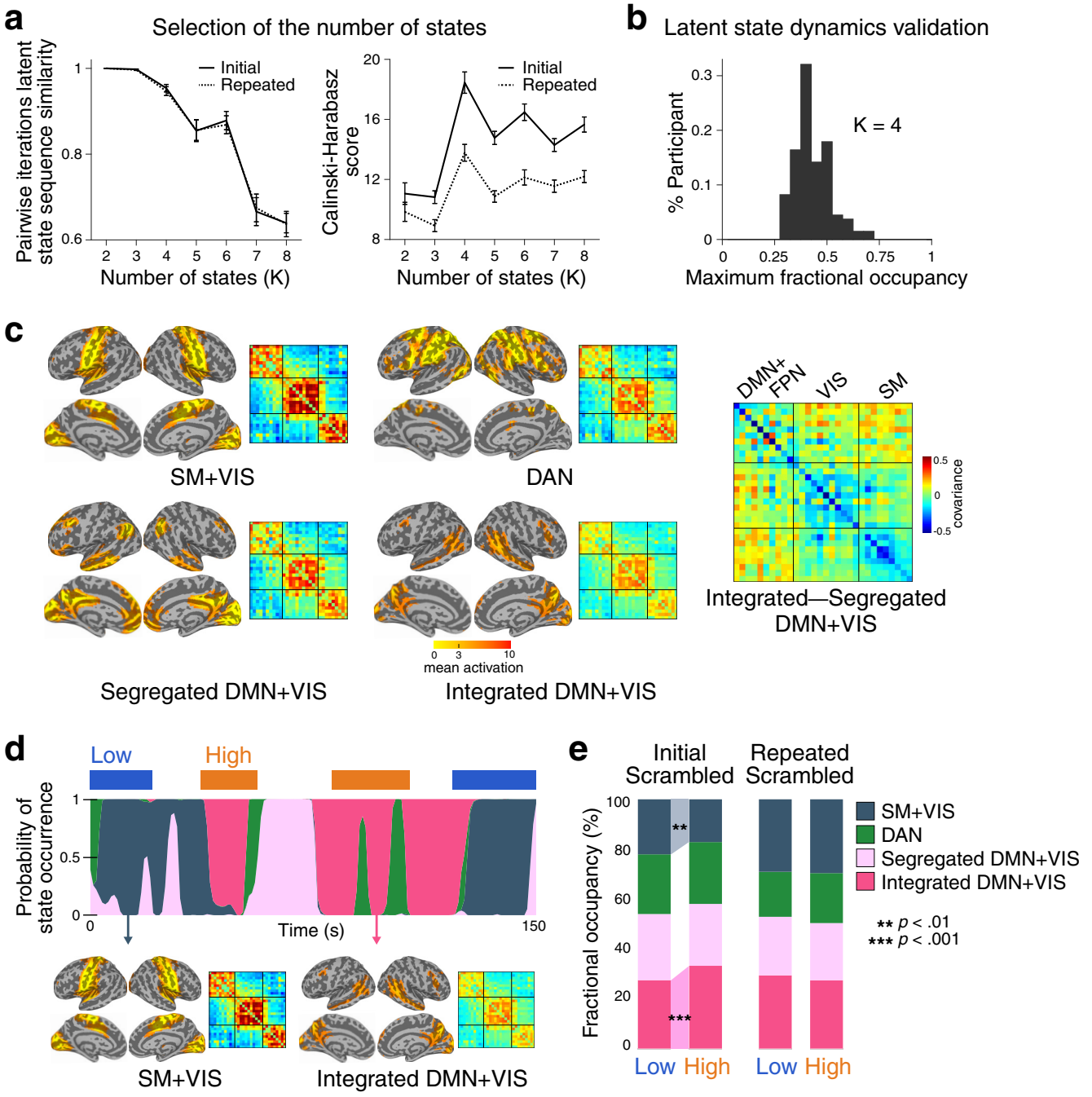

Figure 6. Brain state dynamics underlying narrative comprehension, derived from HMM. ***a***, Selection of the number of latent states (*K*) based on the model's consistency and clustering performance. Left, Model consistency is represented by the mean pairwise similarities of the predicted latent state sequence over five repeated iterations. Right, The model's clustering performance is represented by the ratio between the within-cluster and between-cluster dispersions of the observed IC time series ($N = 67$), with the clusters predicted by the HMM. A *K* value of 4 was chosen as the optimal number of states, given their high consistency and clustering performance among possible *K* values. ***b***, Validation of the dynamics in the inferred sequence from the HMM. The fractional occupancy of the highest emerging state was calculated as per the participant's inferred sequence. ***c***, Activation patterns and functional covariance of the four latent states, identified by training the HMM with the Original condition. The SM+VIS, DAN, Segregated DMN+VIS, and Integrated DMN+VIS were labeled based on their spatial activation patterns corresponding to the eight predefined functional networks. Right, The covariance matrix shows the difference between the covariance matrices of the Integrated DMN+VIS and Segregated DMN+VIS. ***d***, State occupancy and transition dynamics of a representative participant during exemplar moments of the Initial Scrambled condition. Occurrences of the states were probabilistically inferred at each time point (Vidaurre et al., 2017, 2018). Discrete latent states, assigned from the state with the highest probability of occurrence at respective time points, were related to the binary group measure of comprehension. ***e***, The average fractional occupancy of the four latent states during the moments of high and low comprehension, in the Initial and Repeated Scrambled conditions (results from all three movie stimuli, $N = 67$). Highlighted background between the colored bars represents significant differences in fractional occupancies, FDR-corrected for the number of states.

of a particular state's occurrence across all time points per participant, was <50% for most of the participants (one-sample Wilcoxon signed-rank test, $p < 0.001$; Fig. 6*b*), indicating that the transitions occurred from one latent state to more than one other state (Vidaurre et al., 2018).

Next, we examined whether the fractional occupancy of each neural state was modulated as the cognitive states traversed between different states of comprehension. Figure 6*d* illustrates the dynamics of the state occurrence probabilities of an exemplar participant, which is mapped in time to the binary group

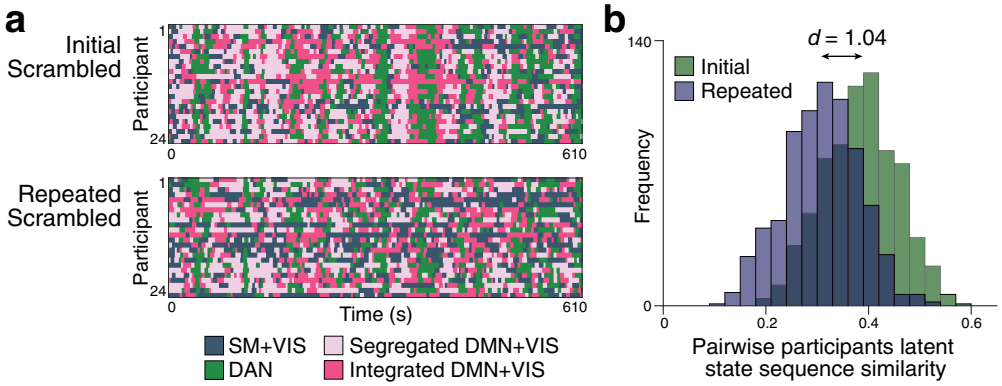

Figure 7. Synchrony of the latent neural states across individuals. ***a***, The HMM-derived neural state dynamics of exemplar movie watching participants in the Initial (top) and Repeated (bottom) Scrambled conditions. ***b***, Histograms of the similarities in state dynamics for all pairwise participants in the two conditions (results from all three movie stimuli, $N = 67$). The neural state dynamics were more similar across individuals in the Initial compared with the Repeated Scrambled condition.

measure of comprehension. In the Initial Scrambled condition, the SM+VIS had a higher occupancy during low-comprehension moments (paired Wilcoxon signed-rank test, $z_{(66)} = 3.32$, FDR $p = 0.002$, Cohen's $d = 0.39$), whereas the Integrated DMN+VIS had a higher occupancy during high-comprehension moments ($z_{(66)} = 3.81$, FDR $p < 0.001$, Cohen's $d = 0.44$) (Fig. 6*e*). None of the brain states differed in fractional occupancy across the high- and low-comprehension moments in the Repeated Scrambled condition (all FDR p values > 0.4). Significant interaction effects between Scrambled conditions and comprehension states were observed for both the fractional occupancies of the SM+VIS state ($F_{(1,66)} = 6.45$, $p = 0.013$) and Integrated DMN+VIS state ($F_{(1,66)} = 16.83$, $p < 0.001$). Results were consistent when six latent states instead of four were used in the HMM. In the Initial Scrambled condition, the Integrated DMN+VIS had a higher fractional occupancy during high comprehension ($z_{(66)} = 3.41$, FDR $p = 0.004$, Cohen's $d = 0.51$), and the visual sensory network state had a higher fractional occupancy during low comprehension ($z_{(66)} = 3.06$, FDR $p = 0.007$, Cohen's $d = 0.30$), whereas in the Repeated Scrambled condition, we observed a significant decrease in the fractional occupancy of the Segregated DMN+VIS ($z_{(66)} = 3.30$, FDR $p = 0.006$, Cohen's $d = 0.47$) during high-comprehension moments compared with low. Interaction effects between Scrambled conditions and comprehension states were significant for the three states (Integrated DMN+VIS: $F_{(1,66)} = 15.81$, Visual: $F_{(1,66)} = 6.41$, Segregated DMN+VIS: $F_{(1,66)} = 4.48$, all p values < 0.05), but not for others (all p values > 0.12). These results imply that the DMN, in tight connection with sensory networks, is highly involved when the narratives are actively being integrated. In contrast, when one focuses on accumulating information from external inputs with a lesser degree of comprehension, the low-level sensory and motor networks take over its role.

Synchronization of underlying brain state dynamics across individuals during novel movie watching

Furthermore, we investigated whether the latent neural state dynamics were synchronized across individuals as they comprehended the same narratives. Figure 7*a* illustrates the inferred neural state sequence of the fMRI participants as they watched an exemplar movie stimulus in the Initial and Repeated Scrambled conditions. The proportion of the moments when the inferred state was identical was calculated for all pairwise participants per movie. The neural state dynamics were more synchronized across participants in the Initial than in the Repeated Scrambled condition (paired Wilcoxon signed-rank test, $z_{(718)} = 18.08$, $p < 0.001$, Cohen's $d = 1.04$; Fig. 7*b*), which was replicated when six latent states were used ($z_{(718)} = 19.29$, $p < 0.001$, Cohen's $d = 1.23$). The higher neural synchrony during novel compared with repeated movie watching was replicated with an ISC analysis, an ROI-based measure of across-subject neural synchrony (Hasson et al., 2004; Nastase et al., 2019). All eight functional networks exhibited a lesser degree of ISC during repeated movie watching compared with participants' first time watching the same movies (paired Wilcoxon signed-rank test between the ISCs of the Initial and Repeated Scrambled conditions; all FDR p values < 0.001, corrected for the number of functional networks). These results suggest that individuals share similar neural dynamics when actively trying to comprehend novel narratives, yet the synchrony decreases when the narratives are no longer novel. The idiosyncratic neural states during repeated movie watching imply that cognitive states may vary across individuals when an active comprehension is no longer required.

Functional brain connectivity predicts evolving cognitive states of comprehension across narratives

Last, we examined whether changes in comprehension, a higher-order cognition that is shared across narratives, can be predicted from patterns of functional brain signatures. Compared with a static predictive model, where a single pattern of brain signature of an individual is related to that person's behavioral score or phenotypic trait (Finn et al., 2015; Rosenberg et al., 2016; Shen et al., 2017), we conducted dynamic predictive modeling, which maps time-resolved brain patterns to the time-resolved behavioral measures (Fig. 8*a*). Dynamically changing brain patterns were captured by the sliding-window-applied FC matrices (i.e., FC pattern-based prediction), or BOLD activation time series of the 122 ROIs (i.e., activation pattern-based prediction). Dynamically changing cognitive states were represented by the group-aggregate measure of comprehension, which was collected from behavioral reports in a separate study. The brain features are different for every fMRI participant, whereas the group behavioral measure is shared across participants who watched the same movie (Fig. 8*a*). To exclude across-participant variance while retaining the temporal variance within an individual, we normalized the time course,

respectively, for each brain feature of each participant. We trained a linear SVR model to predict the moment-to-moment degree of comprehension, given the multivariate neural features at the corresponding time point. To isolate cognitive states associated with general narrative comprehension and exclude narrative-specific, stimulus-driven properties, we adopted across-movie, leave-one-subject-out cross-validation where we trained the model on the data collected from two of the three movies on all participants except one and tested on a held-out participant who watched the held-out movie (Fig. 8*a*). An additional feature selection procedure was used, such that the features that were correlated with the group measure of comprehension, consistent across the training participants (one-sampled *t* test on the correlation coefficients, $p < 0.01$), were selected in each cross-validation fold (Shen et al., 2017).

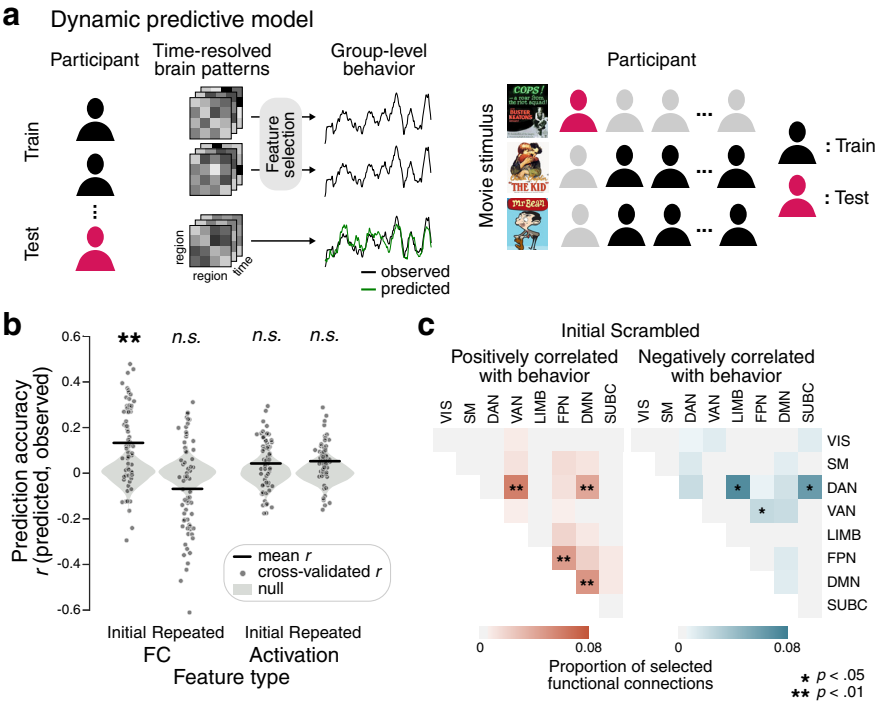

Figure 8. Prediction of the moment-to-moment cognitive states of narrative comprehension from brain patterns. ***a***, Schematic illustrations of dynamic predictive modeling. The model learns the relationship between time-resolved brain patterns (i.e., FC patterns or BOLD activation patterns) and time-resolved cognitive states (i.e., a group-aggregate behavioral measure of comprehension). The brain patterns and a behavioral estimate at every time point from all training participants are treated as independent observations during model training. Feature selection is conducted such that the brain patterns that are significantly correlated with behavioral measures are selected as model features. The trained model is then applied to a held-out individual to predict evolving cognitive states from selected brain features. Prediction accuracy is computed as the Pearson's correlation between the predicted (green line) and observed (black line) behavioral time courses, averaged across cross-validation folds. A cross-validated model is applied to a held-out participant's held-out movie watching scan, from a model trained from the rest of the participants' movie watching scans of different movie stimuli. ***b***, Linear SVR model prediction accuracy (results from all three movie stimuli, $N = 67$). FC pattern-based and activation pattern-based predictions were conducted for the Initial and Repeated Scrambled conditions. Sixty-seven cross-validated prediction accuracies (gray dots) were averaged, and the mean accuracy (black lines) was compared with the null distribution (gray violin plots) in which the same model predicted phase-randomized group measures of comprehension (one-tailed test). ***c***, The proportions of functional connections that were selected in every cross-validation fold during the Initial Scrambled condition, grouped by predefined functional networks. The triangular matrices represent the proportion of functional connections that were positively (left) or negatively (right) correlated with comprehension measures. The functional network pairs of which the proportion of selections was significantly higher than chance are indicated with asterisks (one-tailed test, FDR-corrected for the number of network pairs). LIMB, Limbic network; SUBC, subcortical network.

The results of the predictive modeling are illustrated in Figure 8*b*. When the FC patterns were used as features, the cross-validated model predicted changes in narrative comprehension above chance during the Initial Scrambled ($r = 0.133$, one-tailed $p = 0.003$, nonparametric permutation test with phase-randomized behavioral measures, iteration = 1000), but not during the Repeated Scrambled condition ($r = -0.069$, $p = 0.950$). The difference in cross-validation accuracies between the two Scrambled conditions was significant (paired Wilcoxon signed-rank test, $z_{(66)} = 4.57$, $p < 0.001$, Cohen's $d = 1.02$). However, when the BOLD activation patterns were used as features, we observed no significant prediction performance in both the Initial ($r = 0.044$, $p = 0.184$) and Repeated ($r = 0.054$, $p = 0.113$) Scrambled conditions, with no significant difference in cross-validation accuracies ($z_{(66)} = 0.77$, $p = 0.439$, Cohen's $d = 0.10$). A repeated-measures ANOVA showed a significant interaction effect between the Scrambled conditions and neural feature types ($F_{(1,66)} = 26.26$, $p < 0.001$). The results were replicated using a linear SVM that predicted binary indices of high and low comprehension (FC pattern-based prediction: Initial, 56.01%, $p = 0.010$, Repeated, 47.28%, $p = 0.916$, $z_{(66)} = 4.01$, $p < 0.001$, Cohen's $d = 0.85$; Activation pattern-based prediction: Initial, 51.36%, $p = 0.320$, Repeated, 52.63%, $p = 0.129$, $z_{(66)} = 1.30$, $p = 0.195$, Cohen's $d = 0.22$; interaction, $F_{(1,66)} = 19.91$, $p < 0.001$). The prediction performance was also not significant when the BOLD activation patterns of the entire ROIs were used as features, without feature selection (Initial Scrambled: $r = 0.050$, $p = 0.116$, Repeated Scrambled: $r = 0.040$, $p = 0.139$; $z_{(66)} = 0.306$, $p = 0.745$, Cohen's $d = 0.10$). The results suggest that the cognitive states of comprehension that are generalizable across narratives are robustly predicted by the functional interaction between brain regions, as opposed to the regional activation patterns.

To examine which of the functional connections in the FC pattern-based analysis contributed to predicting cognitive states, we extracted functional connections that were consistently selected in every cross-validation fold (Fig. 8*c*). Among the selected connections, 113 functional connections were positively correlated, and 79 functional connections were negatively correlated with the comprehension measures in the Initial Scrambled condition, whereas only four functional connections were positively correlated and none of the functional connections was negatively correlated with the comprehension measures in the Repeated Scrambled condition. Next, we assigned these consistently selected functional connections to the predefined functional networks (Yeo et al., 2011) and asked whether any of the functional network pairs were selected above chance. The connections between the DAN and VAN, DAN and DMN, and the within-network connections of the FPN and the DMN were selected to be positively correlated with the comprehension measures above chance, in which the null distribution was

computed from the size-matched random networks (all FDR p values < 0.01, one-tailed test, iteration = 10,000, corrected for the number of functional network pairs). The connections between the DAN and limbic network, DAN and subcortical network, and FPN and VAN were selected to be negatively correlated with the comprehension measures during the Initial Scrambled condition (all FDR p values < 0.05). None of the functional network pairs were consistently selected during the Repeated Scrambled condition. The results support our hypothesis that the cognitive states related to narrative comprehension are not restricted to an operation of a particular functional system; rather, they are driven by the dynamic interactions of distributed networks in the brain, in particular the DAN, FPN, and the DMN.

Discussion

Narrative comprehension entails the constant accumulation of information and integration into a causally coherent situational model. We identified the dynamic fluctuation of these cognitive states from participants' behavioral reports of comprehension moments as they watched temporally scrambled movies (Fig. 2). By assessing the causal relationships of the events, we demonstrated that comprehension occurred when an incoming event was strongly causally related to past events, implicating the association of memory reinstatement and narrative integration (Fig. 3). Using fMRI, we showed how large-scale functional brain networks adaptively reconfigure their activity and connective states when individuals engage in comprehending narratives. The systematic modulation of BOLD responses was observed, with higher DMN activity during high comprehension and DAN activity during low comprehension (Fig. 4). Additionally, network-level reconfiguration was aligned to cognitive state changes, such that the functionally integrated and efficient network state occurred during high comprehension, supported by the across-modular connections of the DMN and FPN (Fig. 5). Using a latent state analysis, we showed that the DMN, in tight connection to the sensory processing network, becomes dominant during narrative integration (Fig. 6). The underlying brain states were synchronized across participants when comprehending novel narratives (Fig. 7). We further demonstrated that the evolving comprehension of unseen individuals watching unseen narratives can be predicted by time-resolved functional brain connectivity patterns, but not by regional activation patterns (Fig. 8).

Collectively, the study suggests that narrative comprehension can be characterized by adaptive switches in the dominant information processing modes, transitioning between the accumulation of information when comprehension is low (external mode) and an internal integration into a structured narrative representation when comprehension is high (internal mode) (Dixon et al., 2014; Honey et al., 2018). Although narrative comprehension entails the constant accumulation of incoming events and simultaneous integration of these events into the situational model, our study implies that proportional dominance of information processing modes differs depending on the degree to which narratives are represented in a causally coherent manner. Our behavioral results suggest that the mode switches are driven by the causal dependencies between narrative events, which are connected across time by long-term memory. Our fMRI results also illustrate that alternations between segregated and integrated states of the functional brain network, which were previously suggested to reflect the efficiency and flexibility of the brain in the context of controlled tasks or rest (Bullmore and Sporns, 2012; Zalesky et al., 2014; Gonzalez-Castillo et al., 2015; Shine et al., 2016), are further associated with the externally and internally directed modes of information processing during naturalistic cognition.

During high narrative comprehension, we observed functionally integrated and efficient brain states, with a prominent role of the DMN. In particular, modulation of within- and across-modular functional connections of the DMN drove large-scale reconfiguration of the functional networks (Fig. 5), and a latent state of the tightly connected DMN and sensory processing network was evident during high-comprehension moments (Fig. 6). A recent framework characterized the DMN as a synthesizing network that integrates extrinsic and intrinsic information to construct a contextually coherent situational model (Ranganath and Ritchey, 2012; van den Heuvel and Sporns, 2013; Chang et al., 2021; Yeshurun et al., 2021). Our findings support this functional role of the DMN, such that it integrates incoming events into the accumulated contexts, thus constructing a coherent situational model that is subjectively experienced as a feeling of comprehension. Previous studies suggested that the DMN regions represent discrete narrative events (Baldassano et al., 2017; Chen et al., 2017), modulated by the degree of narrative coherence (Lerner et al., 2011; Honey et al., 2012; Simony et al., 2016) and attention (Dmochowski et al., 2012, 2014; Schmälzle et al., 2015; Ki et al., 2016). Our study adds to these findings and argues that narrative comprehension is not achieved via the workings of specific local regions but rather achieved by a collective operation of large-scale, distributed functional brain networks. The DMN plays a critical role in mediating cooperative connections and adaptive reconfiguration. This hypothesis is supported by our results with dynamic predictive models, which suggest that narrative comprehension can be predicted only with the FC patterns but not with the regional activation patterns.

Notably, previous literature has also related DMN activity to an optimal attentional state, or to moments when stable behavioral performance is maintained during monotonous but attentionally taxing tasks (Esterman et al., 2013; Kucyi et al., 2016a,b, 2020; Van Calster et al., 2017; Zhang et al., 2019; Yamashita et al., 2021). How do we reconcile the recruitment of the DMN during moments of the optimal attentional state (or moments of "in the zone," ease of processing, or effortless attention) (Smallwood et al., 2008; Csikszentmihalyi and Nakamura, 2010; Esterman and Rothlein, 2019), with the view of the DMN as an integrator of external and internal information? In this study, we characterized moments of narrative integration as moments when participants experience subjective feelings of comprehension and moments when the incoming event is causally linked with the past events. Thus, a feeling of comprehension may arise when the relevant past information is retrieved with ease. In that, the workings of the DMN during high-comprehension moments can be broadly related to "an optimal state of information processing" in which information can be integrated, structured, and retrieved with ease. This framework further hints at the mechanism of how distant events in memory that are linked through causal chain are integrated during comprehension. The causal integration hypothesis assumes an ongoing memory retrieval process, where the memory of the causally related events acts as a context to the incoming inputs. We hypothesize that memory retrieval during narrative comprehension would not necessarily require conscious awareness or effortful search of memory by the perceiver. Rather, memory retrieval would occur naturally because the external (i.e., incoming event) and internal (i.e., related events in memory) information is causally chained—acting as a

contextual cue to one another—thereby, eliciting context reinstatement (Polyn et al., 2009; DuBrow et al., 2017; Manning, 2021).

In our study, we assumed that the cognitive state changes related to narrative comprehension would be minimal during the Repeated Scrambled compared with the Initial Scrambled condition because participants no longer had to actively construct a new storyline after having watched the movies in their original order. However, it is important to note the possibility that the cognitive states involved in the two conditions may be qualitatively different in other aspects; for example, participants may have been overall less engaged or aroused during the second viewing, attending to different aspects of the story, such as examining background details or nonimportant characters in the scenes, or they may have been making predictions about the upcoming scenes (Michelmann et al., 2021; C. S. Lee et al., 2021). The result that the latent neural states were less synchronized across individuals during the Repeated Scrambled condition (Fig. 7) reflects that participants may have been involved in a variety of idiosyncratic cognitive states during repeated movie watching. Although the degree to which these other cognitive factors are related to narrative comprehension was not addressed in this study, future research is required to characterize cognitive processes that may differ between the novel and repeated viewing of the narratives.

A limitation of the current study is that we did not consider variance across individuals related to narrative comprehension. A general ability to comprehend narratives largely differs among individuals, and the moments of comprehension during temporally scrambled movie watching may likewise be different. We did not concurrently collect individual-specific measures of comprehension during scans, to exclude possible confounds that may have been caused by performing an additional task. However, the similarities in participants' behavioral reports on moments of comprehension provided a rationale to use the group measure of comprehension to infer cognitive states of fMRI participants. A relevant study (van der Meer et al., 2020) overcame this problem by collecting concurrent physiological measures of heart rate and pupil diameter during scans that acted as dynamic proxies of subjective engagement to the movies. Future work may characterize individual-specific changes in cognitive states and across-individual differences in comprehension, using relevant physiological measures, concurrent behavioral measures, or a *post hoc* behavioral study.

Another consideration is whether a causal reinstatement that occurs during scrambled movie watching reflects comprehension of real-world narratives that progress in order. In our experiments, we artificially scrambled the temporal sequence of the narratives to amplify fluctuations in comprehension. However, real-world narratives do not accompany such drastic changes in cognitive states because comprehension occurs gradually as contextually chained events progressively unfold in time (H. Lee and Chen, 2021). Thus, long-range connections between temporally distant events (Fig. 3*a*) are less likely (though present in narratives) because event representations accumulate and integrate gradually (Franklin et al., 2020). Nevertheless, although fluctuations of cognitive states, as well as the necessity of causal reinstatement, would not be as prominent as in experimental context, similar cognitive and neural mechanisms—transitions between external and internal modes of information processing, which are enacted by functional segregation and integration of brain networks—would also be involved during comprehension of real-world narratives.

Lastly, future studies are encouraged to investigate the underlying computational mechanisms explaining the accumulation and integration of narratives (Chien and Honey, 2020; Franklin et al., 2020; Lu et al., 2020) in relation to the neural representation formed on a moment-to-moment basis (Wehbe et al., 2014; Huth et al., 2016; Vodrahalli et al., 2018). Our study provides insights on how representations of discrete events across distant times in memory, chained with causal relations, may be integrated with accumulated inputs in real time to form coherent narratives. Communication of information between functional modules of the brain during comprehension may be implemented via modulation of the FC and global reconfiguration of the dynamic brain states.

References

Achard S, Bullmore E (2007) Efficiency and cost of economical brain functional networks. PLoS Comput Biol 3:e17.

Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD (2014) Tracking whole-brain connectivity dynamics in the resting state. Cereb Cortex 24:663–676.

Aly M, Chen J, Turk-Browne NB, Hasson U (2018) Learning naturalistic temporal structure in the posterior medial network. J Cogn Neurosci 30:1345–1365.

Andrews-Hanna JR, Smallwood J, Spreng RN (2014) The default network and self-generated thought: component processes, dynamic control, and clinical relevance. Ann NY Acad Sci 1316:29–52.

Baker AP, Brookes MJ, Rezek IA, Smith SM, Behrens T, Probert Smith PJ, Woolrich M (2014) Fast transient networks in spontaneous human brain activity. Elife 3:e01867.

Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, Norman KA (2017) Discovering event structure in continuous narrative perception and memory. Neuron 95:709–721.e5.

Baldassano C, Hasson U, Norman KA (2018) Representation of real-world event schemas during narrative perception. J Neurosci 38:9689–9699.

Barttfeld P, Uhrig L, Sitt JD, Sigman M, Jarraya B, Dehaene S (2015) Signature of consciousness in the dynamics of resting-state brain activity. Proc Natl Acad Sci USA 112:887–892.

Bassett DS, Porter MA, Wymbs NF, Grafton ST, Carlson JM, Mucha PJ (2013) Robust detection of dynamic community structure in networks. Chaos 23:013142.

Beckmann CF, DeLuca M, Devlin JT, Smith SM (2005) Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci 360:1001–1013.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188.

Betzel RF, Byrge L, Esfahlani FZ, Kennedy DP (2020) Temporal fluctuations in the brain's modular architecture during movie-watching. Neuroimage 213:116687.

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Mech 2008:P10008.

Braga RM, Buckner RL (2017) Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. Neuron 95:457–471.e5.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198.

Bullmore E, Sporns O (2012) The economy of brain network organization. Nat Rev Neurosci 13:336–349.

Calinski T, Harabasz J (1974) A dendrite method for clustering analysis. Comm Stats Theory Methods 3:1–27.

Cer D, Yang Y, Kong S, Hua N, Limtiaco N, John RS, Constant N, Guajardo-Cespedes M, Yuan S, Tar C, Sung YH, Strope B, Kurzweil R (2018) Universal sentence encoder. arXiv 1803.11175.

Chang CH, Lazaridi C, Yeshurun Y, Norman KA, Hasson U (2021) Relating the past with the present: information integration and segregation during ongoing narrative processing. J Cogn Neurosci 33:1106–1128.

Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, Hasson U (2017) Shared memories reveal shared structure in neural activity across individuals. Nat Neurosci 20:115–125.

Chien HY, Honey CJ (2020) Constructing and forgetting temporal context in the human cerebral cortex. Neuron 106:675–686.e11.

Csikszentmihalyi M, Nakamura J (2010) Effortless attention in everyday life: a systematic phenomenology. In: Effortless attention: a new perspective in the cognitive science of attention and action (Bruya B, eds), pp 179–189. Cambridge, MA: Massachusetts Institute of Technology.

Danek AH, Wiley J (2017) What about false insights? Deconstructing the aha! experience along its multiple dimensions for correct and incorrect solutions separately. Front Psychol 7:2077.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B 39:1–38.

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

Dixon ML, Fox KC, Christoff K (2014) A framework for understanding the relationship between externally and internally directed cognition. Neuropsychologia 62:321–330.

Dmochowski JP, Bezdek MA, Abelson BP, Johnson JS, Schumacher EH, Parra LC (2014) Audience preferences are predicted by temporal reliability of neural processing. Nat Commun 5:4567.

Dmochowski JP, Sajda P, Dias J, Parra LC (2012) Correlated components of ongoing EEG point to emotionally laden attention: a possible marker of engagement? Front Hum Neurosci 6:112.

DuBrow S, Rouhani N, Niv Y, Norman KA (2017) Does mental context drift or shift? Curr Opin Behav Sci 17:141–146.

Esterman M, Noonan SK, Rosenberg M, Degutis J (2013) In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. Cereb Cortex 23:2712–2723.

Esterman M, Rothlein D (2019) Models of sustained attention. Curr Opin Psychol 29:174–180.

Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, Yang Z, Chu C, Xie S, Laird AR, Fox PT, Eickhoff SB, Yu C, Jiang T (2016) The Human Brainnetome Atlas: a new brain atlas based on connectional architecture. Cereb Cortex 26:3508–3526.

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci 18:1664–1671.

Finn ES, Corlett PR, Chen G, Bandettini PA, Constable RT (2018) Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. Nat Commun 9:2043.

Fischl B (2012) FreeSurfer. Neuroimage 62:774–781.

Fortunato S (2010) Community detection in graphs. Phys Rep 486:75–174.

Franklin NT, Norman KA, Ranganath C, Zacks JM, Gershman SJ (2020) Structured event memory: a neuro-symbolic model of event cognition. Psychol Rev 127:327–361.

Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9:432–441.

Gao S, Mishne G, Scheinost D (2021) Nonlinear manifold learning in functional magnetic resonance imaging uncovers a low-dimensional space of brain dynamics. Hum Brain Mapp 42:4510–4524.

Gonzalez-Castillo J, Hoy CW, Handwerker DA, Robinson ME, Buchanan LC, Saad ZS, Bandettini PA (2015) Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. Proc Natl Acad Sci USA 112:8762–8767.

Gordon EM, Laumann TO, Marek S, Raut R, Gratton C, Newbold DJ, Greene DJ, Coalson RS, Snyder AZ, Schlaggar BL, Petersen SE, Dosenbach NU, Nelson SM (2020) Default-mode network streams for coupling to language and control systems. Proc Natl Acad Sci USA 117:17308–17319.

Graesser AC, Singer M, Trabasso T (1994) Constructing inferences during narrative text comprehension. Psychol Rev 101:371–395.

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM (2014) ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage 95:232–247.

Griffanti L, Douaud G, Bijsterbosch J, Evangelisti S, Alfaro-Almagro F, Glasser MF, Duff EP, Fitzgibbon S, Westphal R, Carone D, Beckmann CF, Smith SM (2017) Hand classification of fMRI ICA noise components. Neuroimage 154:188–205.

Guimerà R, Nunes Amaral LA (2005) Functional cartography of complex metabolic networks. Nature 433:895–900.

Handwerker DA, Roopchansingh V, Gonzalez-Castillo J, Bandettini PA (2012) Periodic changes in fMRI connectivity. Neuroimage 63:1712–1719.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. Science 303:1634–1640.

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A hierarchy of temporal receptive windows in human cortex. J Neurosci 28:2539–2550.

Honey CJ, Newman EL, Schapiro AC (2018) Switching between internal and external modes: a multiscale learning principle. Netw Neurosci 1:339–356.

Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U (2012) Slow cortical dynamics and the accumulation of information over long timescales. Neuron 76:423–434.

Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Della Penna S, Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker DA, Keilholz S, Kiviniemi V, Leopold DA, de Pasquale F, Sporns O, Walter M, Chang C (2013) Dynamic functional connectivity: promise, issues, and interpretations. Neuroimage 80:360–378.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532:453–458.

Ki JJ, Kelly SP, Parra LC (2016) Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. J Neurosci 36:3092–3101.

Kucyi A, Esterman M, Riley CS, Valera EM (2016a) Spontaneous default network activity reflects behavioral variability independent of mind-wandering. Proc Natl Acad Sci USA 113:13899–13904.

Kucyi A, Hove MJ, Esterman M, Hutchison RM, Valera EM (2016b) Dynamic brain network correlates of spontaneous fluctuations in attention. Cereb Cortex 27:1831–1840.

Kucyi A, Daitch A, Raccah O, Zhao B, Zhang C, Esterman M, Zeineh M, Halpern CH, Zhang K, Zhang J, Parvizi J (2020) Electrophysiological dynamics of antagonistic brain networks reflect attentional fluctuations. Nat Commun 11:325.

Landauer TK, Dumais ST (1997) A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol Rev 104:211–240.

Langston M, Trabasso T (1999) Modeling causal integration and availability of information during comprehension of narrative texts. In: The construction of mental representations during reading, pp 29–69. Mahwah, NJ: Lawrence Erlbaum.

Latora V, Marchiori M (2001) Efficient behavior of small-world networks. Phys Rev Lett 87:198701.

Lee CS, Aly M, Baldassano C (2021) Anticipation of temporally structured events in the brain. Elife 10:e64972.

Lee H, Chen J (2021) Narratives as networks: predicting memory from the structure of naturalistic events. bioRxiv 441287. doi: 10.1101/2021.04.24.441287.

Leonardi N, Van De Ville D (2015) On spurious and real fluctuations of dynamic functional connectivity during rest. Neuroimage 104:430–436.

Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J Neurosci 31:2906–2915.

Liégeois R, Ziegler E, Phillips C, Geurts P, Gómez F, Bahri MA, Yeo BT, Soddu A, Vanhaudenhuyse A, Laureys S, Sepulchre R (2016) Cerebral functional connectivity periodically (de)synchronizes with anatomical constraints. Brain Struct Funct 221:2985–2997.

Lu Q, Hasson U, Norman KA (2020) Learning to use episodic memory for event prediction. bioRxiv 422882. doi: 10.1101/2020.12.15.422882.

Manning J (2021) Episodic memory: mental time travel or a quantum 'memory wave' function? Psychol Rev 128:711–725.

Mar RA (2004) The neuropsychology of narrative: story comprehension, story production and their interrelation. Neuropsychologia 42:1414–1434.

Margulies DS, Ghosh SS, Goulas A, Falkiewicz M, Huntenburg JM, Langs G, Bezgin G, Eickhoff SB, Castellanos FX, Petrides M, Jefferies E, Smallwood J (2016) Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc Natl Acad Sci USA 113:12574–12579.

Mesulam MM (1998) From sensation to cognition. Brain 121:1013–1052.

Michelmann S, Price AR, Aubrey B, Doyle WK, Friedman D, Dugan PC, Devinsky O, Devore S, Flinker A, Hasson U, Norman KA (2021) Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. Nat Commun 12:5394.

Murphy C, Jefferies E, Rueschemeyer SA, Sormaz M, Wang H, ting Margulies DS, Smallwood J (2018) Distant from input: evidence of regions within the default mode network supporting perceptually-decoupled and conceptually-guided cognition. Neuroimage 171:393–401.

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

Newman ME (2004) Fast algorithm for detecting community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys 69:066133.

Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci USA 103:8577–8582.

Nishida S, Nishimoto S (2018) Decoding naturalistic experiences from human brain activity via distributed representations of words. Neuroimage 180:232–242.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442.

Polyn SM, Norman KA, Kahana MJ (2009) A context maintenance and retrieval model of organization. Psychol Rev 116:129–156.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154.

Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE (2014) Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84:320–341.

Preti MG, Bolton TA, Van De Ville D (2017) The dynamic functional connectome: state-of-the-art and perspectives. Neuroimage 160:41–54.

Quinn AJ, Vidaurre D, Abeysuriya R, Becker R, Nobre AC, Woolrich MW (2018) Task-evoked dynamic network analysis through hidden Markov modeling. Front Neurosci 12:603.

Rabiner LR, Juang BH (1986) An introduction to hidden Markov models. IEEE ASSP Mag 3:4–16.

Ranganath C, Ritchey M (2012) Two cortical systems for memory-guided behaviour. Nat Rev Neurosci 13:713–726.

Rezek I, Roberts S (2005) Ensemble hidden Markov models with extended observation densities for biosignal analysis. In: Probabilistic modeling in bioinformatics and medical informatics (Husmeier D, Dybowski R, Roberts S, eds), pp 419–450. London: Springer.

Ritchey M, Cooper RA (2020) Deconstructing the posterior medial episodic network. Trends Cogn Sci 24:451–465.

Rosenberg MD, Finn ES, Scheinost D, Papademetris X, Shen X, Constable RT, Chun MM (2016) A neuromarker of sustained attention from whole-brain functional connectivity. Nat Neurosci 19:165–171.

Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52:1059–1069.

Rubinov M, Sporns O (2011) Weight-conserving characterization of complex functional brain networks. Neuroimage 56:2068–2079.

Sadaghiani S, Poline JB, Kleinschmidt A, D'Esposito M (2015) Ongoing dynamics in large-scale functional connectivity predict perception. Proc Natl Acad Sci USA 112:8463–8468.

Sakoğlu Û, Pearlson GD, Kiehl KA, Wang YM, Michael AM, Calhoun VD (2010) A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. Magn Reson Mater Phy 23:351–366.

Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM (2014) Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90:449–468.

Schmälzle R, Häcker FE, Honey CJ, Hasson U (2015) Engaged listeners: shared neural processing of powerful political speeches. Soc Cogn Affect Neurosci 10:1137–1143.

Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, Constable RT (2017) Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat Protoc 12:506–518.

Shine JM, Bissett PG, Bell PT, Koyejo O, Balsters JH, Gorgolewski KJ, Moodie CA, Poldrack RA (2016) The dynamics of functional brain networks: integrated network states during cognitive task performance. Neuron 92:544–554.

Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD (2012) Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex 22:158–165.

Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, Hasson U (2016) Dynamic reconfiguration of the default mode network during narrative comprehension. Nat Commun 7:12141.

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

Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, Beckmann CF (2009) Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci USA 106:13040–13045.

Tononi G, Sporns O, Edelman GM (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc Natl Acad Sci USA 91:5033–5037.

Trabasso T, Sperry LL (1985) Causal relatedness and importance of story events. J Mem Lang 24:595–611.

Van Calster L, D'Argembeau A, Salmon E, Peters F, Majerus S (2017) Fluctuations of attentional networks and default mode network during the resting state reflect variations in cognitive states: evidence from a novel resting-state experience sampling method. J Cogn Neurosci 26:1–19.

van den Heuvel MP, Sporns O (2013) Network hubs in the human brain. Trends Cogn Sci 17:683–696.

van der Meer JN, Breakspear M, Chang LJ, Sonkusare S, Cocchi L (2020) Movie viewing elicits rich and reliable brain state dynamics. Nat Commun 11:5004.

Vidaurre D, Abeysuriya R, Becker R, Quinn AJ, Alfaro-Almagro F, Smith SM, Woolrich MW (2018) Discovering dynamic brain networks from big data in rest and task. Neuroimage 180:646–656.

Vidaurre D, Smith SM, Woolrich MW (2017) Brain network dynamics are hierarchically organized in time. Proc Natl Acad Sci USA 114:12827–12832.

Vodrahalli K, Chen PH, Liang Y, Baldassano C, Chen J, Yong E, Honey C, Hasson U, Ramadge P, Norman KA, Arora S (2018) Mapping between fMRI responses to movies and their natural language annotations. Neuroimage 180:223–231.

Walther D, Koch C (2006) Modeling attention to salient proto-objects. Neural Netw 19:1395–1407.

Wehbe L, Murphy B, Talukdar P, Fyshe A, Ramdas A, Mitchell T (2014) Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. PLoS One 9:e112575.

Wolfe MB, Magliano J, Larsen B (2005) Causal and semantic relatedness in discourse understanding and representation. Discourse Process 39:165–187.

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

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

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fisch B, Liu H, Buckner RL (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106:1125–1165.

Yeo BT, Tandi J, Chee MW (2015) Functional connectivity during rested wakefulness predicts vulnerability to sleep deprivation. Neuroimage 111:147–158.

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

Zacks JM, Swallow KM (2007) Event segmentation. Curr Dir Psychol Sci 16:80–84.

Zalesky A, Fornito A, Cocchi L, Gollo LL, Breakspear M (2014) Time-resolved resting-state brain networks. Proc Natl Acad Sci USA 111:10341–10346.

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

Zwaan RA, Langston MC, Graesser AC (1995) The construction of situation models in narrative comprehension: an event-indexing model. Psychol Sci 6:292–297.