Brain meta-state transitions demarcate spontaneous thoughts in movie-viewing and resting cognition

Tseng, Julie[1], & Poppenk, Jordan[1,2,3,*]

1. Centre for Neuroscience Studies, Queen's University
2. Department of Psychology, Queen's University
3. School of Computing, Queen's University

* Corresponding author, 62 Arch St., Kingston, ON, K7L 3N6, 613-533-6009, jpoppenk@queensu.ca.

Abstract

Using new methods to capture transitions in brain network meta-state, we characterize the psychological meaning of these regular neural events. Drawing on a large fMRI dataset, we found participants' meta-state transitions were triggered primarily by the progression of meaning across events within movie stimuli, and that movies with more events induced greater alignment of transitions among participants. Meta-state transitions also shared many characteristics across movie and resting fMRI data, including trait-like rates of occurrence, concurrence with activation of brain regions associated with spontaneous thought, and suppression by engagement of attention regions. Based on these features, as well as the centrality of semantics to thought, we argue that the mental events marked by meta-state transitions correspond to the initiation of new thoughts, and that meta-state transitions therefore may serve as an implicit neurobiological marker for demarcating thoughts. As larger anterior hippocampi predicted lower transition rates – possibly reflecting more dwell time in memories following hippocampal recall – our findings illustrate how this approach can contribute insights to emerging research on thought dynamics.

Research on thought, such as investigation of mind-wandering, has emerged into the mainstream of cognitive neuroscience[1]. Problems do remain to be overcome, including the unreliable and disruptive nature of meta-cognition[2,3], and uncertainty about how to pinpoint a thought's beginning and end[4]. However, as most definitions of a thought concern its contents (with some definitions emphasizing spontaneous production[5]), we propose that implicit measurement of changes in semantic content offers an interesting alternative to meta-cognition. Researchers can readily decode neurally discernable object categories (e.g., faces and houses) from spatial patterns in participants' functional magnetic resonance imaging (fMRI) data[6]; accordingly, one could infer a new thought has arisen in a participant by observing a switch from one active category to another. One possible reason this strategy has not been employed is that it is challenging to reliably distinguish large numbers of categories, and object categories alone may be inadequate for representing the complexity of our cognitions. However, recent perspectives emphasize measurement of thought dynamics instead of contents[1]. What if, rather than tracking the rise and fall of particular object categories, we found a way to track semantic transitions more globally?

Here we describe a new method to identify breaks between stable periods of brain network configuration (i.e., meta-state transitions), and present a psychological validation based on the correspondence of these transitions with features of movie stimuli during participant movie-viewing. We further assess the neural landmark's temporal and spatial characteristics, and explore its generalization from movie-viewing to unconstrained, task-free brain data. Based on the observed patterns, we argue that the psychological events that co-occur with meta-state transitions correspond to new thoughts. We then illustrate the utility of our approach for understanding thought dynamics by reporting a correlation between transition rate, which we argue aligns with the rate of new thoughts, and neuroanatomical features of the hippocampus.

We conducted our analysis on the 7T Human Connectome Project dataset, which features movie-viewing fMRI (mv-fMRI) and resting-state fMRI (rs-fMRI) data gathered from 184 participants[7–9]. We converted each fMRI run into the expression of 15 known brain networks over time (see Fig. S1 in *Methods*), then reduced its dimensionality from (*15 x time)* to (*2 x time)* using t-SNE[10,11]. In this reduced space, epochs with similar patterns of network activity fall in proximity. We hypothesized a pattern of spatiotemporal organization reflecting progression through a series of discrete thoughts, each centred around its own semantic focal point (e.g., what one will be having for dinner) serving as an attractor. Unstable network meta-states would yield dispersion in this space, whereas an attractor would cause points to cluster, yielding a worm-like series (arising from limited drift as thoughts "evolve"; Fig. 1a vs 1b). We found strong evidence for spatial contiguity with participant trajectories reliably forming into series (see Fig. 1c, Fig. S2, and *Methods*).

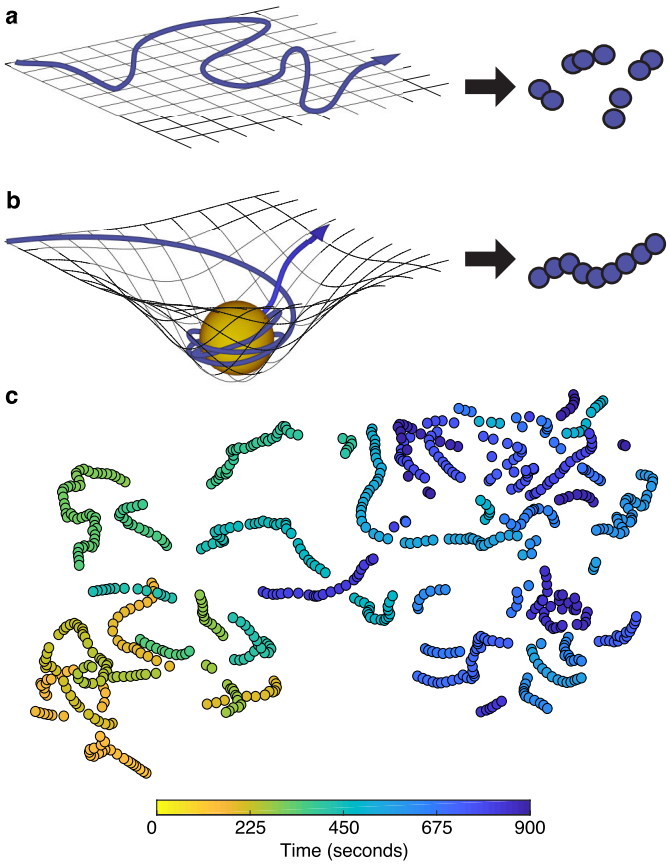


Fig. 1. Network space representation. (**a**) Continuous, random passage through the space of possible network configurations generates fragments in t-SNE space, as opposed to (**b**) contiguous, worm-like segments when an attractor holds network configurations in relative meta-stability. (**c**) An example reduced t-SNE representation involving both segment types, as observed from one participant's 15-minute rs-fMRI scan.

Next, we identified changes in network activity by taking the squared Mahalanobis distance[12] between successive timepoints in t-SNE space for each fMRI run, obtaining a measure of meta-state change that we label a *step distance vector*. To stabilize the step distance vector, we repeated the dimensionality reduction and step distance vector creation process 100 times for each participant and each functional run. Peaks within the resulting *mean* step distance vector represent prominent reconfigurations of network meta-states, thus we called them network meta-state transitions (henceforth *transitions*). For purposes of baseline comparison, we also identified local minima in the mean step distance vector, which each represent a relatively stable network meta-state (henceforth *meta-stable*) (Fig. 2a).

To assess whether these discovered moments of network reorganization held psychological relevance, we validated our approach using mv-fMRI data, examining their alignment to the onset of new semantic or perceptual movie features. As a starting point, visual inspection of the set of all participant transitions during movie-viewing revealed substantial alignment in transitions relative to rest (Fig. 2b, c). To quantify this, we obtained each participant's group alignment for each resting state and movie-viewing run, which describes the correlation between the individual's step distance vector and the corresponding median group step distance vector (i.e., conformity; see *Methods*). We separated conformity values into runs of the same type, resulting in 723 movie conformity values and 722 rest conformity values. After feeding each set of values into a group bootstrap analysis, we found higher conformity for movie runs, mean $r = 0.27$, 95% CI: [0.26, 0.27], than for rest, mean $r = 0.04$, 95% CI: [0.03, 0.04], r difference = 0.23, 95% CI: [0.22, 0.25]. This finding reflects past observations of film's unique ability to induce similar activity across participants in a wide variety of brain areas[13–15], and provides a preliminary link between transitions and naturalistic cognition.

As an alternate means of evaluating the influence of plot progression (i.e., progression of meaning) over transitions, we attempted to predict transition alignment based on the number of narrative events in each clip. Two expert raters came to a consensus on boundaries between events[16], which we defined as timepoints where a change in the movie triggers a new semantic focal point or evolves viewer understanding of the movie narrative. We correlated the number of these events in each clip with group alignment within each participant. This yielded 184 correlation coefficients that we entered into a group bootstrap analysis. Clips with more events per minute had higher group alignment, mean $r = 0.25$, 95% CI: [0.21, 0.29].

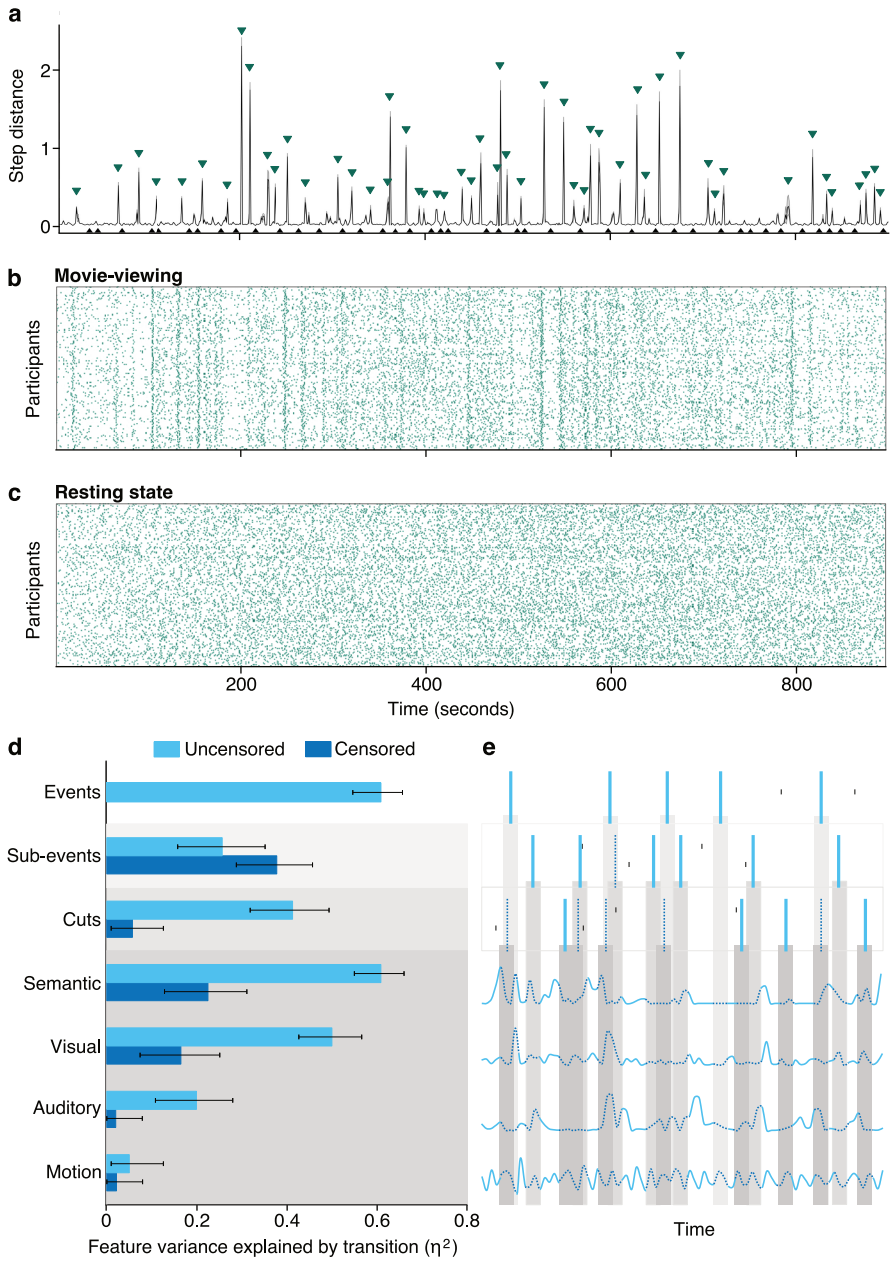


Fig. 2. Explaining network meta-state transitions using movie features. (**a**) A participant's mean step distance vector during one mv-fMRI run, with 95% CI ribbon (narrow ribbon indicates stability over t-SNE iterations). *Transition* timepoints (green triangles) and *meta-stable* timepoints (black triangles) identified by a peak finding algorithm. (**b**) All participants' transition timepoints for the same mv-fMRI run, with many peaks overlapping those of the example participant in **a**. (**c**) All participants' transitions for one rs-fMRI run. Alignment of peaks in **a** but not **c** reveals stimulus control over transitions. (**d**) Eta-squared values describing the proportion of variance in movie features explained by alignment to transition vs. meta-stable timepoints. Error bars designate 95% confidence intervals. (**e**) Diagram illustrating censorship of lower-level feature vectors by higher-level features for greater independence of test statistics.

We also directly examined the temporal correspondence of transitions with other movie features within-participant. Consensus labels of sub-events (consisting of individual actions) and cuts were obtained from two expert raters. We also obtained lower-level feature timeseries for each clip that describe semantic, visual, and amplitude change, as well as change in head motion. No other features were tested. Some features were correlated; for example, an event's end often coincides with a cut between shots, which in turn coincides with semantic and perceptual stimulus changes. These correlations could produce a result wherein transitions appear to reflect lower-level features, but only because they peak concurrently with the onset of new cuts or events. To disentangle high- and low-level feature contributions, we "censored" epochs of lower-level feature timeseries where higher-level event boundaries co-occurred (Fig. 2e). We reasoned that if lower-level features induce transitions, this should remain the case outside of censored epochs.

To determine the proportion of feature variance accounted for by transition versus meta-stable timepoints, we calculated eta-squared values for both uncensored and censored features after accounting for lag in the hemodynamic response function. Transitions (and specifically t-SNE-derived transitions; Fig. S3) were strongly associated with movie features, with up to 60.8% of feature variance explained (event feature). However, with the exception of sub-events, eta-squared values for all features decreased markedly after removing timepoints that could be confounded with event boundaries. Features corresponding to non-visual perceptual change, in particular, dropped to near-zero values (Fig. 2d). Thus, transitions showed a clear, albeit non-exclusive alignment to features pertaining to plot progression, with the top three predictors being event, sub-event and semantic features (see Movie S1 for illustration of this specificity). These results complement those of Baldassano and colleagues[15], who used a Hidden Markov Model (HMM) approach to find shifts in movie-viewing brain activity and found that these shifts often co-occurred with event boundaries.

We interpret this co-occurrence with various movie features as evidence that meta-state transitions correspond with psychologically meaningful "mental events". To learn whether these mental events were observable even outside of stimulus control, we interrogated possible similarities in the properties of transitions found in mv-fMRI and rs-fMRI data. First, we explored whether any consistent spatial pattern differentiated transition and meta-stable timepoints. Although transitions were defined on the basis of change in network activation, they captured both activation and deactivation of each network, so there was no circularity or bias towards the spatial characteristics of any one network. We performed a conjunction analysis, searching for voxel activation clusters that independently distinguished transition and meta-stable timepoints for both mv-fMRI and rs-fMRI runs. We found that regardless of whether participants were engaged in movie-viewing or at rest, transitions were associated with regions previously implicated in spontaneous thought (Fig. 3a-c, Table S1).

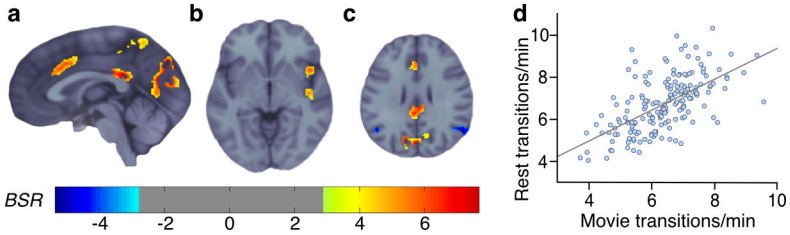


Fig. 3. Transition features generalize from stimulus-driven to resting cognition. (**a-c**) Conjunction analysis revealed brain activity that consistently differentiated transition from meta-stable timepoints, regardless of task. Images are shown in neurological orientation. (**a**) Anterior and posterior cingulate, precuneus, and visual association area were implicated in transitions. (**b**) Anterior and posterior insula were implicated in transitions. (**c**) Midline anterior, posterior cingulate, and visual association cortex were implicated in transitions; angular gyrus was implicated in meta-stability. (**d**) Participants' movie-viewing transition rates correlated with their resting transition rates.

We also evaluated whether an individual's *transition rate* generalized across task and rest. First, we tested its internal stability across the four rs-fMRI runs (as each was sampled in a separate scanner session). Transition rate was correlated across runs (mean $r = 0.45$, SD = 0.06), revealing it to be a trait-like characteristic that can be adequately sampled using 15 minutes of rs-fMRI data. We then averaged participant-wise transition rates across rs-fMRI runs to stabilize the trait measurement, and regressed them against the average participant-wise transition rates computed from mv-fMRI runs. Transition rate at rest was correlated to transition rate during movie-viewing, $r(181) = 0.60$, 95% CI: [0.51, 0.70] (Fig. 3d), thereby further linking its properties across stimulus-driven and resting cognition.

Finally, based on past reports that the hippocampus is sensitive to similar stimulus features (including both movie onset/offset[17] and event boundaries[15]), we investigated hippocampus volumetry as a potential correlate of transition rate. Our group has previously proposed that, whereas posterior hippocampus (pHPC) helps to distinguish similar events in memory[18], anterior hippocampus (aHPC) supports gist-like representations[19]. Accordingly, we hypothesized that because aHPC is associated with greater pattern completion[20] because of a relatively small dentate gyrus[21], a larger aHPC would more readily cue memories that capture participants' thoughts, serving as an attractor that delays further mind-wandering (Fig. 1b). We therefore tested whether a larger aHPC would predict a lower transition rate. We averaged rs-fMRI and mv-fMRI transition rates for a maximally stable and singular trait assessment of transition rate. Consistent with our predictions, we found that a larger anterior hippocampus and smaller pHPC:aHPC ratio predicted a lower transition rate (Fig. 4).

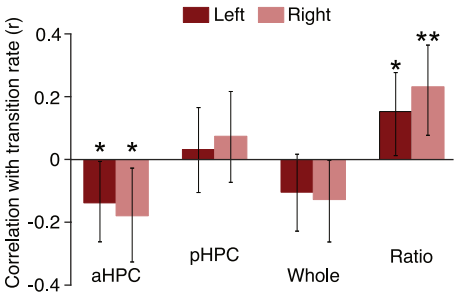


Fig. 4. Grounding of transitions in neurobiology of event cognition. Prediction of transition rate by volumes of anterior hippocampus (aHPC), posterior hippocampus (pHPC), whole hippocampus, and the pHPC to aHPC volume ratio using transition rates derived from movie-viewing and rest data. Error bars designate 95% confidence intervals. * designates correlations significant at $p < .05$; ** designates correlations significant at $p < 0.005$.

To summarize, we used novel methods to characterize a neurocognitive landmark based on transitions between brain network meta-states. Our approach is distinctive for requiring no prior knowledge about the complexity of brain data under interrogation or the meta-states to be visited, and is straightforward to implement. We found transitions to be highly responsive to the progression of meaning across events within movie stimuli, and also found transition rate to be a reliable, trait-like characteristic across tasks and imaging sessions. Characteristics of movie-viewing transitions were shared by resting-state transitions, supporting a unified interpretation of transitions across task.

We also found transitions co-occurred with activation of regions associated with spontaneous thought. In particular, anterior cingulate and insular cortices are considered members of a salience network that shifts attention to novel external and internal events[22]; the posterior cingulate cortex is a key region in the default mode network that is more active during task-unrelated than task-related thought[23,24]. By contrast, we found that angular gyrus activation, which has been associated with both event representation[25] and sustained attention[26], predicted meta-*stability,* perhaps reflecting decreased likelihood of transitions while participants focally contemplated an event.

A thought is grounded in its contents. Therefore, drawing on our findings that meta-state transitions predicted plot progression, that they evoke neural activity resembling spontaneous thought, and that they occur at a steady rate across naturalistic and resting forms of cognition, we argue that transitions may serve as a biological marker of new thoughts. Moreover, through negative correlation of transition rate with anterior hippocampus volume – perhaps a result of increased dwell time within attractors tied to more readily accessible hippocampal memories – our findings illustrate how this marker could also contribute to a better understanding of thought dynamics.

We found meta-state transitions to be a prominent, trait-like feature in cognition, and although our analysis has focused on movie and resting fMRI data from healthy adults, our methods are applicable to a wide range of tasks, populations, and even case studies. By offering far better reliability than past (introspective) approaches to measuring thought, they make new questions accessible: for example, how many thoughts do we experience each waking day? Extrapolating from a median transition rate across movie-viewing and rest of 6.5 transitions / min, and a recommended sleep time of 8 hours, one could estimate over six thousand daily thoughts for healthy adults of a young-adult demographic similar to the one used in our analysis. Although further interrogation of meta-state transitions would be needed to employ such a measure with confidence, availability of a tentative answer highlights how the current approach may be fruitful in advancing how we think about thought.

References

1. Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N. & Andrews-Hanna, J. R. Mind-wandering as spontaneous thought: a dynamic framework. *Nat. Rev. Neurosci.* **17**, 718–731 (2016).
2. Nisbett, R. E. & Wilson, T. D. Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* **84**, 231–259 (1977).
3. Seli, P. *et al.* How pervasive is mind wandering, really? *Conscious. Cogn.* **66**, 74–78 (2018).
4. James, W. *The principles of psychology*. **2**, (H. Holt and Company, 1890).
5. Smallwood, J. Distinguishing how from why the mind wanders: a process-occurrence framework for self-generated mental activity. *Psychol. Bull.* **139**, 519–535 (2013).
6. Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* **10**, 424–430 (2006).
7. Van Essen, D. C. *et al.* The Human Connectome Project: A data acquisition perspective. *Neuroimage* **62**, 2222–2231 (2012).
8. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
9. Ugurbil, K. *et al.* Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* **80**, 80–104 (2013).
10. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
11. Billings, J. C. W. *et al.* Instantaneous brain dynamics mapped to a continuous state space. *Neuroimage* **162**, 344–352 (2017).
12. Gnanadesikan, R. & Kettenring, L. R. Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data. *Biometrics* **28**, 81–124 (1972).
13. Hasson, U. *et al.* Neurocinematics: The Neuroscience of Film. *Projections* **2**, 1–26 (2008).
14. Chen, J. *et al.* Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).
15. Baldassano, C. *et al.* Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709–721 (2017).
16. Zacks, J. M. & Swallow, K. M. Event Segmentation. **16**, 80–84 (2007).
17. Ben-Yakov, A. & Dudai, Y. Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. *J. Neurosci.* **31**, 9032–9042 (2011).
18. Poppenk, J. & Moscovitch, M. A hippocampal marker of recollection memory ability among healthy young adults: Contributions of posterior and anterior segments. *Neuron* **72**, 931–937 (2011).
19. Poppenk, J., Evensmoen, H. R., Moscovitch, M. & Nadel, L. Long-axis specialization of the human hippocampus. *Trends Cogn. Sci.* **17**, 230–240 (2013).
20. Yassa, M. A. & Stark, C. E. L. Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–525 (2011).
21. Malykhin, N. V., Lebel, R. M., Coupland, N. J., Wilman, A. H. & Carter, R. In vivo

quantification of hippocampal subfields using 4.7 T fast spin echo imaging. *Neuroimage* **49**, 1224–1230 (2010).
22. Seeley, W. W. *et al.* Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. **27**, 2349–2356 (2007).
23. Leech, R. & Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease. *Brain* **137**, 12–32 (2013).
24. Smallwood, J. *et al.* Representing representation: Integration between the temporal lobe and the posterior cingulate influences the content and form of spontaneous thought. *PLoS One* **11**, 1–19 (2016).
25. Seghier, M. L. The Angular Gyrus. *Neurosci.* **19**, 43–61 (2012).
26. Singh-Curry, V. & Husain, M. The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia* **47**, 1434–1448 (2009).
27. Nickerson, L. D., Smith, S. M., Öngür, D. & Beckmann, C. F. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Front. Neurosci.* **11**, 1–18 (2017).
28. Beckmann, C., Mackay, C., Filippini, N. & Smith, S. Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *Neuroimage* **47**, S148 (2009).
29. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
30. Moeller, S. *et al.* Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn. Reson. Med.* **63**, 1144–1153 (2010).
31. Feinberg, D. A. *et al.* Multiplexed Echo Planar Imaging for Sub-Second Whole Brain FMRI and Fast Diffusion Imaging. *PLoS One* **5**, e15710 (2010).
32. Setsompop, K. *et al.* Blipped-Controlled Aliasing in Parallel Imaging (blipped-CAIPI) for simultaneous multi-slice EPI with reduced g-factor penalty. *Magn. Reson. Med.* **67**, 1210–1224 (2012).
33. Xu, J. *et al.* Highly Accelerated Whole Brain Imaging Using Aligned-Blipped-Controlled-aliasing Multiband EPI. in *Proceedings of the 20th Annual Meeting of ISMRM* **2306**, (2012).
34. Miller, R. L. *et al.* Higher Dimensional Meta-State Analysis Reveals Reduced Resting fMRI Connectivity Dynamism in Schizophrenia Patients. *PLoS One* **11**, e0149849 (2016).
35. Team, D. Datavyu: A Video Coding Tool. (2014).
36. Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **14**, 1210–1224 (2012).
37. Nishimoto, S. *et al.* Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* **21**, 1641–1646 (2011).
38. Jenkinson, M. *Measuring Transformation Error by RMS Deviation*. (1999).
39. Aguirre, G. K., Zarahn, E. & D'Esposito, M. The variability of human, BOLD hemodynamic responses. *Neuroimage* **8**, 360–369 (1998).
40. Bollen, K. A. & Stine, R. Direct and Indirect Effects : Classical and Bootstrap

Estimates of Variability. *Sociol. Methodol.* 115–140 (1990).
41. Beyer, K., Beyer, K., Goldstein, J., Ramakrishnan, R. & Shaft, U. When Is 'Nearest Neighbor' Meaningful? *Int. Conf. Database Theory* 217–235 (1999). doi:10.1007/3-540-49257-7_15
42. Zimek, A., Schubert, E. & Kriegel, H.-P. A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data. *Stat. Anal. Data Min. ASA Data Sci. J.* **5**, 363–387 (2012).
43. Mikl, M. *et al.* Effects of spatial smoothing on fMRI group inferences. *Magn. Reson. Imaging* **26**, 490–503 (2007).
44. Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P. & Lancaster, J. A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). *Neuroimage* **2**, 89–101 (1995).
45. Poppenk, J. & Norman, K. A. Briefly Cuing Memories Leads to Suppression of Their Neural Representations. *J. Neurosci.* **34**, 8010–8020 (2014).
46. McIntosh, A. R. & Mišić, B. Multivariate Statistical Analyses for Neuroimaging Data. *Annu. Rev. Psychol.* **64**, 499–525 (2013).
47. Lieberman, M. D. & Cunningham, W. A. Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Soc. Cogn. Affect. Neurosci.* **4**, 423–428 (2009).
48. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
49. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
50. Cocosco, C. A., Kollokian, V., Kwan, R. K., Pike, G. B. & Evans, A. C. BrainWeb : Online Interface to a 3D MRI Simulated Brain Database. *Neuroimage* **5**, S425 (1997).

Acknowledgements

We thank Y. Lu for help with event, sub-event, and cut segmentations, S. Smith for manual hippocampal segmentations, and K. Norman, M. Sabbagh and G. Blohm for helpful comments. This research was funded by Natural Sciences & Engineering Research Council Discovery Grant 03637 (J.P.), which also supported J.T. Infrastructure funding was provided by the Canada Foundation for Innovation – John R. Evans Leaders Fund (J.P.). and a Queen's University Research Initiation Grant to J.P., who was supported by the Canada Research Chairs program. Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Data and code availability

Data are available from the Human Connectome Project at humanconnectome.org. Our analysis relied upon algorithms implemented in MATLAB (v.2017a), FreeSurfer (v.5.3), and FSL (v.5.0.10), and we have no new software to report.

Methods

Neuroanatomical and functional data were collected by the WU-Minn Human Connectome Project (HCP) consortium[8]. In a prior analysis, a 3T resting state fMRI (rs-fMRI) dataset of 1003 participants (age $M = 28.7$ years, $SD = 3.7$ years; 534 female) was used by the HCP group to generate spatial maps of typical brain networks that may be found in rs-fMRI participants through a process involving group-PCA and group-ICA[27,28]. As this dataset was the source of these maps, we refer to it as the *training* dataset. Our analyses involved applying these maps to a *testing* HCP dataset (7T) that contained both mv-fMRI and rs-fMRI scans, and that was the subject of all of our own analyses. Although this 7T dataset is described elsewhere[7–9,29], briefly, it consists of fMRI scans from 184 participants (age $M = 29.4$ years, $SD = 3.4$ years; 112 female). Each participant underwent four 15-minute mv-fMRI and four 15-minute rs-fMRI runs; functional images were acquired using a multiband gradient echo-planar imaging (EPI) pulse sequence (TR 1000 ms, TE 22.2 ms, flip angle 45°, multiband factor 5, whole-brain coverage 85 slices of 1.6 mm thickness, in-plane resolution 1.6 x 1.6 mm^2, FOV 208 x 208 mm^2)[30–33]. During each movie run, participants watched three or four movie clips interspersed with 20-second rest periods as well as an 84-second validation clip repeated at the end of each run (due to its repetition, we did not include this clip in our analyses). In addition, high-resolution T1-weighted and T2-weighted scans were gathered (TR 2400 ms and 3200 ms, TE 2.14 ms and 565 ms, flip angle 8° and variable, 0.7 mm thickness, in-plane resolution 0.7 x 0.7 mm^2, FOV 224 x 224 mm^2) for purposes of group anatomical alignment and creation of volume segmentations. Data collection was approved by the Washington University institutional review board[7], and performed by the HCP consortium, which also gathered informed consent from all participants at the time of data acquisition. Access to these datasets was granted by the HCP consortium, and acknowledged by the Health Sciences Research Ethics Board at Queen's University. No participants were excluded from analysis.

<u>Step 1: Timepoint discovery</u>

We mapped the 15 spatial maps resulting from the training set group-ICA decomposition onto each 7T participant's resting state and movie-viewing data using FSL's dual regression function[27,28], a method in which known spatial configurations are regressed against new data to transform 4D functional data into a set of timeseries (one per spatial map, Fig. S1a).

Although larger network set sizes were available from the HCP group (ranging from 15- to 300-brain-network solutions), we observed that set size had little material impact on trajectory estimates, and therefore selected the simplest available (15-network) solution. To increase signal-to-noise ratio, we temporally smoothed each resulting timeseries with a moving average filter (span = 5 s). This procedure yielded a smoothed timeseries for each brain network reflecting that network's activation over time.

We combined these timeseries to create separate (time x network) matrices for all four mv-fMRI and all four rs-fMRI runs for each participant (Fig. S1b). In preparation

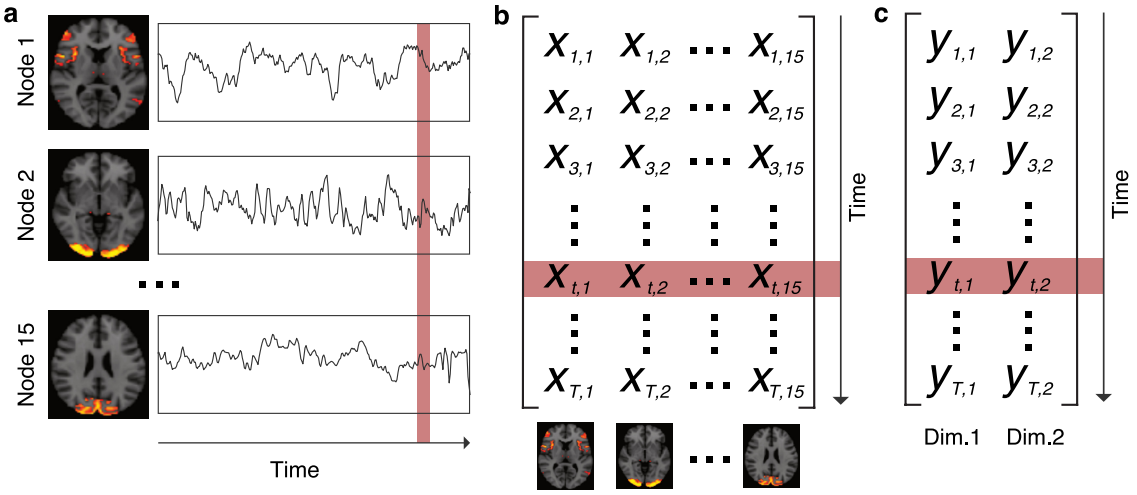


Fig. S1. Expression of brain network activity. (**a**) Each fMRI run of each participant was characterized as the activation of 15 brain networks over time. (**b**) Representation of A as a two-dimensional matrix of (time x network). (**c**) Reduced representation obtained after applying t-SNE algorithm.

for using the Mahalanobis distance metric on the data, we applied the t-distributed stochastic neighbour embedding (t-SNE) algorithm to reduce the dimensionality of each matrix from 15 dimensions to 2 dimensions at the default perplexity setting of 30 (Fig. 1c, Fig. S1c)[10]. We found the perplexity setting had little impact on our analysis, and therefore selected what is regarded as a moderate (and default) value.

Notably, Miller and colleagues[34] used a somewhat similar approach, creating a low-dimensional representation by first defining the space of possible meta-states as a discrete 5-dimensional state space, with each dimension representing a distinct group temporal ICA component derived from participants' functional data (i.e., connectivity patterns). Whole brain activity was expressed as a weighted combination of these components over time. To map each timepoint onto their meta-state space, they then discretized each weight at each timepoint according to its signed quartile. In contrast, our approach relies on dimensionality reduction algorithms to discover changes in meta-state directly from the continuous-valued 15-brain-network representation. We selected this approach because it affords flexibility in the designation of each meta-state by mapping each one onto a continuous two-dimensional space instead of a discrete 5-dimensional state space. Just as the Greek philosopher Heraclitus noted, "No man ever steps in the same river twice, for it's not the same river and he's not the same man", our approach is aligned to the very likely possibility that meta-states are continually evolving. Our approach also differs by drawing on published reference networks derived from a static group of 1003 participants (i.e., the training dataset described above[8,27,28]), rather than a set of networks derived from the specific dataset under interrogation.

To derive from our t-SNE representation a global measure sensitive to changes in network meta-state, we computed the Mahalanobis distance in position within this low-dimensionality t-SNE space across subsequent timepoints. The resulting *step distance vector* for each fMRI run of each participant should peak at points in the timeseries where shifts in network meta-state occur. To address potentially divergent results across repeated t-SNE algorithm runs, we repeated the dimensionality reduction and step distance vector creation process 100 times for each participant and each functional run. Then, we took the mean across the 100 step distance vectors for that run (Fig. 2a). However, even 95% confidence intervals were tightly constrained.

We applied a peak finding algorithm on each mean step distance vector to identify *transition* timepoints at which the step distance satisfied a minimum peak prominence threshold of 0.06, the value at which approximately 80% of all step distance values fell under the 5th percentile transition-associated step distance value. Setting a prominence value rather than applying a high pass filter allows the algorithm to consider step distances in the neighbourhood surrounding the peak being evaluated and results in more robust transition selection. To find *meta-stable* timepoints, we inverted the signal and specified a minimum peak width of 10; this parameter ensured that timepoints would only be identified within persistently meta-stable periods. One example of a participant's identified transitions and meta-stable timepoints are shown in Fig. 2a with green and black triangles, respectively. Under these parameters, the median step distance value associated with transition timepoints was 0.48 (5th and 95th percentile: [0.09, 1.79]). The median exceeds 94% of all values, while the lower bound exceeds

about 80% of all values (Fig. S2b). By contrast, the median step distance associated with meta-stable timepoints was 0.02 (5th and 95th percentile: [0.01, 0.03]), falling below about 94% of all values.

We examined the degree of contiguity in the t-SNE representations by comparing the standardized step distance distribution to an analogous distribution based on Gaussian noise. First, we generated simulated noise data, where each simulation consisted of 900 pairs of values taken at random from a normal distribution with the mean and standard deviation of a random participant's movie or rest t-SNE representation (Fig. S2d). We then calculated a step distance vector for this noise sample. The distribution of real data (described in Fig. S2b) was comprised of 1,311,791 values collapsed across participants and all available functional runs, where each run had 900 to 921 timepoints. Thus, we repeated the noise sampling procedure 1,460 times (the number of actual values divided by 900), generating a similar number of simulated values (Fig. S2e). To better compare the dispersion in the actual data against the dispersion in the noise samples, we standardized both distributions (Fig. S2c, f). Standardized step distance values for real data were predominantly dispersed closely around zero, whereas the noise standardized values were dispersed over a wider range. Also, the standardized step distance distribution for real data had a skewness of 5.0 and a kurtosis of 34.3, reflecting a long tail of extreme values, whereas the standardized noise distribution's skewness and kurtosis metrics indicated a fairly normal distribution (skewness of 0.6 and kurtosis of 3.2). Together, these distinctions help to formalize the temporal contiguity of real data that is visually apparent in Fig. S2a.

<u>Step 2: Psychological relevance</u>

To assess the relationship between our identified neural transitions to participants' experience, we investigated how transitions mapped onto well-characterized movie features (events, sub-events, cuts, semantic, visual, and auditory), as well as each participant's recorded head motion during the scan. Sub-events describe directly what an actor or multiple actors are doing or saying ("he is walking down the stairs", "they are riding in a car"), or they describe the motion of important objects ("a meteorite strikes the ground", "a car drives down the street"). We defined an event as a meaningful cluster of sub-events that describes a larger, overarching goal achieved by the sum of its parts (i.e., fine-grained and coarse-grained events[16]. Cuts referred to boundaries between two separate camera shots. Two raters used a video coding tool (Datavyu[35]) to independently identify boundaries demarcating events and sub-events, then met to discuss any differences in their ratings and achieve consensus event boundary timepoints. Using this approach, raters created consensus event and sub-event boundary segmentations for all four movie runs (14 clips). Only one rater logged cuts, as the objective nature of camera positions left little to discussion. Across the 14 clips found during the four movie runs, raters identified an average of 7.5 events, 37.3 sub-events, and 46.2 cuts per clip. A binarized timeseries was created for each logged feature, with onsets allocated to the nearest 1 s time bin (corresponding to the 1000 ms TR during which each fMRI volume was gathered).

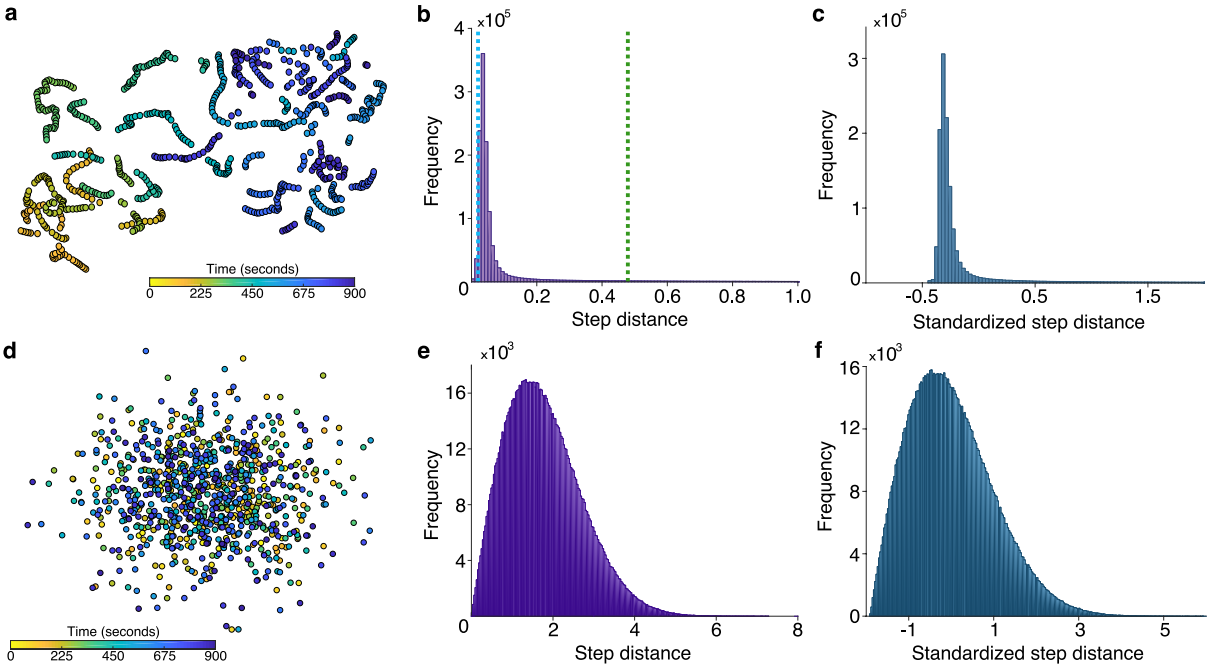


Fig. S2. Distribution of step distance values. (**a**) An example reduced t-SNE representation involving both segment types, as observed from a 15-minute rs-fMRI scan of one participant. (**b**) A histogram depicts the mean step distances within 7T participants' four rs-fMRI and four mv-fMRI runs. Approximately 94% of the observed values fell below the median *transition* distance (green dashed line), and 94% above the median *meta-stable* distance (blue dashed line). (**c**) A histogram depicts the standardized step distance values. Note that the values are mostly distributed close to zero, but feature a heavily skewed tail. (**d**) An example of noise simulated by sampling pairs from a normal distribution with characteristics mirroring actual reduced t-SNE representations. (**e**) A histogram analogous to the one in panel B that depicts the step distances across simulated noise samples. (**f**) A histogram depicts the standardized step distance values of noise samples. Note that relative to the analogous real data depicted in C, the simulated noise values are not as tightly distributed around zero, and the y-axis scale is two orders of magnitude less than the corresponding distribution drawn from real data in C.

The HCP group supplied two types of feature labels for the movie stimuli: semantic-category labels that described the high-level semantic features contained in each one-second epoch of the film[36], and motion-energy labels that described low-level structural features in the same epochs[37]. There were 859 semantic features and 2031 motion energy channels that expressed changes in the semantic content and motion-energy of each epoch, respectively. By summing across all semantic features and taking its derivative, we obtained a measure of overall magnitude of change in semantic content at each epoch. Similarly, we summed across all motion energy channels and took its derivative to obtain a measure of overall magnitude of change in perceptual features at each epoch. We also took the absolute value of the auditory amplitude vector derivative for each movie run as a measure of magnitude of change in volume. Finally, to rule out the possibility that transitions are a motion artifact, we obtained, for each participant, relative root-mean-square (RMS) change in head position[38], corresponding to a vector of head movement over time.

Whereas the feature vectors obtained above were correlated, we derived from them a set of independent feature vectors by "censoring" epochs where event and sub-event boundaries, as well as cuts were present. This was done by dropping values in a 3-second window around feature boundaries and ensured that apparent effects in lower-level features would not be explained by correlation to higher-level features. Event boundaries censored all other vectors; sub-event boundaries censored all vectors other than events; and cuts censored all vectors other than events and sub-events (Fig. 2e). To prevent spurious effects related to clip onset within movie runs, we also censored the first six seconds of each clip for both the uncensored and censored feature vectors.

For each transition and meta-stable point found in the mv-fMRI data, we next computed the average level of each feature accounting for hemodynamic response function (HRF) lag (working backwards based on the canonical HRF[39], sampling from a feature window 3-6 s prior to each transition and meta-stable point). We then averaged across all onsets of the same type for each participant. Thus, each participant ultimately had two values for each feature vector: one representing the average feature vector value at a transition timepoint, and an analogous value at a meta-stable timepoint.

We ran a t-test comparing these two values across participants for each feature, and used each resulting t-statistic to compute the proportion of variance that was explained in the feature vector (i.e., eta-squared) by the presence of a transition. Next, we used a non-parametric bootstrapping analysis to obtain a 95% CI for each feature[40]. Using participants' transition-baseline feature value difference as input data, this approach constructs a sampling distribution of the mean by resampling 1000 times with replacement across participants. Then, the t-statistics corresponding to the upper and lower bounds of each CI were converted to eta-squared values (Fig. 2d).

Having identified strong prediction of various features, we next sought to determine how crucial our specific embedding approach (i.e., t-SNE) was to identifying transitions that mapped strongly onto movie features. To this end, we derived three additional sets of transition and meta-stable timepoints: one from the unreduced (time x network) representations of each participant's mv-fMRI data (Fig. S1b), a second and

third set from a 2-dimensional representation of their mv-fMRI data obtained through principal components analysis (PCA) and independent components analysis (ICA). Specifically, we calculated the Mahalanobis distance across subsequent timepoints and applied the peak finding algorithm to identify transition timepoints; the minimum peak prominence threshold selection strategy and peak finding parameters for meta-stable timepoints remained unchanged. As above, we used these three additional sets of transition and meta-stable timepoints to obtain the proportion of variance that was explained in the uncensored feature vectors by the presence of a transition for each embedding approach (Fig. S3).

Transitions derived directly from the 15-dimensional (time x network) representations predicted a moderate proportion of variance in events, but not for lower-level event-based features such as sub-events or cuts. PCA transitions also performed worse than t-SNE transitions for all feature categories except auditory and motion. ICA transitions performed worst of all four embedding strategies. We interpret the strong performance of t-SNE in predicting features (especially semantic features) as reflecting its unique ability to distill important local and global structure from the data. Consistent with this idea, as dimensionality increases, distance metrics are understood to lose their usefulness as the distances to the nearest and furthest point from any reference point approach equality[41]. Zimek and colleagues further showed that if the dimensions are correlated rather than independent and identically distributed, then considering subsets of dimensions can improve the performance of distance metrics such as Euclidean distance[42]. This likely explains the poor unreduced feature prediction in our analysis, and supports the usage of a dimensionality reduction algorithm, as the spatial networks in the 15-dimensional (time x network) representations that we submitted to analysis were not independent from each other. Substituting Mahalanobis for Euclidean distance can additionally account for residual correlations between reduced dimensions[12]. In spite of this, in our analysis, poor PCA performance signalled that the *kind* of dimensionality reduction approach also matters. PCA is a linear method that seeks to preserve global structure, whereas t-SNE can balance a trade-off between local and global structure. Upon exploring the variable loadings for PCA results, we observed that the two principal components were weighted heavily towards visual networks that likely explained substantial global variance, but at the cost of sensitivity to changes in networks related to higher-level conceptual processing. As a result, PCA-based transitions performed comparably to unreduced transitions in the visual category and correlated categories (e.g., auditory), but worse in semantic categories.

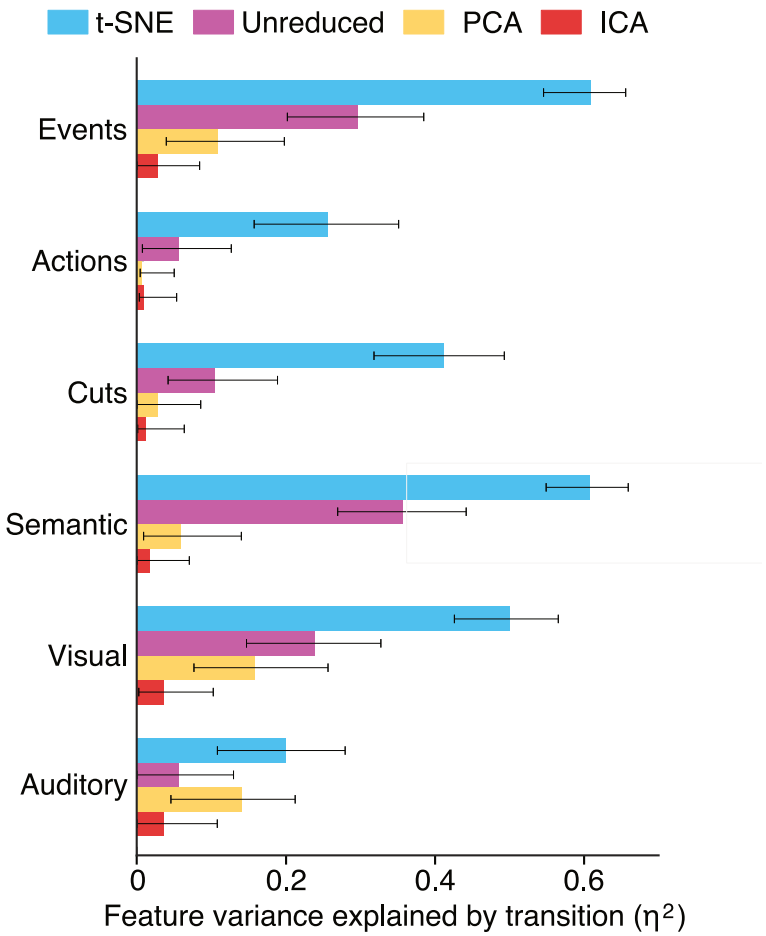


Fig. S3. Impact of embedding approach on transitions' predictive power for movie features. Eta-squared values describing the proportion of variance in uncensored movie features explained by alignment to transition vs. meta-stable timepoints. Colours indicate the state space from which Mahalanobis distances were calculated to identify transition and meta-stable timepoints.

Visual inspection revealed considerable structure in transitions across participants during movie-viewing (Fig. 2b), complementing prior findings of local coordination of brain activity as participants watch well-made films. To formalize this observation, we tested for higher group alignment in movie runs than rest runs. For each participant and each run, we took the Fisher transformation of the correlation between the log of their step distance vector and the log of the median group signal (excluding the step distance vector of the participant in question). To avoid potential effects resulting from the onset and offset of the run, we excluded the first and last five epochs of the step distance vectors. Then, we compared all group alignment values from the four mv-fMRI runs against zero. The same bootstrapping analysis was carried out for all group alignment values for the four rs-fMRI runs.

To provide an alternate way of testing the influence of narrative events over transitions, we correlated the number of events in each movie clip with the degree of group alignment of transitions for that clip. For each clip, we divided the number of events in that clip by its duration in minutes. For each participant, we obtained their clip conformity by taking the Fisher transformation of the correlation between the log of their step distance vector and the log of the median group signal (excluding the step distance vector of the participant in question). This was repeated for each clip within a movie run rather than the movie run in its entirety; thus, each participant had a vector describing their clip conformity, wherein each element corresponded to a conformity value for a particular clip. We calculated the Pearson correlation between the events per minute and conformity vector within each participant, resulting in a distribution of 184 correlation coefficients. We fed this distribution into a bootstrapping analysis (as described previously when we used this strategy to obtain 95% confidence intervals) with 1000 samples to test whether the correlation between events per minute and group alignment was significantly different from 0.

Step 3: Linking movie and rest

Using the transition and baseline onsets, we next performed a voxel-wise conjunction analysis in which we sought to identify stable spatial correlates of network meta-state transitions that could be found across rest and movie runs. Through this analysis, we wished to learn whether any set of transition predictors could link brain activity for which we have insight into psychological relevance (mv-fMRI) with brain activity for which we do not (rs-fMRI). To this end, we sampled the average fMRI image at transition and baseline onsets for each participant in the same manner as with feature vectors above, but without correcting for HRF lag (since the predictor and dependent variables were affected by the same delay). In this case, however, we created participant-wise transition and meta-stable timepoint averages not only for mv-fMRI runs, but also (separately) for rs-fMRI runs. We also spatially smoothed each image using a 6 mm FWHM gaussian kernel (at the lower bound of optimal parameters for overcoming inter-subject variability[43]).

We first subtracted meta-stable from transition images independently for movie and rest, such that a movie difference image and rest difference image was available for each participant. We masked these images using a grey matter mask, performing

comparisons only on those voxels with at least a 50% probability of being grey matter based on an MNI anatomical atlas in the same space[44]. Again, we used a non-parametric bootstrapping analysis, but this time obtained a bootstrap ratio *image* for each of the movie and rest difference images (see e.g., [45]). As the bootstrap ratio approximates a z-distribution[46], we used a cumulative distribution function to convert it into a map of voxel-wise *p*-statistics. To perform conjunction analysis, we then thresholded both the movie and rest *p*-maps at a $P < 0.05$, setting all voxels above this value to infinity, and computed the product of the *p*-maps. The resulting conjunction *p*-map represented the probability of obtaining a supra-threshold result not only in map A, but also in map B. Because of the initial thresholding step, it had an implied *p*-value threshold of 0.0025. We then further suppressed supra-threshold voxels within the conjunction *p*-map that did not satisfy a minimum cluster extent threshold of 150 voxels (614.4 mm^3). We selected these voxel-wise and extent thresholds to achieve a balance between Type I and Type II error rate[47].

To ascertain whether the frequency of neural transitions was stable within individuals, we also derived a *transition rate* (per minute) for each participant and run by dividing the total number of transitions in each run by the total scan time in minutes. As all four rs-fMRI scans were collected during different sessions, we were able to quantify the reliability of our transition rate metric within participant and across scanning sessions. To do so, we computed the correlation between participant transition rates derived from each pair of rs-fMRI scans. After repeating this for each pair of scans, we computed the average correlation coefficient obtained from all pairs. For each participant, we then computed the within-participant average transition rate for rs-fMRI runs to obtain a more stable trait estimate of each participant's intrinsic transition rate. We used these values to predict their average mv-fMRI transition rate for corresponding participants using the bootstrap correlation approach described earlier.

Step 4: Assessing hippocampal predictors

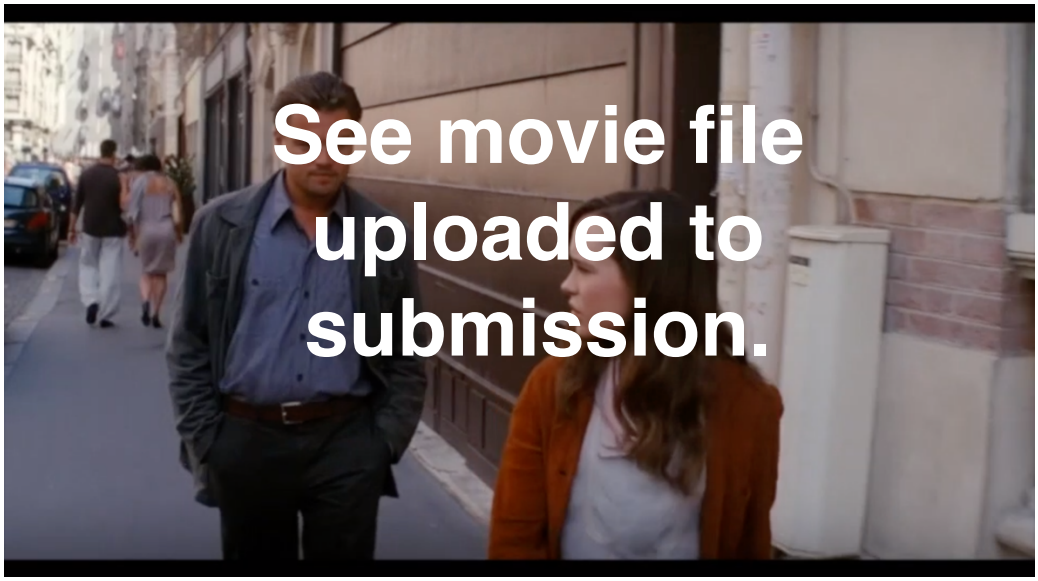
We next tested whether grand mean (rs-fMRI and mv-fMRI) transition rates were related to hippocampal volume configurations as a proxy for memory type. Specifically, we tested anterior, posterior, overall, and the ratio of posterior to anterior hippocampal volume for the left and right hippocampus separately using a bootstrap correlation analysis (as above). To obtain volumes of these structures, we submitted high-resolution T1-weighted MRI scans to FreeSurfer (v.5.3) for automated analysis[48,49]. An expert rater then conducted a manual segmentation of the structure along its longitudinal axis, using the uncal apex as a landmark[19]. For validation purposes, a second rater scored a subset of 40 hippocampi. Agreement concerning the exact coronal slice containing the uncal apex was high (Dice coefficient = 0.85).

Extended Data

Table S1. Voxel clusters showing reliable activity differences at network transition vs. meta-stable timepoints in both movie-viewing and resting fMRI runs.

Region	BA	Hemi.	Peak MNI coordinates			Peak *z* estimate	Spatial extent (mm^3)
			X	Y	Z		
Transition > Meta-stable							
Frontal lobe							
Dorsal anterior cingulate ctx.	32	L/R	-3	13	37	6.59	2232
Ventrolateral prefrontal ctx.	6	R	56	0	11	5.06	819
Insular Lobe							
Anterior insula	13	R	43	11	-2	6.18	1282
Posterior insula	13	R	45	-11	-6	5.53	754
Parietal lobe							
Posterior cingulate ctx.	23	L/R	-5	-38	24	7.69	1962
Precuneus	7	L/R	-6	-43	58	6.24	2597
Visual lobe							
Visual assoc. ctx.	18	L/R	2	-70	21	7.15	6230
Meta-stable > Transition							
Frontal lobe							
Angular g.	39	R	53	-61	27	-5.37	819
Angular g.	39	L	-40	-67	48	5.29	1442

Note: Peak coordinates are displayed in MNI space[50].



Movie S1. Movie-based elicitation of meta-state transitions. A bell sound overlaid on the movie audio denotes times at which transitions were frequently elicited (with correction for the hemodynamic response function). These times were derived from peaks in the group median step distance vector, which arose when many participants experienced transitions at the same time. The clips highlight specific alignment of meta-state transitions to semantic change in the film stimulus.