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 identify transitions in brain network meta-state, we characterize a neurocognitive landmark of conceptual shifts. Drawing on a large fMRI dataset, we found meta-state transitions robustly aligned to new concepts during movie-viewing, with the highest organization when movie stimuli had more numerous conceptual shifts. Meta-state transitions featured common characteristics across movie and rest, including a spatial pattern that involved brain regions associated with spontaneous thought, as well as a stable transition rate. Based on our findings, we suggest meta-state transitions could serve as an objective biological landmark for implicit demarcation of new thoughts. As larger anterior hippocampi predicted lower transition rates – possibly reflecting more dwell time in memories following hippocampal recall – our findings also 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 conceptual contents (with some definitions emphasizing spontaneous production, *5*), we propose that implicit measurement of conceptual shifts offers an interesting alternative to meta-cognition. Researchers can readily decode neurally discernable concepts (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 concept to another. One possible reason this strategy has not been employed is that it is challenging to reliably distinguish large numbers of concepts. However, recent perspectives emphasize measurement of thought dynamics instead of contents (*1*). What if, rather than tracking the rise and fall of particular concepts, we found a way to track conceptual shifts themselves?

Here we describe a neurocognitive landmark that tracks such conceptual shifts. In particular, we report a new method to identify breaks between stable periods of brain network configuration, and present a psychological validation based on the correspondence of these network events with conceptual shifts in films. We further assess the neural landmark's reliability and correspondence with known regional markers of spontaneous thought, and explore its generalization to task-free brain data in which thought is unconstrained. Finally, we 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*, *8*). We converted each fMRI run into the expression of 15 known brain networks over time (see Fig. S1 in *9*), then reduced its dimensionality from (*15 x time)* to (*2 x time)* using t-SNE (*10*). 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 concept 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 *9*).

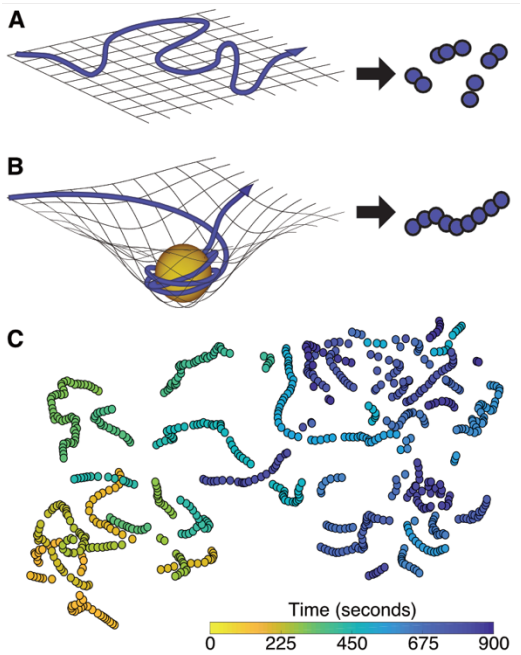

Fig. 1. Network space representation. (**A**) Continuous, random passage through the space of possible network configurations generates fragments in t-SNE space, as opposed to (**B**) contiguous, worm-like segments when an attractor holds network configurations in relative meta-stability. (**C**) An 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 Euclidean distance 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 steps 100 times for each participant and each functional run. Peaks within the resulting *median* 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 median 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 conceptual or perceptual features being viewed by participants. 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. 2 B,C). To quantify this, we obtained each participant's group alignment for each resting state and movie-viewing run, which describes the similarity between the individual's step distance vector and the corresponding median group step distance vector (*9*). By averaging group alignment across participants, then across runs of the same type, we found higher group alignment for movie runs, mean $r = 0.30$, 95% CI: [0.29, 0.31], than for rest, mean $r = 0.06$, 95% CI: [0.05, 0.07], 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 (*11*, *12*; see also *13*), and provides a preliminary link between transitions and naturalistic cognition.

As an alternate means of evaluating the influence of conceptual information over transitions, we attempted to predict transition alignment based on the amount of conceptual structure in each clip. Two expert raters came to a consensus on boundaries between events (i.e., introduction of new concepts), and 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 $r = 0.27$, 95% CI: [0.24, 0.31].

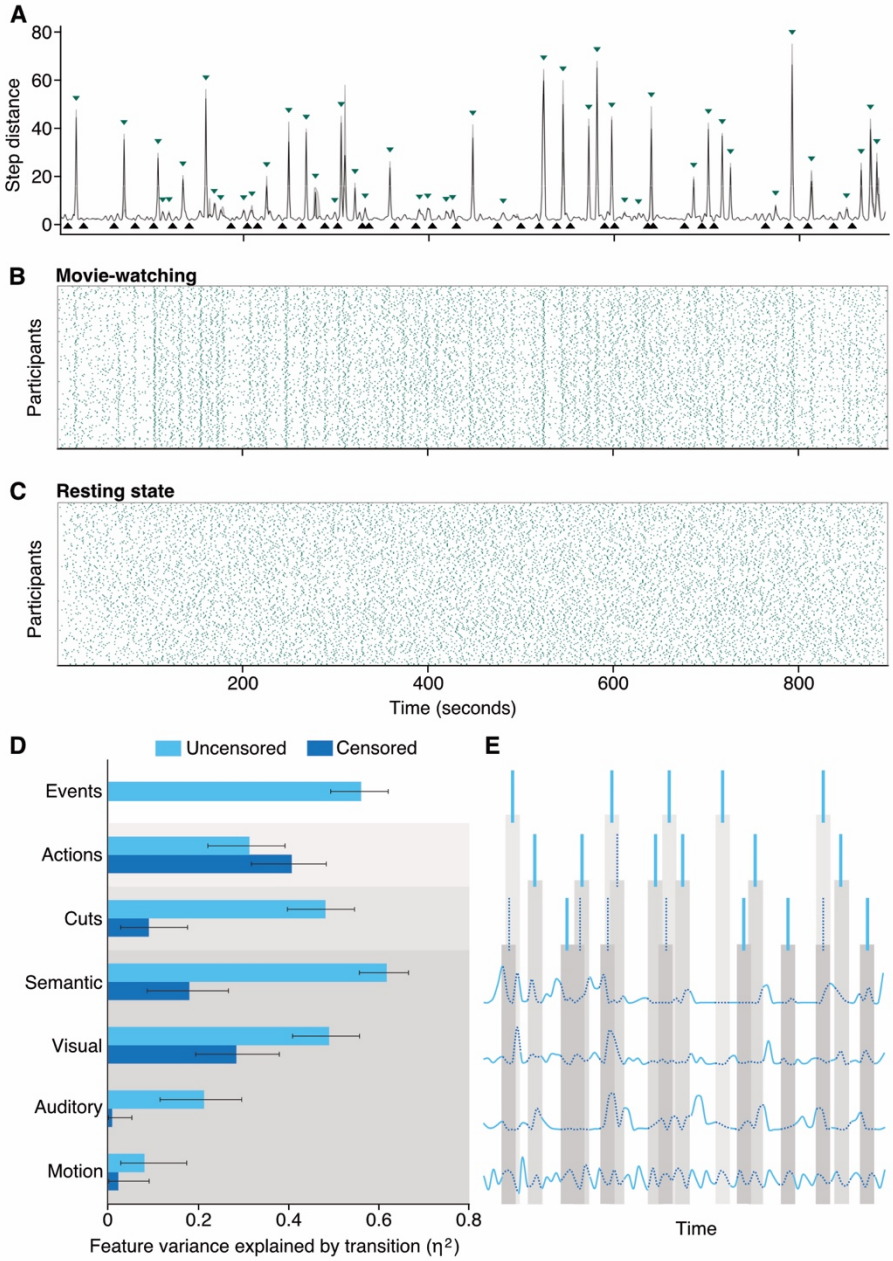

Fig. 2. Explaining network meta-state transitions using movie features. (**A**) A participant's median step distance vector during one mv-fMRI run, with 95% CI ribbon (narrow ribbon indicates stability over t-SNE iterations). *Transition* timepoints (green triangles) and *meta-stable* timepoints (black 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 B 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. (**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 conceptual and other features within-participant. Consensus labels of 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. 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 lower-level features appear to predict transitions, 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 were most successful at predicting the onset of new concepts: they accounted for 56.0% of variance in events and 61.7% of variance in semantic change. However, only 17.9% of semantic change variance was accounted for by transitions after censorship. In fact, eta-squared values for all lower-level movie features decreased markedly after removing timepoints that could be confounded with event boundaries (Fig. 2D). Thus, transitions (and specifically t-SNE-derived transitions; Fig. S3) were most prominently markers of new concepts (see Movie S1 for illustration of this specificity). These results complement those of Baldassano and colleagues (*13*), 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.

To help generalize our psychological validation to resting cognition, where conceptual shifts might be observed outside of stimulus influence, we sought to establish a link between transitions 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. 3 A-C, Table S1).

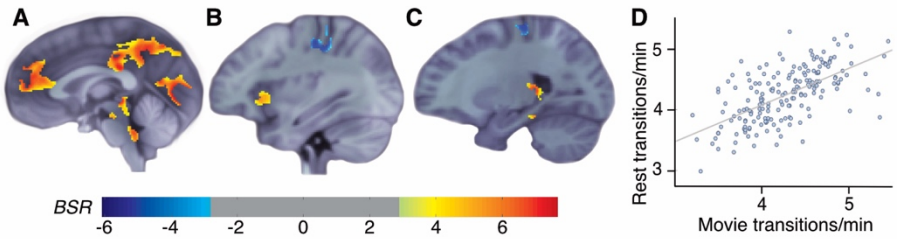

Fig. 3. Transition features generalize from stimulus-driven to resting cognition. (**A**-**C**) Conjunction analysis revealed brain activity that consistently differentiated transition from meta-stable timepoints, regardless of task. (**A**) Anterior cingulate, posterior cingulate, and precuneus implicated in transitions; lateral parietal cortex in meta-stability. (**B**) Anterior insula implicated in transitions. (**C**) Posterior hippocampus and parahippocampal cortex implicated in transitions. (**D**) Participants' movie-viewing transition rate predicts resting transition rate.

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 $r = 0.43$), 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 significantly predicted transition rate during movie-viewing, $r(181) = 0.52$, 95% CI: [0.41, 0.63] (Fig. 3D), thereby further linking its properties across stimulus-driven and resting cognition.

Finally, we investigated the hippocampus as a predictor of transition rate, based on its appearance in our conjunction analysis (Fig. 3C) as well as past reports that the hippocampus is sensitive to movie onset/offset (*14*) and conceptual shifts (*13*). Our group has previously proposed that, whereas posterior hippocampus (pHPC) helps to distinguish similar events in memory (*15*), anterior hippocampus (aHPC) supports gist-like representations (*16*). Accordingly, we interpret pHPC activation alongside transitions as reflecting a role in separating two events in time. We further hypothesized that because aHPC is associated with greater pattern completion (*17*) because of a relatively small dentate gyrus (*18*), 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 right anterior hippocampus and smaller pHPC:aHPC ratio predicted a lower transition rate (Fig. 4).

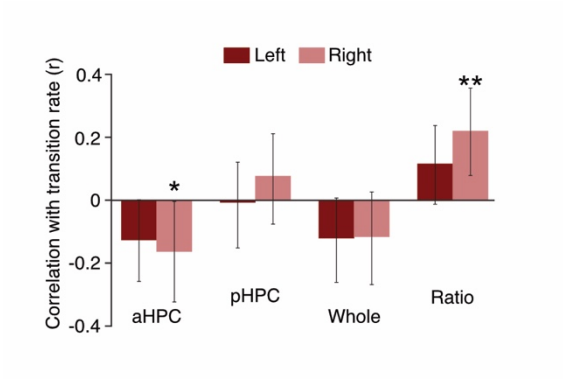

Fig. 4. Grounding of transitions in neurobiology of event cognition. Prediction of transition rate by volumes of anterior hippocampus (aHPC), posterior hippocampus (pHPC), whole hippocampus, and the pHPC to aHPC volume ratio using transition rates derived from movie-viewing and rest data. * designates 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 new concepts in movie stimuli, and also found the resting-state transition rate to be a reliable, trait-like characteristic across imaging sessions. Characteristics of movie-viewing transitions were shared by resting-state transitions, supporting a unified interpretation of transitions across task.

A thought is grounded in a concept. We found transitions not only predicted conceptual shifts, but also activity in 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 (*19*); the posterior cingulate cortex is a key region in the default mode network that is more active during task-unrelated than task-related thought (*20*, *21*); and parahippocampal cortex and the hippocampus have been associated with unaware (and therefore unconstrained) cognitions during task-unrelated thought (*22*). Taken together, our findings therefore suggest transitions may serve as an effective marker of new thoughts. Moreover, through negative correlation of transition rate with anterior hippocampus volume – perhaps a result of increased dwell time within concepts contained by more readily accessible hippocampal memories – our findings illustrate how this marker could also contribute to a better understanding of thought dynamics.

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. Analysis of transitions may offer an empirical solution to meta-cognitive challenges associated with demarcating thoughts: what if we were to measure them using neurobiology, whether by reconfiguration of brain networks, activation of particular nodes, or other observable phenomena? This strategy sparks questions previously inaccessible without introspection: for example, how many thoughts do we experience per waking day? Extrapolating from a median transition rate across movie-viewing and rest of 4.3 transitions / min, and a recommended sleep time of 8 hours, one could tentatively estimate just over four thousand daily thoughts for healthy adults of a 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, the availability of even a tentative answer highlights how such an approach could be fruitful in refining how we think about thought.

1. K. Christoff, Z. C. Irving, K. C. R. Fox, R. N. Spreng, J. R. Andrews-Hanna, Mind-wandering as spontaneous thought: a dynamic framework. *Nat. Rev. Neurosci.* **17**, 718–731 (2016).
2. R. E. Nisbett, T. D. Wilson, Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* **84**, 231–259 (1977).
3. P. Seli, R. E. Beaty, J. A. Cheyne, D. Smilek, J. Oakman, D. L. Schacter, How pervasive is mind wandering, really? *Conscious. Cogn.* **66**, 74–78 (2018).
4. W. James, *The principles of psychology* (H. Holt and Company, New York, 1890), vol. 20.
5. J. Smallwood, Distinguishing how from why the mind wanders: a process-occurrence framework for self-generated mental activity. *Psychol. Bull.* **139**, 519–535 (2013).
6. K. A. Norman, S. M. Polyn, G. J. Detre, J. V. Haxby, Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* **10**, 424–430 (2006).
7. D. C. Van Essen et al., The Human Connectome Project: A data acquisition perspective. *Neuroimage*. **62**, 2222–2231 (2012).
8. D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, K. Ugurbil, WU-Minn Consortium, The WU-Minn Human Connectome Project: an overview. *Neuroimage*. **80**, 62–79 (2013).
9. Materials and methods are available as supplementary materials.
10. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
11. U. Hasson, O. Landesman, B. Knappmeyer, I. Vallines, N. Rubin, D. J. Heeger, Neurocinematics: The Neuroscience of Film. *Projections*. **2**, 1–26 (2008).
12. J. Chen, Y. C. Leong, C. J. Honey, C. H. Yong, K. A. Norman, U. Hasson, B. Sciences, Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).
13. C. Baldassano, J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, K. A. Norman, Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*. **95**, 709–721 (2017).
14. A. Ben-Yakov, Y. Dudai, Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. *J. Neurosci.* **31**, 9032–9042 (2011).
15. J. Poppenk, M. Moscovitch, A hippocampal marker of recollection memory ability among healthy young adults: Contributions of posterior and anterior segments. *Neuron*. **72**, 931–937 (2011).
16. J. Poppenk, H. R. Evensmoen, M. Moscovitch, L. Nadel, Long-axis specialization of the human hippocampus. *Trends Cogn. Sci.* **17**, 230–240 (2013).
17. M. A. Yassa, C. E. L. Stark, Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–525 (2011).
18. N. V. Malykhin, R. M. Lebel, N. J. Coupland, A. H. Wilman, R. Carter, In vivo quantification of hippocampal subfields using 4.7 T fast spin echo imaging. *Neuroimage*. **49**, 1224–1230 (2010).
19. W. W. Seeley, V. Menon, A. F. Schatzberg, J. Keller, G. H. Glover, H. Kenna, A.

L. Reiss, M. D. Greicius, Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. **27**, 2349–2356 (2007).
20. R. Leech, D. J. Sharp, The role of the posterior cingulate cortex in cognition and disease, 12–32 (2014).
21. J. Smallwood, T. Karapanagiotidis, F. Ruby, B. Medea, I. De Caso, M. Konishi, H. T. Wang, G. Hallam, D. S. Margulies, E. Jefferies, 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).
22. K. Christoff, A. M. Gordon, J. Smallwood, R. Smith, J. W. Schooler, Experience sampling during fMRI reveals default network and executive system contributions to mind wandering (2009).
23. L. D. Nickerson, S. M. Smith, D. Öngür, C. F. Beckmann, Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Front. Neurosci.* **11**, 1–18 (2017).
24. C. Beckmann, C. Mackay, N. Filippini, S. Smith, Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *Neuroimage*. **47**, S148 (2009).
25. M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, M. Jenkinson, The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*. **80**, 105–124 (2013).
26. S. Moeller, E. Yacoub, C. a Olman, E. Auerbach, J. Strupp, N. Harel, K. Uğurbil, 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).
27. D. A. Feinberg, S. Moeller, S. M. Smith, E. Auerbach, S. Ramanna, M. Gunther, M. F. Glasser, K. L. Miller, K. Ugurbil, E. Yacoub, Multiplexed Echo Planar Imaging for Sub-Second Whole Brain FMRI and Fast Diffusion Imaging. *PLoS One*. **5**, e15710 (2010).
28. K. Setsompop, B. A. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, L. L. Wald, 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).
29. J. Xu, S. Moeller, J. Strupp, E. J. Auerbach, L. Chen, D. A. Feinberg, K. Ugurbil, E. Yacoub, in *Proceedings of the 20th Annual Meeting of ISMRM* (2012), vol. 2306.
30. R. L. Miller, M. Yaesoubi, J. A. Turner, D. Mathalon, A. Preda, G. Pearlson, T. Adali, V. D. Calhoun, Higher Dimensional Meta-State Analysis Reveals Reduced Resting fMRI Connectivity Dynamism in Schizophrenia Patients. *PLoS One*. **11**, e0149849 (2016).
31. J. M. Zacks, K. M. Swallow, Event Segmentation. **16**, 80–84 (2007).
32. D. Team, Datavyu: A Video Coding Tool (2014), (available at http://datavyu.org).
33. A. G. Huth, S. Nishimoto, A. T. Vu, J. L. Gallant, A continuous semantic space describes the representation of thousands of object and action categories across

the human brain. *Neuron*. **14**, 1210–1224 (2012).
34. S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, J. L. Gallant, Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* **21**, 1641–1646 (2011).
35. M. Jenkinson, “Measuring Transformation Error by RMS Deviation” (Oxford, 1999), (available at https://www.fmrib.ox.ac.uk/datasets/techrep/tr99mj1/tr99mj1.pdf).
36. G. K. Aguirre, E. Zarahn, M. D’Esposito, The variability of human, BOLD hemodynamic responses. *Neuroimage*. **8**, 360–369 (1998).
37. K. A. Bollen, R. Stine, Direct and Indirect Effects : Classical and Bootstrap Estimates of Variability. *Sociol. Methodol.*, 115–140 (1990).
38. K. Beyer, K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, When Is “Nearest Neighbor” Meaningful? *Int. Conf. Database Theory*, 217–235 (1999).
39. A. Zimek, E. Schubert, H.-P. Kriegel, A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data. *Stat. Anal. Data Min. ASA Data Sci. J.* **5**, 363–387 (2012).
40. M. Mikl, P. Hluštík, A. Drastich, R. Mareček, M. Brázdil, P. Krupa, P. Chlebus, M. Pavlicová, Effects of spatial smoothing on fMRI group inferences. *Magn. Reson. Imaging*. **26**, 490–503 (2007).
41. J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, J. Lancaster, 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).
42. J. Poppenk, K. A. Norman, Briefly Cuing Memories Leads to Suppression of Their Neural Representations. *J. Neurosci.* **34**, 8010–8020 (2014).
43. A. R. McIntosh, B. Mišić, Multivariate Statistical Analyses for Neuroimaging Data. *Annu. Rev. Psychol.* **64**, 499–525 (2013).
44. M. D. Lieberman, W. A. Cunningham, Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Soc. Cogn. Affect. Neurosci.* **4**, 423–428 (2009).
45. B. Fischl, FreeSurfer. *Neuroimage*. **62**, 774–781 (2012).
46. B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, A. M. Dale, Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. **33**, 341–355 (2002).
47. C. A. Cocosco, V. Kollokian, R. K. Kwan, G. B. Pike, A. C. Evans, BrainWeb : Online Interface to a 3D MRI Simulated Brain Database. *Neuroimage*. **5**, S425 (1997).

Acknowledgements

We thank Y. Lu for help with event, action, and cut segmentations, S. Smith for manual hippocampal segmentations, and K. Norman and M. Sabbagh 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.

Supplemental 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 820 participants 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 (*23*, *24*). 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*, *8*, *25*), briefly, it consists of fMRI scans from 184 participants (age $M = 29.4$ years, $SD = 3.4$ years). 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) (*26*–*29*). 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.

<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 (*23*, *24*), 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 for using the Euclidean 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.

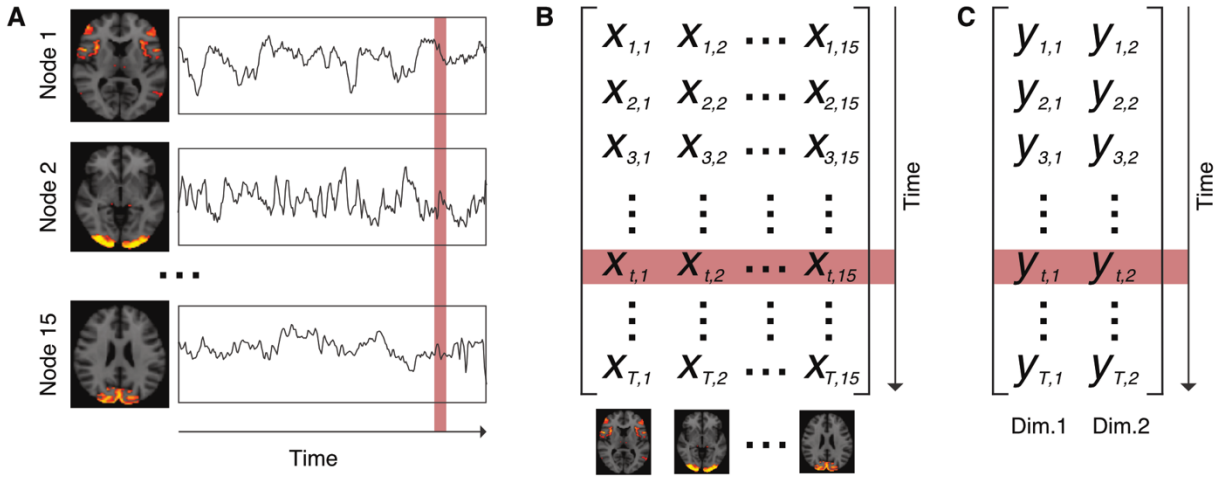

Fig. S1. Expression of brain network activity. (**A**) Each fMRI run of each participant was characterized as the activation of 15 brain networks over time. (**B**) Representation of A as a two-dimensional matrix of (time x network). (**C**) Reduced representation obtained after applying t-SNE algorithm.

Notably, Miller and colleagues (*30*) 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 820 participants (i.e*.,* the training dataset described above; *8, 23, 24*), 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 Euclidean 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. We normalized step distance vectors for each run by dividing it by the median standard deviation of all step distance values from that run, thereby expressing the shift in network meta-state at each point in time as a function of its uniqueness compared to the whole timeseries. To address potentially divergent results across repeated t-SNE algorithm runs, we repeated the dimensionality reduction and step distance vector creation steps 100 times for each participant and each functional run. Then, we took the median across the 100 normalized step distance vectors for that run (Fig. 2A). However, even 95% confidence intervals were tightly constrained.

We applied a peak finding algorithm on each median normalized step distance vector to identify *transition* timepoints at which the step distance satisfied a minimum peak prominence threshold of 3 and a minimum peak spacing of 4 s. 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 normalized step distance associated with transition timepoints was 23.4, and the lower bound was 5.0 (95% CI: [5.0, 79.2]). The median exceeds 95% of all values, while the lower bound exceeds about 80% of all values (Fig. S2B). By contrast, the median

normalized step distance associated with meta-stable timepoints was 1.5 (95% CI: [0.7, 2.1]), falling below about 88% 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 and normalized a step distance vector for this noise sample. The distribution of real data (described in S2B) was comprised of 1,311,785 values collapsed across participants and all available functional runs, where each run had 900 to 921 timepoints. Thus, we repeated the noise sampling procedure 1,458 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. S2 C,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 5.5 and a kurtosis of 47.5, reflecting a long tail of extreme values, whereas the standardized noise distribution's skewness and kurtosis metrics indicated a fairly normal distribution (skewness of 0.7 and kurtosis of 3.5). 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, actions, cuts, semantic, visual, and auditory), as well as each participant's recorded head motion during the scan. Actions 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 up the wall"). We defined an event as a meaningful cluster of actions that describes a larger, overarching goal achieved by the sum of its parts (i.e., fine-grained and coarse-grained events, *31*). Cuts referred to boundaries between two separate camera shots. Two raters used a video coding tool (Datavyu; *32*) to independently identify boundaries demarcating events and actions, then met to discuss any differences in their ratings and achieve consensus event boundary timepoints. Using this approach, raters created consensus event and action 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 actions, 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).

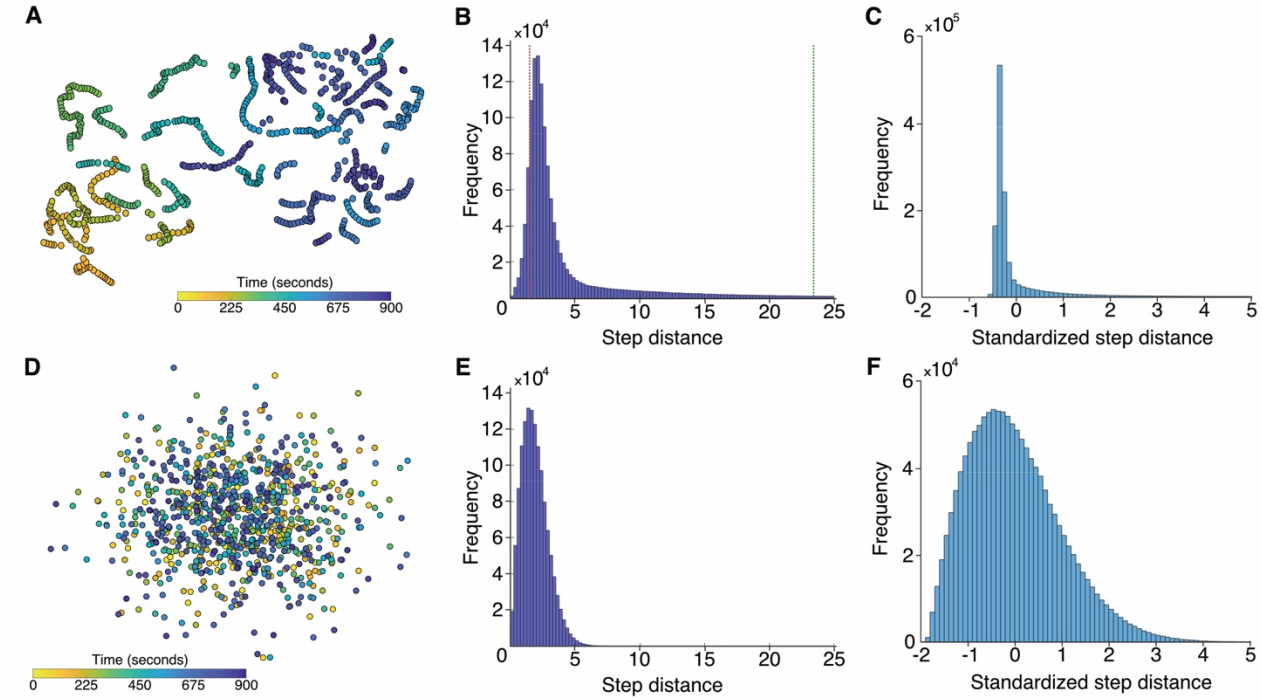

Fig. S2. Median normalized step distance values. (**A**) An example reduced t-SNE representation involving both segment types, as observed from a 15-minute rs-fMRI scan of one participant. (**B**) A histogram depicts the median normalized step distances within 7T participants' four rs-fMRI and four mv-fMRI runs. Approximately 95% of the observed values fell below the median *transition* distance (green dashed line), and 88% above the median *meta-stable* distance (red 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 normalized 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 an order 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 (*33*), and motion-energy labels that described low-level structural features in the same epochs (*34*). 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 high-level 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 (*35*), 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 action boundaries, as well as cuts were present. This was done by dropping values in a 3-second window around feature 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; action boundaries censored all vectors other than events; and cuts censored all vectors other than events and actions (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, *36*, 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 (*37*). 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 two 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), and the other

from a 2-dimensional representation of their mv-fMRI data obtained through principal components analysis (PCA). Specifically, we calculated the Euclidean distance across subsequent timepoints and divided these step distance vectors by their median standard deviation. Then, we applied the peak finding algorithm to identify transition timepoints; the minimum peak prominence was set to the value at which 80% of all step distance values fell under it, and the minimum peak spacing parameter remained unchanged, as did peak finding parameters for meta-stable timepoints. As above, we used these two 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 actions or cuts. PCA transitions also performed worse than t-SNE transitions for all feature categories except auditory and motion. 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 (*38*). 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 (*39*). 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. 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.

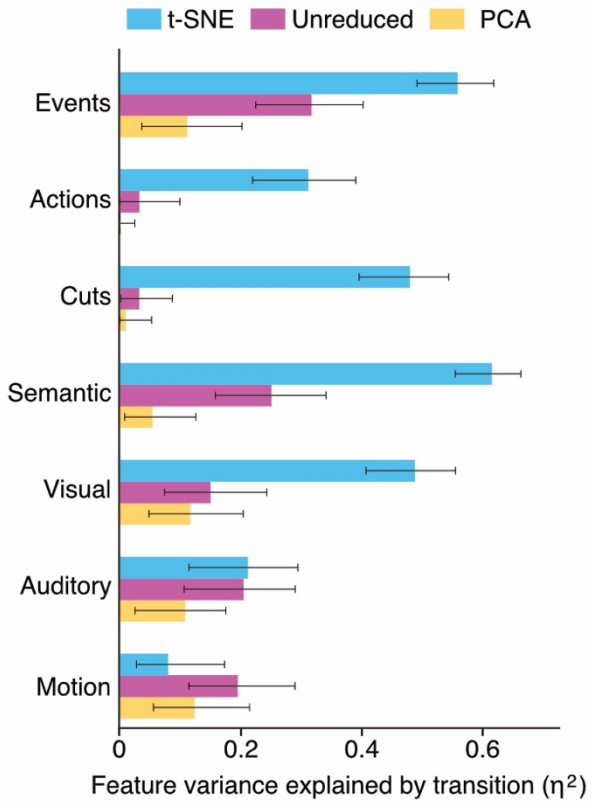

Fig. S3. Impact of embedding approach on transitions' predictive power for movie features. Eta-squared values describing the proportion of variance in uncensored movie features explained by alignment to transition vs. meta-stable timepoints. Colours indicate the state space from which Euclidean distances 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 calculated 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 conceptual content over transitions, we correlated the number of events in each movie clip with the degree of transition alignment across participants 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 group alignment by calculating 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 group alignment, within which each element corresponded to an alignment value for a particular clip. We calculated the Pearson correlation between the events per minute and group alignment 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; *40*).

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 (see *41*). 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., *42*). As the bootstrap ratio approximates a z-distribution (*43*), 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 (*44*).

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, *45*, *46*) for automated analysis, An expert rater then conducted a manual segmentation of the structure along its longitudinal axis, using the uncal apex as a landmark (*16*). 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).

Supplemental Results

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							
Anterior cingulate ctx.	32/24	L/R	0	50	10	6.56	5630
Parietal lobe							
Posterior cingulate ctx.	23/31	L/R	0	-29	27	6.88	1809
Precuneus	7	L/R	0	-53	45	6.43	2244
Occipital lobe							
Visual assoc. ctx	18	L/R	2	-77	0	7.69	4320
Subcortical							
Anterior insular ctx.	13	R	29	13	-2	6.21	918
Caudate n.	48	R	10	8	18	6.73	753
Caudate n.	48	L	-13	5	16	6.20	932
Parahippocampus, Brainstem (sup. colliculus)	-	L/R	-8	-32	-5	7.04	3311
Posterior hippocampus, Thalamus,	-	L	-16	-34	11	6.46	594
Meta-stable > Transition							
Frontal lobe							
Precentral g.	4	L	-27	-26	64	-5.64	1580
Parietal lobe							
Supramarginal g.	40	R	43	-30	45	-6.08	2541

Note: Peak coordinates are displayed in MNI space (*47*).

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 conceptual shifts in the film stimulus.