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 transitions. Drawing on a large fMRI dataset, we found meta-state transitions robustly predicted the onset of new concepts during movie-viewing, and were more organized under the control of movie stimuli with more conceptual transitions. Meta-state transitions held common characteristics across movie-viewing and rest data, including a spatial pattern that involved brain regions associated with mind-wandering, as well as a stable transition rate. We also found meta-state transitions coincided with posterior hippocampus activation, whereas a lower transition rate was predicted by anterior hippocampus volume, consistent with a mnemonic segmentation role for the posterior and mnemonic attractor role for the anterior. On the basis of our findings, and because new thoughts begin with the same conceptual transitions and recruitment of brain regions associated with meta-state transitions, we suggest that meta-state transitions could serve as an objective biological landmark to delineate the temporal boundaries of spontaneous thoughts. In so doing, they may contribute new implicit insights to emerging research on thought dynamics.

Research on thought and topics centred upon it, such as mind-wandering, has emerged into the mainstream of cognitive neuroscience on the strength of new empirical approaches and converging findings (*1*). Problems remain to be overcome, however, including the unreliability of meta-cognition, on which subjective appraisal of thought often relies (*2*, *3*); the disruptive effects of meta-cognition itself; and as thoughts often drift, a lack of clarity about how to pinpoint a thought's beginning and end (*4*). According to most definitions, a thought is fundamentally an idea (or *concept*), with some definitions emphasizing its spontaneous production (*5*). A potential alternative to meta-cognition, therefore, is to employ implicit measures of concept activation: for example, researchers now routinely assess spatial patterns in functional magnetic resonance imaging (fMRI) data to decode neurally discernable concepts such as faces and houses (*6*). In principle, a new thought could be observed as a switch from one active concept to another, without any need for meta-cognition. Unfortunately, only a few categories have been reliably resolved in the brain at a time, whereas this approach would require us to resolve an unlimited number of potential concepts. But recent perspectives have emphasized the importance of dynamics, rather than contents, for best understanding thought (*1*). What if instead of following the rise and fall of particular concepts, we instead found a way to identify the conceptual transitions themselves?

Here we describe a novel neurocognitive landmark that robustly tracks conceptual transitions. To conduct our analysis, we drew upon the 7T Human Connectome Project dataset, which features both movie-viewing fMRI (mv-fMRI) and resting-state fMRI (rs-fMRI) data runs gathered from 184 participants (*7*, *8*). For each fMRI run, we computed the expression of 15 known brain networks over time (see *Supplemental Methods*). We then reduced the dimensionality of this matrix from *15 x time* to *2 x time* using t-SNE (*9*). In this reduced space, scans with similar overall 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 spatially contiguous worm-like series (arising from limited drift as thoughts "evolve"; Fig. 1A vs 1B).

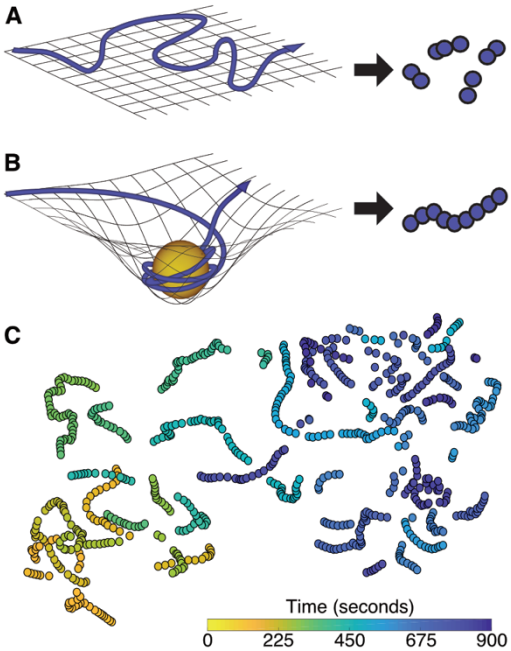

Fig. 1. Network space representation. (**A**) Continuous, random passage through the space of possible network configurations generates fragments in t-SNE space, as opposed to (**B**) contiguous, worm-like segments when an attractor holds network configurations in a relatively meta-stable state. (**C**) An example reduced t-SNE representation involving both segment types, as observed from a 15-minute rs-fMRI scan of one participant.

We found strong evidence for contiguity, with participant t-SNE trajectories reliably forming into series (see Fig. 1C and *Supplemental Methods*). To quantify this, we created for each fMRI run a step distance vector by taking the Euclidean distance in t-SNE space between successive timepoints, and identifying peaks within this distance vector. As each peak represents a relatively prominent reconfiguration of network meta-states, we labeled these points as network meta-state transitions (henceforth *transitions*). For purposes of baseline comparison, we also identified local minima in the distance vector. As each local minimum represents a relatively unchanging network meta-state, we labeled these points as meta-stable (Fig. 2A).

Some researchers have addressed the topic of cognitive dynamics in naturalistic cognition by leveraging film's unique ability to induce similar brain activity across participants (*10*–*13*); this work is often guided by event segmentation theory (*14*, *15*). Accordingly, we began our psychological validation with mv-fMRI data, examining the degree to which transitions were responsive to new concepts in film stimuli. Consistent with these past observations of film-based alignment, visual inspection of the set of all participant transitions during movie-viewing vs. during rest revealed movie-viewing to induce substantial alignment (Fig. 2 B,C).

As transitions are a whole-brain measure, we suspected it was specifically high-level conceptual structure in a clip that induces alignment in transition timing. To test this idea, we ran an analysis comparing the number of events in each movie clip – two expert raters defined events by consensus on the basis of the introduction of new concepts – with group alignment, which we assessed based on the similarity between the step distance vector of an individual participant and the corresponding median group step distance vector (see *Supplemental Methods*). Correlating events per minute with group alignment (separately for each participant), we obtained 184 correlation coefficients describing event rate by alignment relationships. We entered these into a group bootstrap analysis and found that clips with more events per minute had higher group alignment, mean $r = 0.27$, 95% CI: [0.24, 0.31].

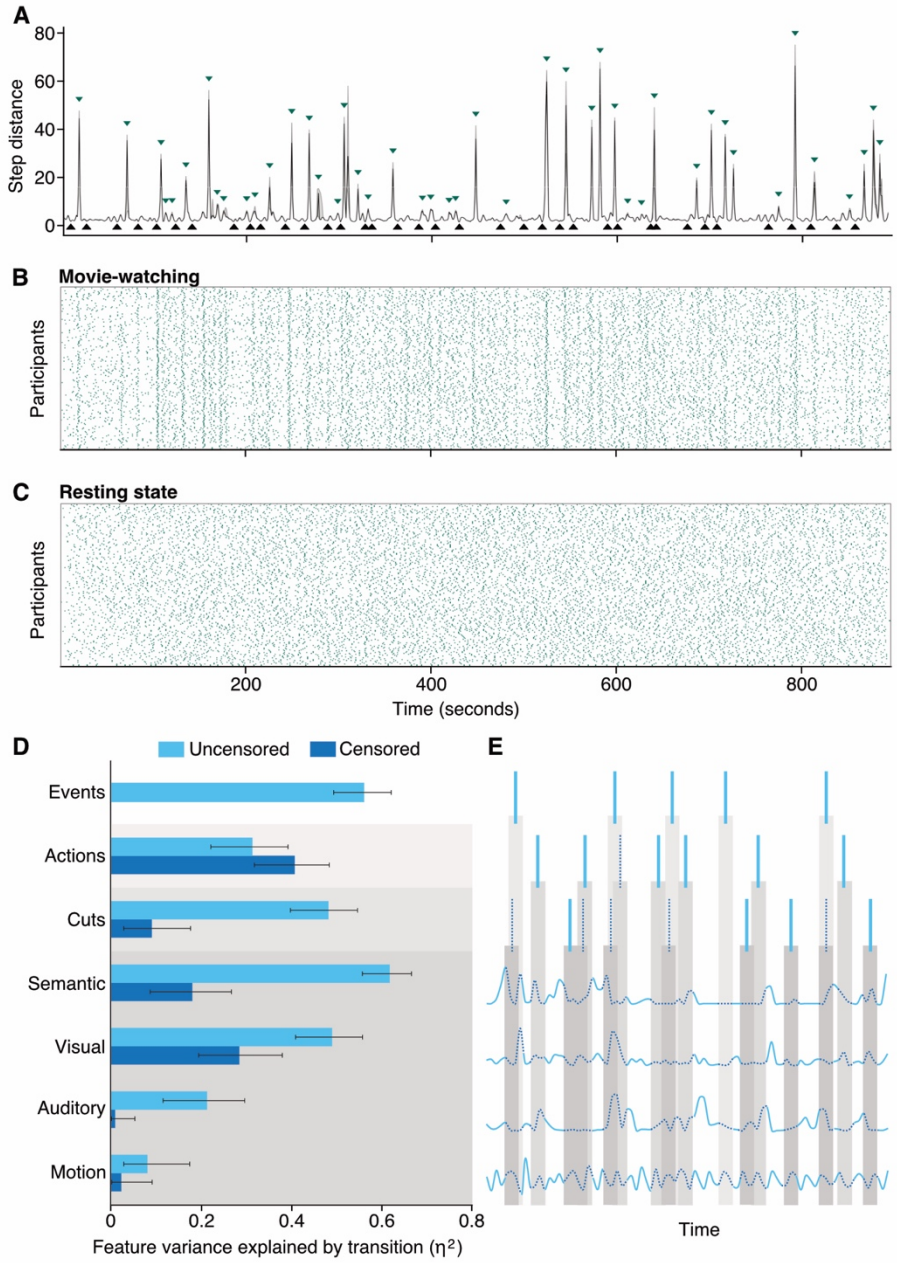

Fig. 2. Explaining network meta-state transitions using movie features. (**A**) A participant's median step distance vector during one mv-fMRI run, with 95% CI ribbon (narrow ribbon indicates stability over t-SNE iterations). *Transition* timepoints (green triangles) and *meta-stable* timepoints (black triangles) were 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' transition timepoints 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 next turned to prediction of movie features within-participant using transitions. As above, two expert raters defined actions and cuts for each clip using a consensus approach (*Supplemental Methods*). We also obtained lower-level feature time vectors for each clip, including semantic change, visual change, amplitude change, and change in head motion. These features are correlated; for example, the end of an event often coincides with a cut between shots, which in turn coincides with the onset of a new set of semantic and perceptual stimulus features. 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 better assess high- and low-level features independently, we therefore "censored" epochs of the timeseries of lower-level features where higher-level event boundaries co-occurred (Fig. 2E). If lower-level features themselves induce transitions, they should do so even outside of censored epochs.

We calculated eta-squared values for each set of features (uncensored and censored) after accounting for lag in the hemodynamic response function (*Supplemental Methods*). Transitions were most successful at predicting the onset of new concepts: they accounted for 56.0% of variance in movie events and 61.7% of variance in semantic change. However, only 17.9% of semantic change variance was accounted for by transitions after boundaries of events, actions, and cuts were censored. In fact, eta-squared values for all lower-level movie features (semantic, perceptual, and auditory), and even participant head motion, decreased markedly after removing timepoints that could be confounded with event boundaries. Accordingly, in our analysis, transitions were most prominently markers of new concepts. These results complement those of Baldassano and colleagues (*12*), 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. Our findings are distinctive in that they endow a biological phenomenon (network meta-state transitions) with psychological meaning. Because our approach relies on landmark identification, it also, unlike HMM, requires no predefinition of the number of stable periods, and therefore uniquely lends itself to tracking activation of conceptual transitions when brain data have unknown complexity, as with data gathered during rest.

We drew on this affordance by turning to rs-fMRI scans gathered from the same participants, where conceptual transitions potentially could be observed outside of the influence of stimulus control. To validate our approach, we assessed the similarity of the characteristics of transitions across mv-fMRI and rs-fMRI. 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 no circularity or bias towards the spatial characteristics of any one network was intrinsic to this analysis. 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-

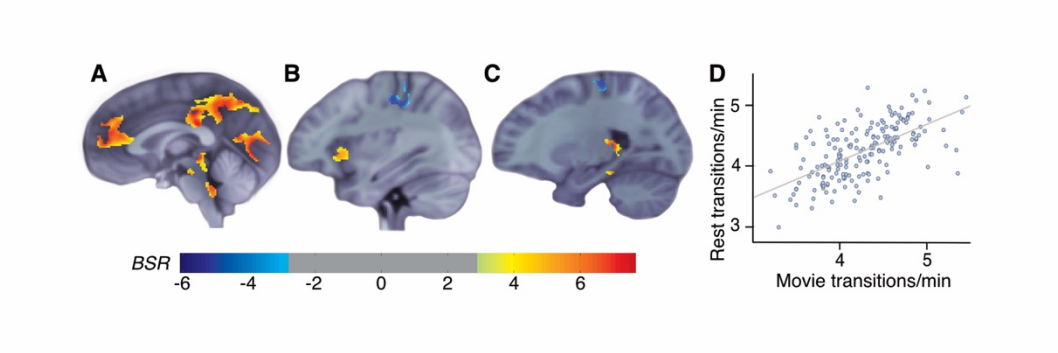

Fig. 3. Transition features generalize from stimulus-driven to resting cognition. (**A**-**C**) Conjunction analysis revealed brain activity that consistently differentiated transition from meta-stable timepoints, regardless of whether the active task was movie-viewing vs. rest. (**A**) Anterior cingulate, 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**) Participant movie-viewing transition rate predicts participant resting cognition transition rate.

viewing or rest, transitions were associated with a group of regions that have been previously associated with spontaneous thought (Fig. 3 A-C).

We also evaluated whether the rate at which transitions occurred for a given individual generalized across task and rest. This first required us to determine whether the rate of transitions was internally stable across the four rs-fMRI runs (each of which was sampled in a separate scanner session). We found transition rate to be correlated across runs (mean $r = 0.43$), revealing it to be a relatively trait-like characteristic that can be adequately sampled using 15 minutes of rs-fMRI data. We then averaged transition rate across the four rs-fMRI runs to further stabilize our measurement of the trait. Finally, we regressed the average participant-wise transition rate computed from rs-fMRI data earlier against the average participant-wise transition rate computed from the four mv-fMRI runs. Transition rate at rest significantly predicted transition rate during movie-viewing, $r(181) = 0.52$, $p < .001$ (Fig. 3D), thereby further linking its properties across stimulus-driven and resting cognition.

In our conjunction analysis, we observed greater posterior hippocampus (pHPC) activity at transition relative to meta-stable timepoints. This finding aligns with reports that the hippocampus is sensitive to the onset and offset of a movie stimulus (*16*), as well as more subtle conceptual transitions within a movie (*12*). It also aligns with a proposal from our group that, whereas pHPC plays a role in distinguishing similar events in memory (*17*), anterior hippocampus (aHPC) plays a distinctive role in gist-like representations (*18*). We hypothesized that because aHPC is associated with greater pattern completion (*19*) because of a relatively small dentate gyrus (*20*), 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 explored the possibility that 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).

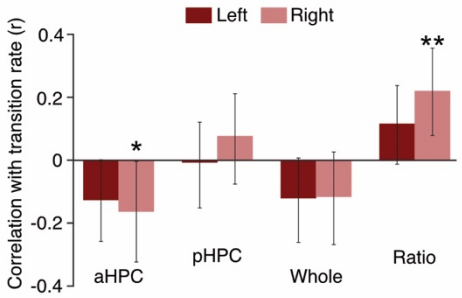

Fig. 4. Grounding of transitions in neurobiology of event cognition. Prediction of transition rate by anterior hippocampus (aHPC) volume, posterior hippocampus (pHPC) volume, whole hippocampus volume, and the pHPC to aHPC volume ratio using transition rates derived from movie-viewing and rest data. *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

To summarize, we used novel methods to characterize a neurocognitive landmark of conceptual change 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 the rate of transitions in rs-fMRI data to be a reliable, trait-like characteristic across imaging sessions, and also found transitions to be highly responsive to new concepts in movie stimuli. Characteristics of transitions observed during movie-viewing data were shared by transitions observed during rest, supporting a unified interpretation of transitions across task contexts.

In light of the importance of conceptual transitions for distinguishing thoughts, our findings further suggest meta-state transitions may serve as a landmark with utility for better understanding the cognitive neuroscience of thought. Further grounding meta-state transitions in this construct, the specific brain regions our analysis linked to transitions are prominent in this literature. In particular, anterior cingulate and insular cortices are considered members of a salience network that detects both external and internal events (*21*). The posterior cingulate cortex is a key region in the default mode network that is more active during task-unrelated than task-related thought (*22*, *23*). Parahippocampal cortex and the hippocampus have been found in association with unaware (and therefore unconstrained) cognitions during task-unrelated thought (*24*). Finally, our findings support the idea that whereas the posterior hippocampus helps to sever events at conceptual transition points, a larger anterior hippocampus increases the density of mnemonic attractors in network state space, contributing to greater attentional capture and dwell time.

Although our analysis has focused on movie and resting fMRI data from healthy adults, the methods we employed are applicable to a wide range of tasks, populations, and even case studies. The landmark characterized here therefore represents an empirical solution to meta-cognitive challenges associated with delineating thoughts: what if we were to measure them using neurobiology, whether by reconfiguration of brain networks, activation of particular nodes, or other observable phenomenon? This strategy opens up opportunities to ask 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 in our analysis, and a recommended sleep time of 8 hours, one could tentatively estimate a mean of 4,128 thoughts for healthy adults of a demographic similar to the one used in our study. Although further discussion and 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 *et al.*, 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 *et al.*, The WU-Minn Human Connectome Project: an overview. *Neuroimage*. **80**, 62–79 (2013).
9. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
10. U. Hasson *et al.*, Neurocinematics: The Neuroscience of Film. *Projections*. **2**, 1–26 (2008).
11. J. Chen *et al.*, Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).
12. C. Baldassano *et al.*, Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*. **95**, 709–721 (2017).
13. A. C. Heusser, P. C. Fitzpatrick, J. R. Manning, “How is experience transformed into memory ?” (2018), , doi:https://doi.org/10.1101/409987.
14. J. M. Zacks, K. M. Swallow, Event Segmentation. **16**, 80–84 (2007).
15. J. M. Zacks *et al.*, Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.* **4**, 651–5 (2001).
16. 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).
17. 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).
18. J. Poppenk, H. R. Evensmoen, M. Moscovitch, L. Nadel, Long-axis specialization of the human hippocampus. *Trends Cogn. Sci.* **17**, 230–240 (2013).
19. M. A. Yassa, C. E. L. Stark, Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–525 (2011).
20. 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).

21. W. W. Seeley *et al.*, Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. **27**, 2349–2356 (2007).
22. R. Leech, D. J. Sharp, The role of the posterior cingulate cortex in cognition and disease, 12–32 (2014).
23. J. Smallwood *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).
24. 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).

Acknowledgements

We thank Y. Lu for help with event, action, and cut segmentations, and S. Smith for manual hippocampal segmentations. 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 (*25*, *26*). 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*, *27*), 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) (*28*–*31*). 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.

Step 1: Timepoint discovery

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 (*25*, *26*), 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.

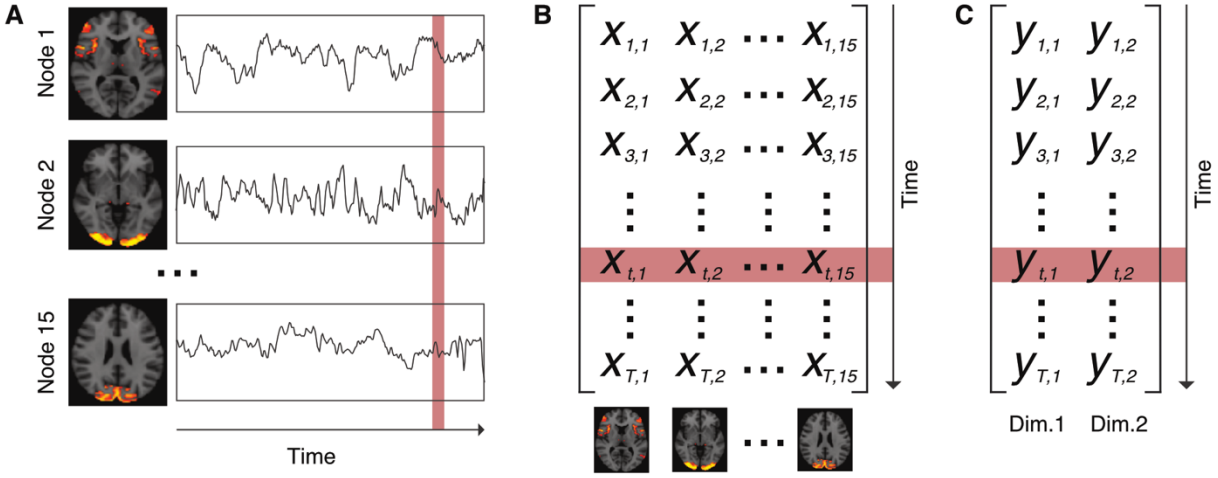

Fig. S1. Expression of brain network activity. (**A**) Each fMRI run of each participant was characterized as the activation of 15 brain networks over time. (**B**) Representation of A as a two-dimensional matrix of (time x network). (**C**) Reduced representation obtained after applying t-SNE algorithm.

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). Next, 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) (*9*). By applying a dimensionality reduction algorithm, we tease out the most salient changes by forcing the algorithm to find a lower-dimensionality representation that prioritizes key structural aspects of its higher-dimensionality representation. We found the perplexity setting had little impact on our analysis, and therefore selected what is regarded as a moderate (and default) value (*9*).

Notably, Miller and colleagues (*32*) 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 (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. Our approach is more flexible, as it relies on dimensionality reduction algorithms to discover changes in meta-state directly from the continuous-valued 15-brain-network representation. In other words, 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 better aligned to the very likely possibility that meta-states are continually evolving. It is more likely to be sensitive to transitions to and from updated meta-states. In spite of this more flexible approach, our approach actually provides a more stable basis for inquiry and comparison across experiments, because it draws on published reference networks derived from 820 participants (i.e., the training dataset described above; 8, 25, 26), rather than running group temporal ICA on the same smaller sample being used for inference. It is also compatible with pilot and case study research, since no group sample is required.

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 vector in 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 baseline 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.

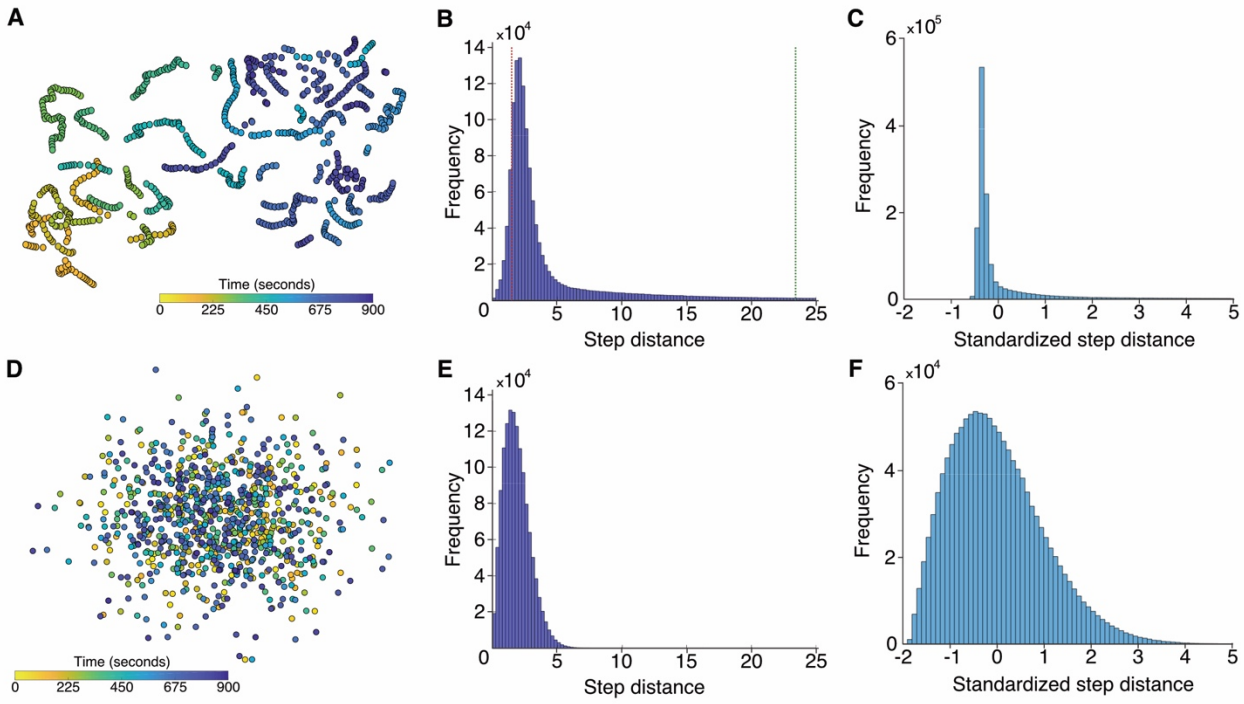

Fig. S2. Median normalized step distance values. (**A**) An example reduced t-SNE representation involving both segment types, as observed from a 15-minute rs-fMRI scan of one participant. (**B**) A histogram depicts the median normalized step distances within 7T participants' four rs-fMRI and four mv-fMRI 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.

Step 2: Psychological relevance

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, *15*). Cuts referred to boundaries between two separate camera shots. Two raters used a video coding tool (Datavyu; *19*) 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).

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 (*34*), and motion-energy labels that described low-level structural features in the same epochs (*35*). 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 (*36*), 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 timepoints in the censored vectors, 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 during the hemodynamic response function (HRF) peak (working backwards based on the canonical HRF, *30*, 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 (*38*). 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).

By visual inspection, it seems that there exists structure in the transitions across participants during movie-viewing (Fig. 2B), which we expected on the basis of findings in other studies of inter-subject correlations between the brain activity of participants watching the same (well-made) film. We investigated the relationship between the number of high-level events in a clip and the degree of group alignment in brain activity. For each clip, we divided the number of events in that clip by its duration in minutes. For each participant, we obtained their 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; 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; *27*).

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 *33*). 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., *34*). As the bootstrap ratio approximates a z-distribution (*42*), 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 (*32*).

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, *44*, *45*) 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 (*18*). 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 (*46*).

Supplemental References

25. 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).
26. 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).
27. M. F. Glasser *et al.*, The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*. **80**, 105–124 (2013).
28. S. Moeller *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).
29. D. A. Feinberg *et al.*, Multiplexed Echo Planar Imaging for Sub-Second Whole Brain FMRI and Fast Diffusion Imaging. *PLoS One*. **5**, e15710 (2010).
30. K. Setsompop *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).
31. J. Xu *et al.*, in *Proceedings of the 20th Annual Meeting of ISMRM* (2012), vol. 2306.
32. R. L. Miller *et al.*, Higher Dimensional Meta-State Analysis Reveals Reduced Resting fMRI Connectivity Dynamism in Schizophrenia Patients. *PLoS One*. **11**, e0149849 (2016).
33. D. Team, Datavyu: A Video Coding Tool (2014), (available at http://datavyu.org).
34. 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).
35. S. Nishimoto *et al.*, Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* **21**, 1641–1646 (2011).
36. M. Jenkinson, “Measuring Transformation Error by RMS Deviation” (Oxford, 1999), (available at https://www.fmrib.ox.ac.uk/datasets/techrep/tr99mj1/tr99mj1.pdf).
37. G. K. Aguirre, E. Zarahn, M. D’Esposito, The variability of human, BOLD hemodynamic responses. *Neuroimage*. **8**, 360–369 (1998).
38. K. A. Bollen, R. Stine, Direct and Indirect Effects : Classical and Bootstrap Estimates of Variability. *Sociol. Methodol.*, 115–140 (1990).
39. M. Mikl *et al.*, Effects of spatial smoothing on fMRI group inferences. *Magn. Reson. Imaging*. **26**, 490–503 (2007).
40. 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).
41. J. Poppenk, K. A. Norman, Briefly Cuing Memories Leads to Suppression of Their Neural Representations. *J. Neurosci.* **34**, 8010–8020 (2014).
42. A. R. McIntosh, B. Mišić, Multivariate Statistical Analyses for Neuroimaging Data.

Annu. Rev. Psychol. **64**, 499–525 (2013).
43. 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).
44. B. Fischl, FreeSurfer. *Neuroimage*. **62**, 774–781 (2012).
45. B. Fischl *et al.*, Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. **33**, 341–355 (2002).
46. 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).