Supplemental Information

Supplemental Methods

Data Acquisition

The data for this study was obtained by leveraging the library of resting-state and task fMRI images from the Human Connectome Project (HCP), a joint project between Washington University and the University of Minnesota (Van Essen et al., 2012b). These data were acquired using a customized Siemens 3T "Connectome Skyra" and the 32 channel, anterior/posterior, head receive coil. T1 weighted anatomical scans were acquired via a 3D MPRAGE sequence with $TR = 2400$ ms, $TE = 2.14$ ms, $TI = 1000$ ms, flip angle of 8°, $FOV = 224 \times 224$ mm, and voxel size 0.7 mm isotropic. BOLD-weighted fMRI images were acquired via a gradient-echo EPI sequence with $TR = 720$ ms, $TE = 33.1$ ms, flip angle of 52°, $FOV = 208 \times 180$ mm, 72 slices, 2.0 mm isotropic voxels, and multiband factor of 8. Functional scans imaged individuals while they adopted a comprehensive battery of states. These states may be subdivided into the 9 scans named as follows: REST1, REST2, EMOTION, GAMBLING, LANGUAGE, MOTOR, RELATIONAL, SOCIAL, and WORKING MEMORY (WM). Each scan was performed twice for each volunteer, each time with an opposite phase encoding gradient (left to right, vs right to left). In total, each individual contributed 8,680 temporal and 91,282 spatial data points. REST scans spanned 4,800 time points. All data were de-identified before download.

Data Preprocessing

Data preprocessing include spatial artifact and distortion removal, surface generation, anatomical registration, and alignment to grayordinate space (gray-matter vertices or voxels). Subsequent use of spatial filters from a separate Independent Component Analysis (ICA) assume that data follow an isotropic noise model; thus, all voxel time series are normalized to zero mean and unit variance. To reduce the influence of edge effects during spectral filtering, contiguous, 300 image segments from a volunteer's REST scans were placed in-between their task scans. The remaining 900 REST images capped the beginning and the end of the concatenated series with 450 time points each. The order of concatenated rest and task images were randomized across volunteers.

Spectral and Spatial Filtering

The BOLD signal bears a log linear relationship between power spectrum and frequency: $\log S(f) = c + \gamma \log f$; alternatively, $S(f) \sim 1/f^{\gamma}$. For the average BOLD signal in brains, the power law exponent, $\gamma \cong -1$. The variable c is a constant. Such '1/f-type' systems denote that the system's high-frequency realizations establish and maintain its low-frequency structure (Wornell, 1993). The simplest 1/f-type systems are termed, 'scale-free,' that is, one observes rescaled versions of some elementary process, a "fractal", at all observable scales. On the other hand, complex 1/f-type systems exhibit emergent properties at multiple scales (Ciuciu et al., 2012; He, 2014; Liu et al., 2014). A theoretically optimal method for observing 1/f-type processes is to transform them using a scale-free, or multiresolution, basis set (Bullmore et al., 2004; Ciuciu et al., 2012). Coefficient variance in the scale-free domain is thus a representation of the emergence of novel signal characteristics resolved to one or more scales (Daubechies, 1992). Wavelet transforms are especially useful multiresolution transforms as their kernel functions reduce to zero over a finite time-span. By convolving a 1/f-type signal with a finite, scale-free

kernel, wavelet transforms highlight the signal's dynamical properties in both the temporal and the spectral domains.

Previous studies demonstrated that BOLD data segment into static FC subnetworks from the application of multiscale filter banks (Billings et al., 2015). In the present study, spectral filtering utilized the continuous wavelet transform with a Daubechies 4 wavelet. This continuous wavelet filterbank segmented BOLD signals into an octave of 8 frequency bands log-spaced across the decade [0.01, 0.1] Hz. This frequency range corresponds to the low-frequency fluctuation range in which BOLD fluctuations bear maximal information about neuronal activity. The mother wavelet, Daubechies 4-tap, was chosen to achieve a relatively short temporal window over each spectral band, while the number of bands is sufficient to capture the inter scale network variation observed in Billings et al. 2015.

Spatial filtering utilized a 50-component ICA decomposition. The ICA transform matric was calculated as part of the HCP beta-release of group-ICA maps (Human Connectome Project, 2014). The number of components was chosen by identifying the intersection between the eigenvalues of a volunteer's real concatenated input data matrix, and a randomly shuffled version of that matrix, and choosing a number of components that just exceeded this point of intersect (data not shown).

Manifold Embedding

Each temporal sample for each volunteer's high-dimensional state descriptor (50 spatial components by 8 spectral components) was pairwise compared using the Pearson correlation distance,

$$d_{ij} = \|x_i - x_j\| = 1 - \frac{(x_i - \bar{x_i})(x_j - \bar{x_j})'}{\sqrt{(x_i - \bar{x_i})(x_i - \bar{x_i})'}\sqrt{(x_j - \bar{x_j})(x_j - \bar{x_j})'}}. \quad (1)$$

Because of the theoretically optimal whitening properties of the wavelet transform, and because we have normalized time series via z-scoring, the Pearson correlation distance highlights coordinated deviations from normative spectral intensities.

Manifold embedding was performed with the algorithm t-Distributed Stochastic Neighborhood Embedding (t-SNE) (Berman et al., 2014; van der Maaten and Hinton, 2008; van der Maaten et al., 2009). The algorithm begins by transforming high-dimensional pairwise distances into conditional probabilities, $p_{j|i}$, along a Gaussian probability distribution,

$$p_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)}. \quad (2)$$

The variable σ_i is equal to the variance of the high-dimensional data when multiplied by a Gaussian centered over point i. The width of each Gaussian is adjusted to cover an equivalent amount of points. Formally, the width is adjusted until the base 2 exponent of the Shannon entropy measured in the stochastic distribution around the i^{th} point achieves a fixed value termed the perplexity. For the present study, we follow the recommendation from van der Maaten & Hinton (2008) of a perplexity equal to 30. Collectively, the transformation from inter-sample distances to conditional probabilities emphasizes the natural associations of each sample to its neighbors. The authors of t-SNE also described a problem with

previous implementations of SNE-based algorithms wherein moderately dissimilar samples, in the high-dimensional space, crowd together in the low-dimensional map (van der Maaten and Hinton, 2008). Therefore, t-SNE calculates the low-dimensional probabilities, Q, using a distribution having a much longer tail than in the high-dimensional case. A good choice to avoid this problem was found to be the Student t-distribution with one degree of freedom:

$$q_{ij} = \frac{\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}}, \qquad (3)$$

where $\left\|y_i - y_j\right\|$ is the Euclidean distance between samples i and j in the low-dimensional space.

A natural cost function, C, to calculate the fidelity of the low-dimensional representation relative to the high-dimensional data is the Kullback-Liebler (KL) divergence which is related to the cross-entropy between the two distributions. A symmetrized version of the KL divergence is used here to expedite computation time and to balance the cost of representing points that are close together in the high-dimensional space as distant points in the low-dimensional space, and vice-versa. Thus,

$$C = \sum_i \mathrm{KL}(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}. \qquad (4)$$

The joint probabilities in the high-dimensional space are calculated as $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$, where n are the number of samples.

This collective description of high-dimensional and low-dimensional spaces, as well as the relationship between them, emphasizes both that similar map points are modeled by small pairwise distances and that dissimilar map points are modeled by large pairwise distances. This is the case at all but the finest scales, at which point, the numerator of equation (3) is dominated by a constant rather than by variations from the input data. The t-SNE algorithm is implemented as a gradient descent process. The form of the gradient, as well as detailed notes on methods to improve the speed of convergence may be found in van der Maaten and Hinton, 2008.

The initial construction of a t-SNE embedding is computationally expensive: $O(n^2)$. For a compute node having 256 GB of RAM, the maximum number of double precision data points that may be included in a single t-SNE embedding is limited to about 90,000 samples. The full complement of 4 resting state scans and 14 task scans contains 8,680 images for each of the 446 included volunteers. To overcome the computational limits of embedding larger datasets, the present study follows the recommendations from Berman et al. for training a low-dimensional embedding space from a subsampling of data points (Berman et al., 2014). Briefly, t-SNE embeddings were generated from each of 446 volunteers, individually. Next, 200 sample points were pulled from each volunteer's embedding, at random, and in proportion to the density of points within the embedding. A group-level embedding was then trained from each volunteer's sample of 200 time points. The best low-dimensional locations of the remaining time points vis-à-vis the trained embedding were then calculated in two steps: 1) Approximate the out-of-sample point's low dimensional location as a weighted sum of its nearest neighbors in the full high-

dimensional space. 2) Determine the local KL divergence minimum by changing only the location of the out-of-sample point. As this minimization is not convex, it is worthwhile to jitter the out-of-sample point's initial low-dimensional location by sampling from a range of its high-dimensional neighbors. This procedure reduces the computational load to O(n). The subsampling procedure greatly increases the interpretability of the resulting map by removing the bias experienced among sequentially sampled—and hence, temporally correlated—points when they are embedded simultaneously.

Sub-Space Identification and Characterization

One method to summarize 2-dimensional point distributions is by convolution with a Gaussian filter. In order to account for both coarse and fine features of the embedded distribution, two filter radiuses were selected for the present study—one at 1/32 the maximum displacement from the map center and the other at 1/256. Particularly dense map regions are segmented from one another, in a data-driven fashion, by taking the watershed transform of the inverse of each density map (Meyer, 1994).

Velocity Field

Instantaneous velocities, were calculated by taking the difference in the embedded location of successive sample points. The group-level displacement magnitude was averaged, separately, in each of the 4 cardinal Euclidean directions, -i, +j, +i, and -j, for each point in a 32x32 grid. Results were normalized to unit magnitude.

Comparing Embeddings

Embeddings were segmenting against the HCP's experimentally defined states, i.e. the resting-state and the task-based scans. To test the inference that scan-segmented maps depicted distinct brain-state distributions, we conducted an ANOVA with multiple comparisons testing using a bootstrapped sample of each experimentally defined state. Points within each bootstrap realization were chosen from segmented group-level datasets. The lower bound to the number of points in each bootstrap realization sample was chosen to ensure a full coverage of the state's embedded range. The upper bound was chosen to ensure that few points were sampled twice in any two bootstrap realizations.

Bootstrap realizations were pairwise compared using the Structural Similarity Index (SSIM) (Zhou et al., 2004). SSIM measures the similarity between two images, *x* and *y*, as the multiplicative combination of three image quantities, the cross-luminance, *I*, cross-variance, *c,* and cross-structure, *s.* Thus:

$$SSIM(x,y) = [I(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma},$$

where

$$I(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

where μ_x, μ_y, σ_x,σ_y, and σ_{xy} are the local means, standard deviations, and cross-covariance for images *x, y*. The values C_3, C_3, and C_3 are small constants given by $C_1 = (K_1 L)^2, C_2 = (K_2 L)^2$, and $C_3 = C_2/2$. Here L is the dynamic range of pixel values. The variable $K_1 \ll 1$, and the variable $K_2 \ll 1$. The exponents over each SSIM term were set to 1 so as to weight each term equally. SSIM values range between 0, no image similarity, and 1, complete image similarity. SSIM statistical testing was conducted simultaneously for all SSIM pairs (50*50/2 independent comparisons). Maps were deemed to provide insignificant segmentation if the 95% confidence interval of the within-state SSIM fell within or below the range of any of its between-state SSIM 95% confidence intervals. The multivariate construction of the SSIM algorithm makes it a useful technique for quantifying the differences between density maps. Density maps contained equal numbers of points to ensure that the SSIM metric to remain balanced.

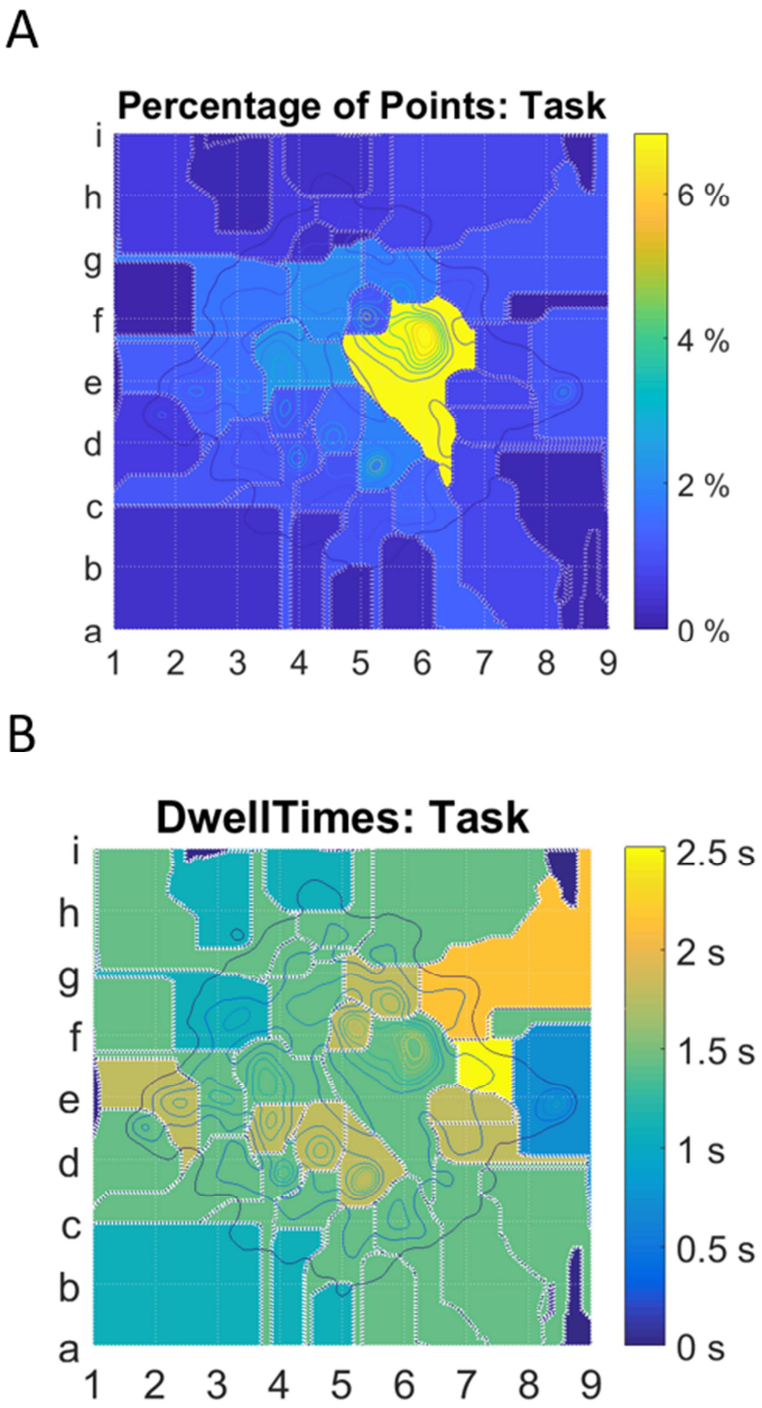
Real-Time Dynamics

Group level brain-state dynamics are characterized through map segmentation at the level of each task's block-design contrasts. For instance, MOTION task blocks are segmented into movement of the tongue, the left hand, right hands, etc. The total set of block-design contrasts, from all individuals and each individual's task repetitions, are aligned at time t = 0s, the start of the block (including the cue, if present). Group level density images are then calculated for each aligned image acquisition. The resting state was aligned to a single time point.

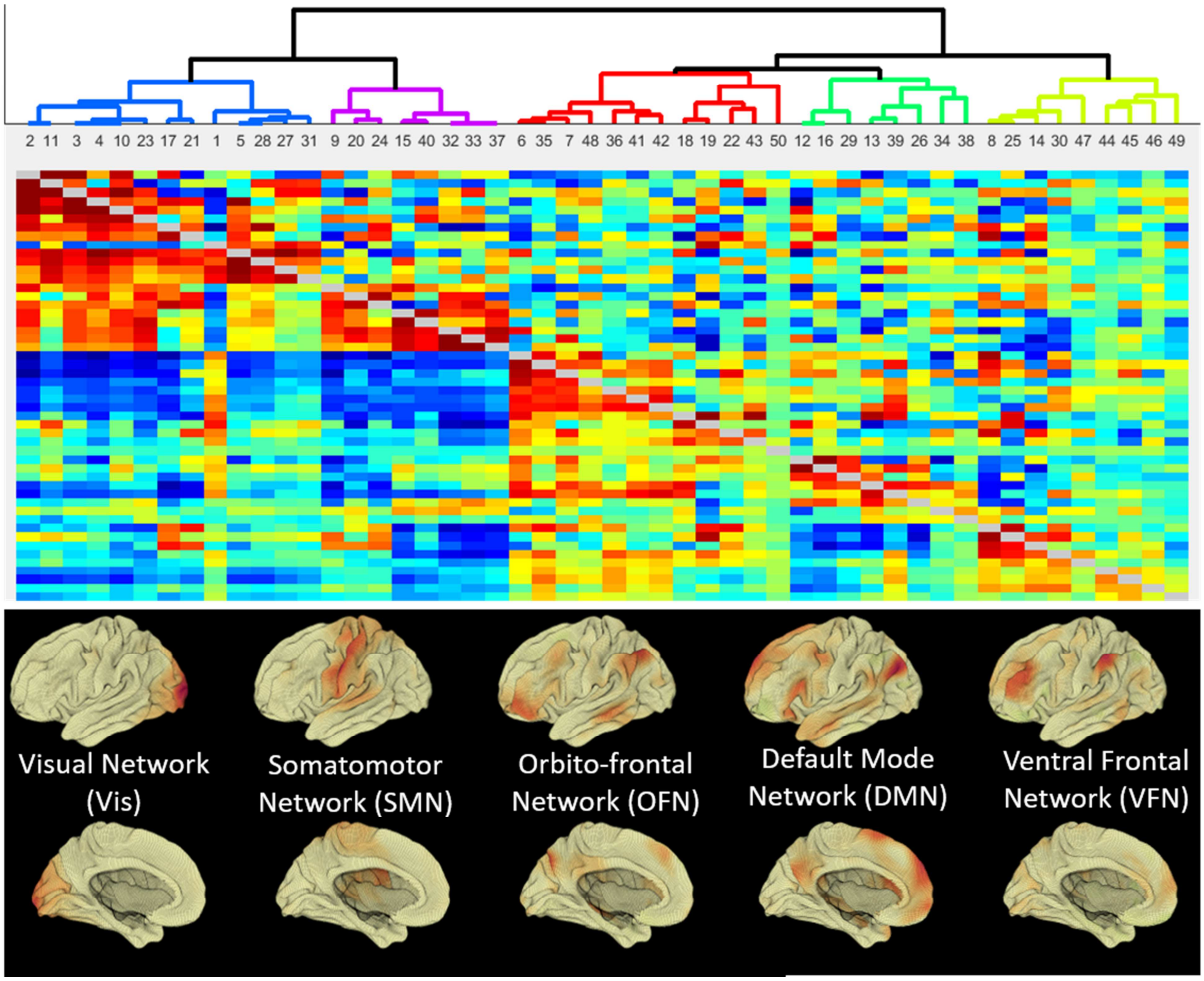
Permutation Testing for Labeling the Embedded map.

To test the preference of labeled times and conditions for particular map regions, a null distribution was constructed by randomly permuting the labels assigned to each embedded point. Thereafter, it is possible to compare the mean number of points randomly assigned to each region, under a particular condition/time, with the actual number of points in that region, at that same condition/time. Map regions may then be labeled in terms of the preference of each region for particular condition/times.

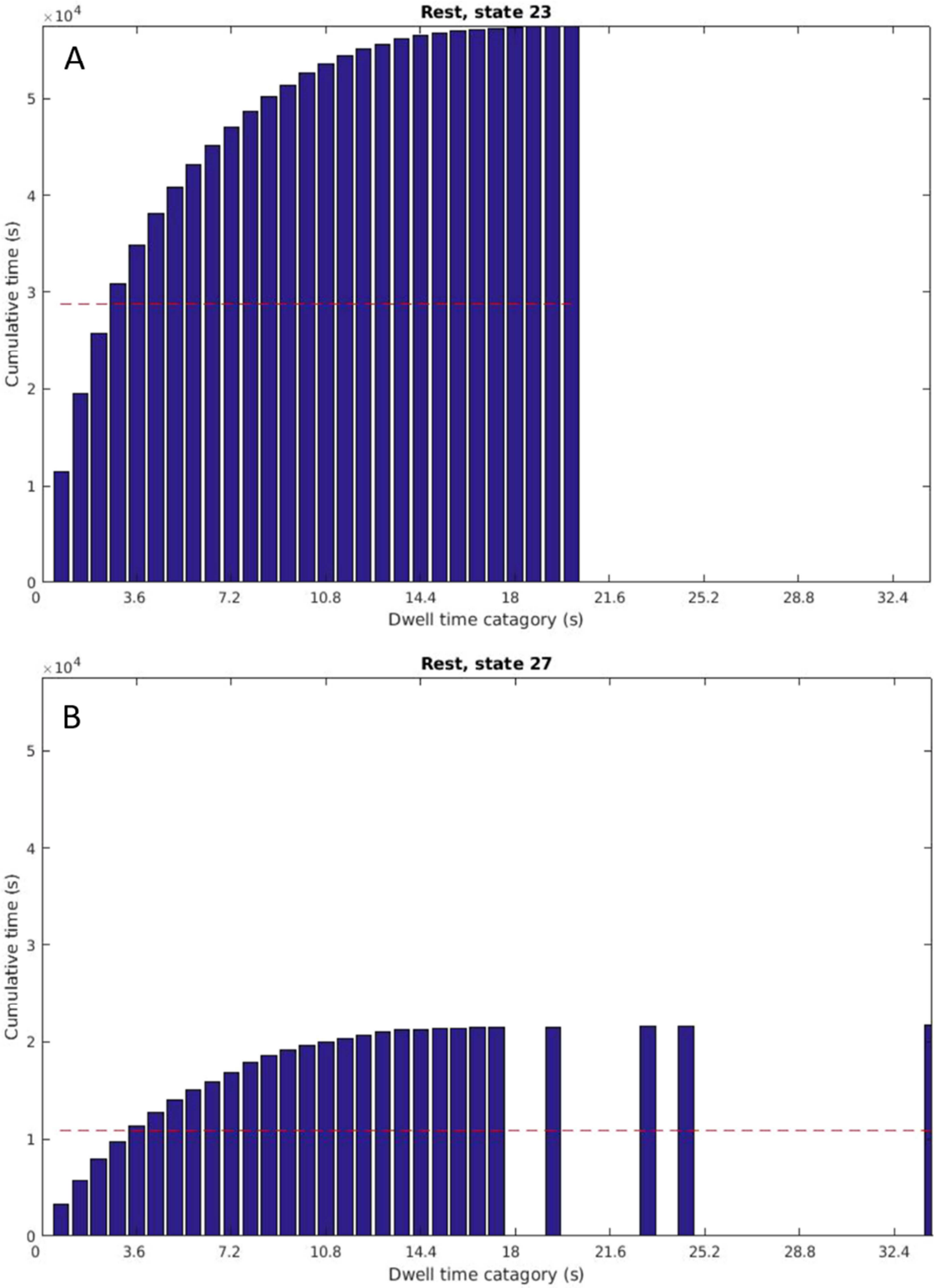
Supplemental Figures



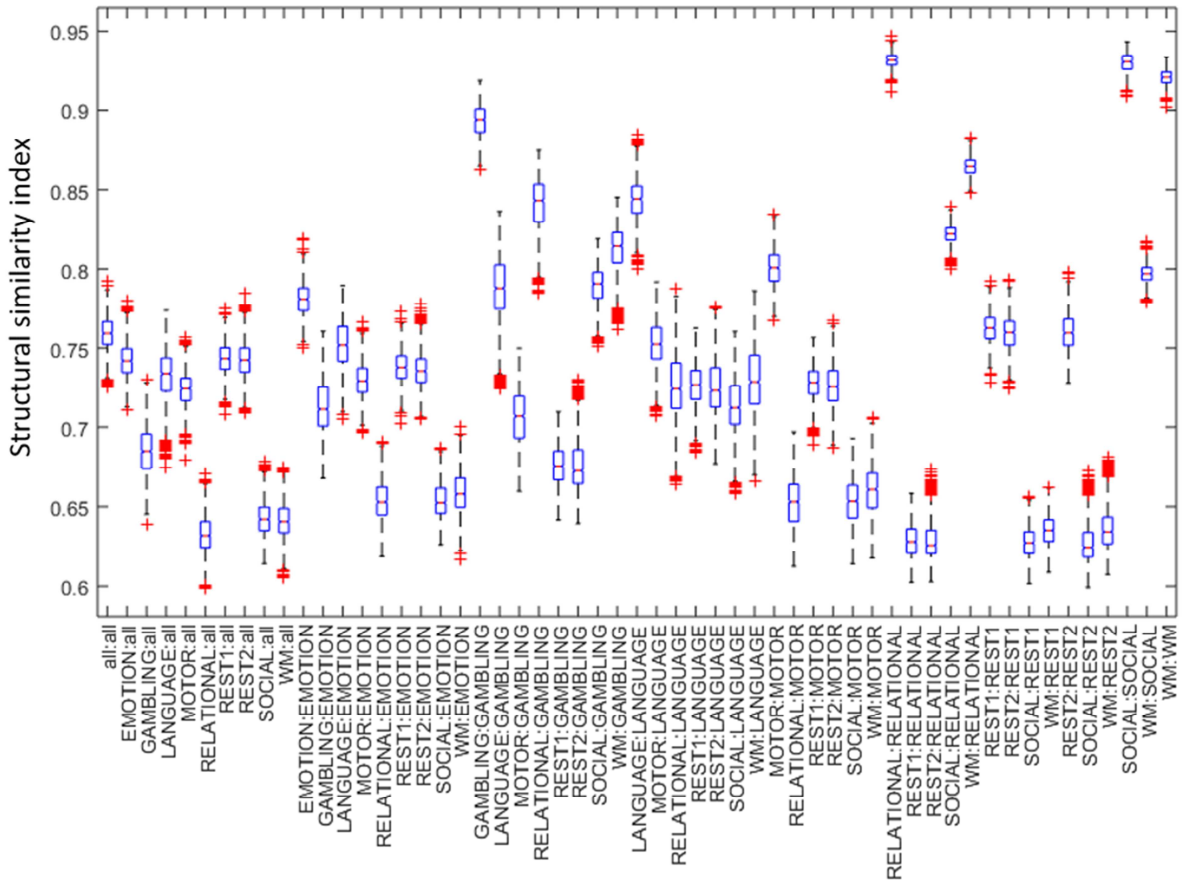
Supplemental figure S1 elaborates on the point distributions within watershed catchment-basins for task-active maps. Part A displays the percentage of points contained within each region. Part B displays the dwell-time for each region, reported as the mean number of temporally contiguous points contained within each region.



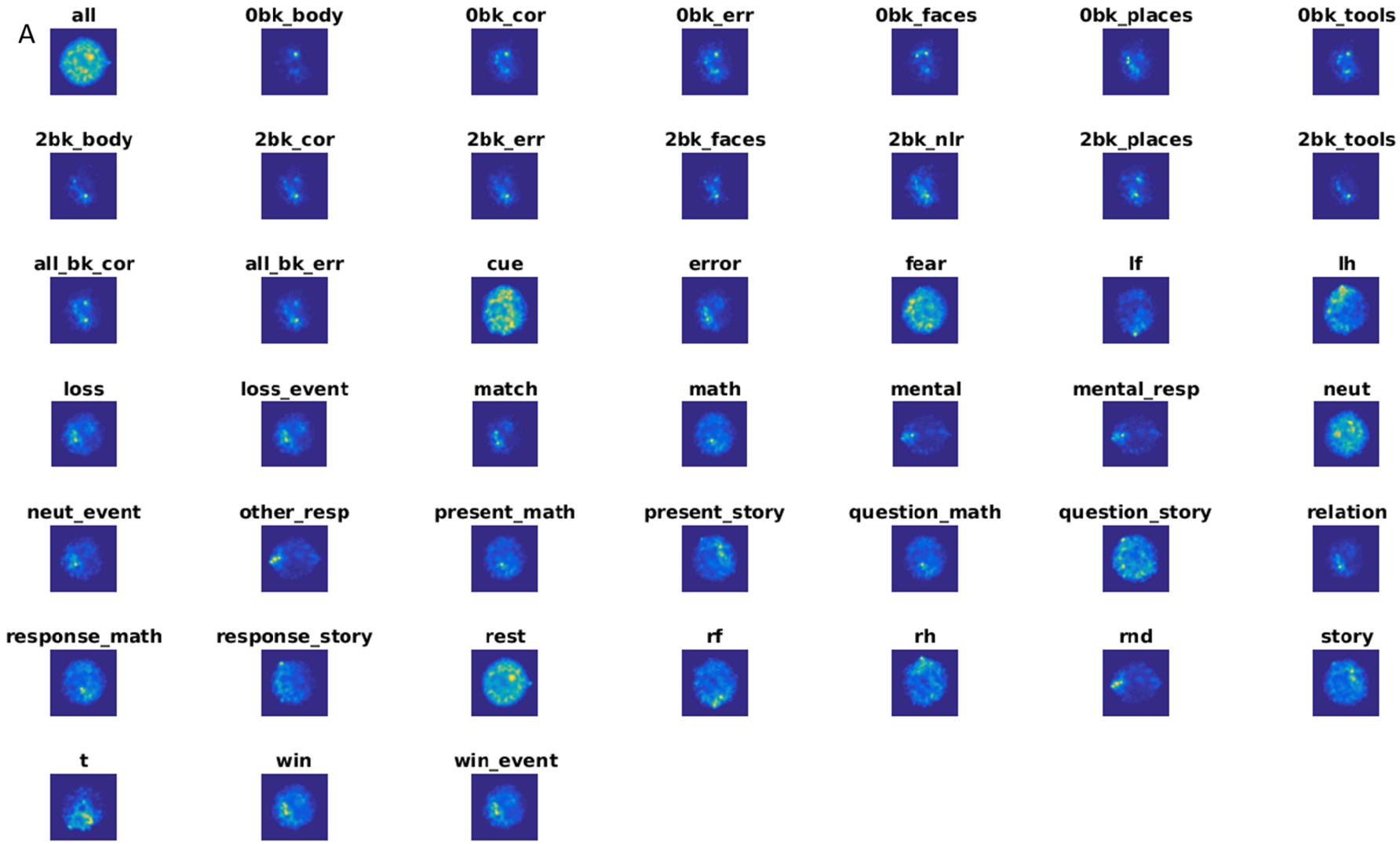
Supplemental figure S2 displays the relatedness between each of the 50 ICA components. Data were generated using the FSLNets toolbox provided through the HCP. The hierarchical clustering map was calculated from time-series from each ICA network back-projected for each volunteer included in the original analysis. The projection onto the brain of each of 5 ICA clusters is also shown.

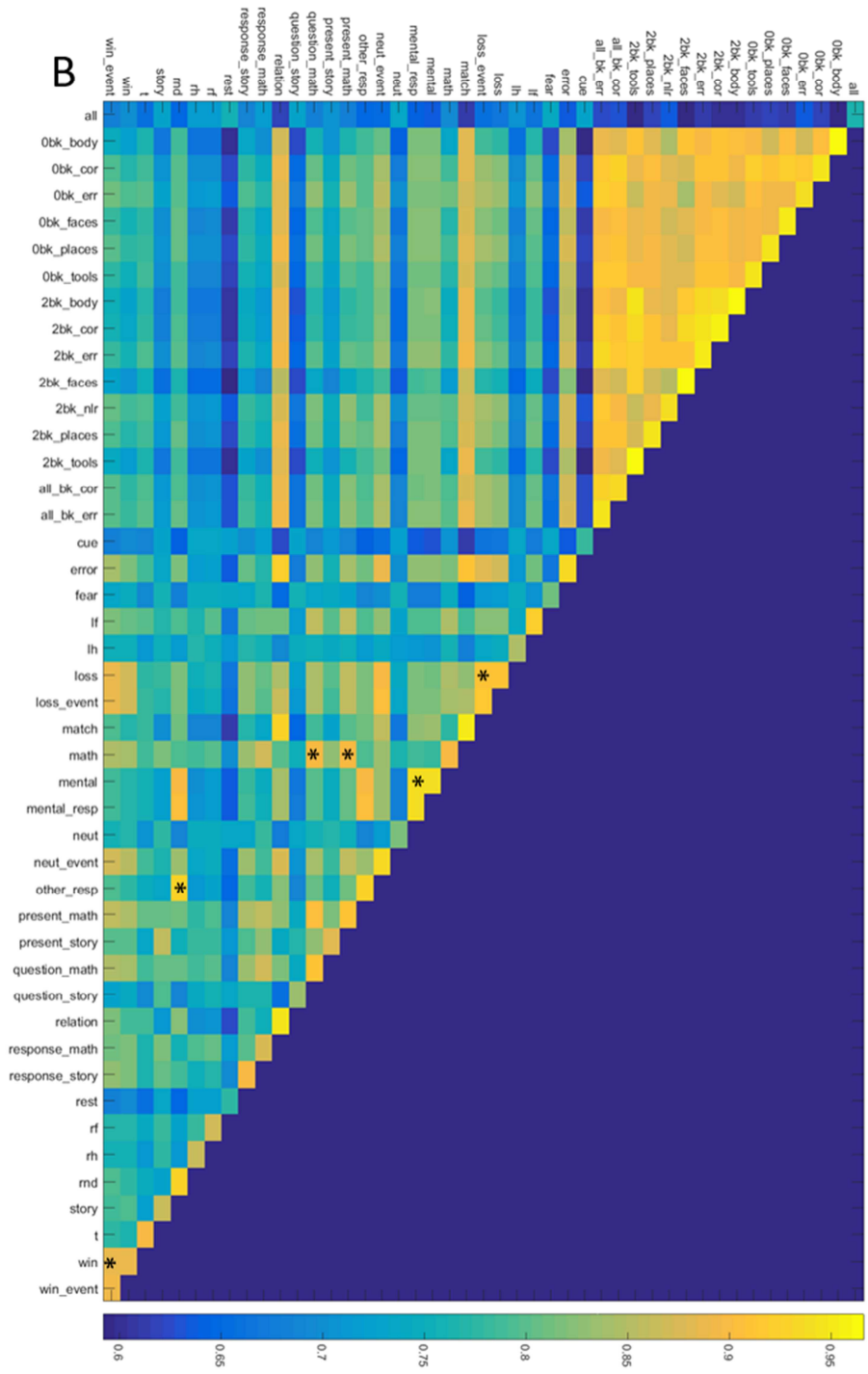


Supplemental figure S2 displays the dwell-time distribution for the highlighted regions in parts A and B (respectively) of figure 2.

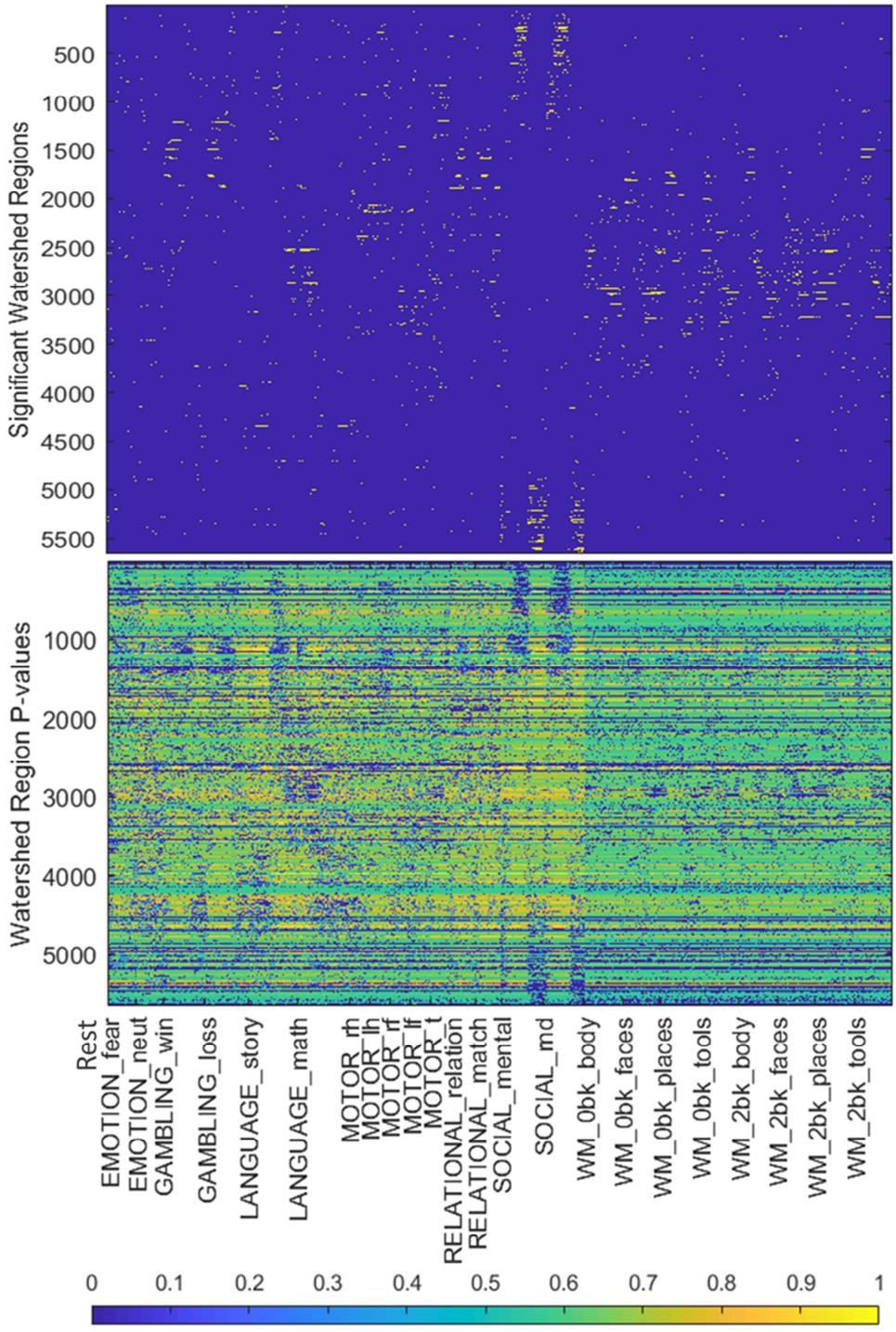


Supplemental figure S4 displays the descriptive statistics from bootstrap, between-scan, structural similarity index testing as a box-stem plot.


A
all
0bk_body
0bk_cor
0bk_err
0bk_faces
0bk_places
0bk_tools
2bk_body
2bk_cor
2bk_err
2bk_faces
2bk_nlr
2bk_places
2bk_tools
all_bk_cor
all_bk_err
cue
error
fear
lf
lh
loss
loss_event
match
math
mental
mental_resp
neut
neut_event
other_resp
present_math
present_story
question_math
question_story
relation
response_math
response_story
rest
rf
rh
rnd
story
t
win
win_event



Supplemental figure S5 displays and compares 2D Euclidean embeddings of instantaneous brain states, segmented by within-scan events. Part A displays the density of each scan's embedded points. Part B displays the mean structural similarity index (SSIM) from 50 bootstrap comparisons, with 2500 points per comparison. Asterisks indicate between-group comparisons whose SSIM was not significantly less than either within-group comparison.



Supplemental figure S6 addresses the likelihood that each of the experimental condition results in any of the brain states. Conditions are aligned task blocks. The resting state is taken as a single condition. Watershed regions are from a fine-grained density map, and resulted in ~5000 regions. A z-statistic was calculated across all possible affinities. The null distribution was generated by randomly permuting the labels associated with each point 100 times. The top plot highlights statistically significant affinities (after Bonferroni correction). The bottom plot displays each comparison's p-value.

Supplemental Movies

Movie 1 depicts mean brain states after segmenting the embedding using the point distribution of the resting state.

Movie 2 depicts mean brain states after segmenting the embedding using the point distribution of all tasks.

Movie 3 depicts the temporal evolution of volunteer brain states through the embedded state space during the *mental* contrast of the SOCIAL task.

Movie 4 depicts the temporal evolution of volunteer brain states through the embedded state space during the *random* contrast of the SOCIAL task.

Movie 5 depicts the temporal evolution of group-level volunteer brain states through the embedded state space during the *random* contrast of the SOCIAL task.