Comparing Functional Connectivity Matrices: A Geometry-Aware Approach applied to Participant Identification

Manasij Venkatesh[a], Joseph Jaja[a], Luiz pessoa[a,b]

[a]*Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA*
[b]*Department of Psychology and Maryland Neuroimaging Center, University of Maryland, College Park, MD, USA*

Abstract

Understanding the correlation structure associated with multiple brain measurements informs about potential "functional groupings" and network organization. The correlation structure can be conveniently captured in a matrix format that summarizes the relationships among a set of brain measurements involving two regions, for example. Such functional connectivity matrix is an important component of many types of investigation focusing on network-level properties of the brain, including clustering brain states, characterizing dynamic functional states, performing participant identification (so-called "fingerprinting), understanding how tasks reconfigure brain networks, and inter-subject correlation analysis. In these investigations, some notion of proximity or similarity of functional connectivity matrices is employed, such as their Euclidean distance or Pearson correlation (by correlating the matrix entries). Here, we propose the use of a *geodesic distance* metric that reflects the underlying non-Euclidean geometry of functional correlation matrices. The approach is evaluated in the context of participant identification (fingerprinting): given a participant's functional connectivity matrix based on resting-state or task data, how effectively can the participant be identified? Using geodesic distance, identification accuracy was over 95% on resting-state data, and exceeded the Pearson correlation approach by 20%. For whole-cortex regions, accuracy improved on a range of tasks by between 2% and as much as 20%. We also investigated identification using pairs of subnetworks (say, dorsal attention plus default mode), and particular combinations improved accuracy over whole-cortex participant identification by over 10%. The geodesic distance also outperformed Pearson correlation when the former employed a fourth of the data as the latter. Finally, we suggest that low-dimensional distance visualizations based on the geodesic approach help uncover the geometry of task functional connectivity in relation to that during resting-state. We propose that the use of the geodesic distance is an effective way to compare the correlation structure of the brain across a broad range of studies.

1. Introduction

Measurements of brain activity are acquired across multiple sensors or spatial locations, such as those obtained by electro/magneto-encephalography, electrophysiology recordings, calcium imaging, or functional magnetic resonance imaging (fMRI) data. Understanding the correlation structure associated with multiple brain measurements is a central goal in neuroscience, as it informs about potential "functional groupings" and network structure [21, 26]. The correlation structure can be conveniently captured in a matrix format that captures the relationships among a set of brain measurements. For example, in the case of fMRI, each entry of the matrix might contain an estimate of the *functional connectivity* (FC) between regions i and j, typically computed as the correlation between the time series data of the two regions in question.

In recent years, the FC matrix has become an important component of many types of investigation focusing on network-level properties of the brain, particularly in fMRI. For example, it has been used to cluster brain states [2], characterize dynamic functional states [14], perform participant identification [10], and understand how tasks reconfigure brain networks [24]. In these applications, some notion of proximity or similarity of FC matrices is employed (Figure 1A). How should similarity be gauged? An intuitive approach is to "unroll" the FC matrix into a vector and compute the Pearson correlation between the matrices themselves. Thus if, say, two brain states captured by FC matrices are similar (for example, during two similar perceptual conditions), their matrices would be (relatively) highly correlated. Indeed, the correlation approach has yielded impressive results, such as successfully identifying a participant out of a large group of participants based on FC matrix similarity (fingerprinting; [10, 9]). Related approaches include computing the Euclidean (L^2) distance between the vectorized matrices [22], or using the so-called Manhattan (L^1) distance [2].

FC matrices computed by Pearson correlating time series data are objects that lie on a non-linear surface (technically known as a manifold) called the *positive semidefinite cone*: their geometry is non-Euclidean. Accordingly, distances between Pearson FC matrices must be measured along the surface of the cone (Figure 1B). In addition, FC matrices are often high dimensional, and the proximity measure adopted is critical since noisy dimensions can contribute substantially to the measure [1].

Email address: manasij@umd.edu (Manasij Venkatesh)

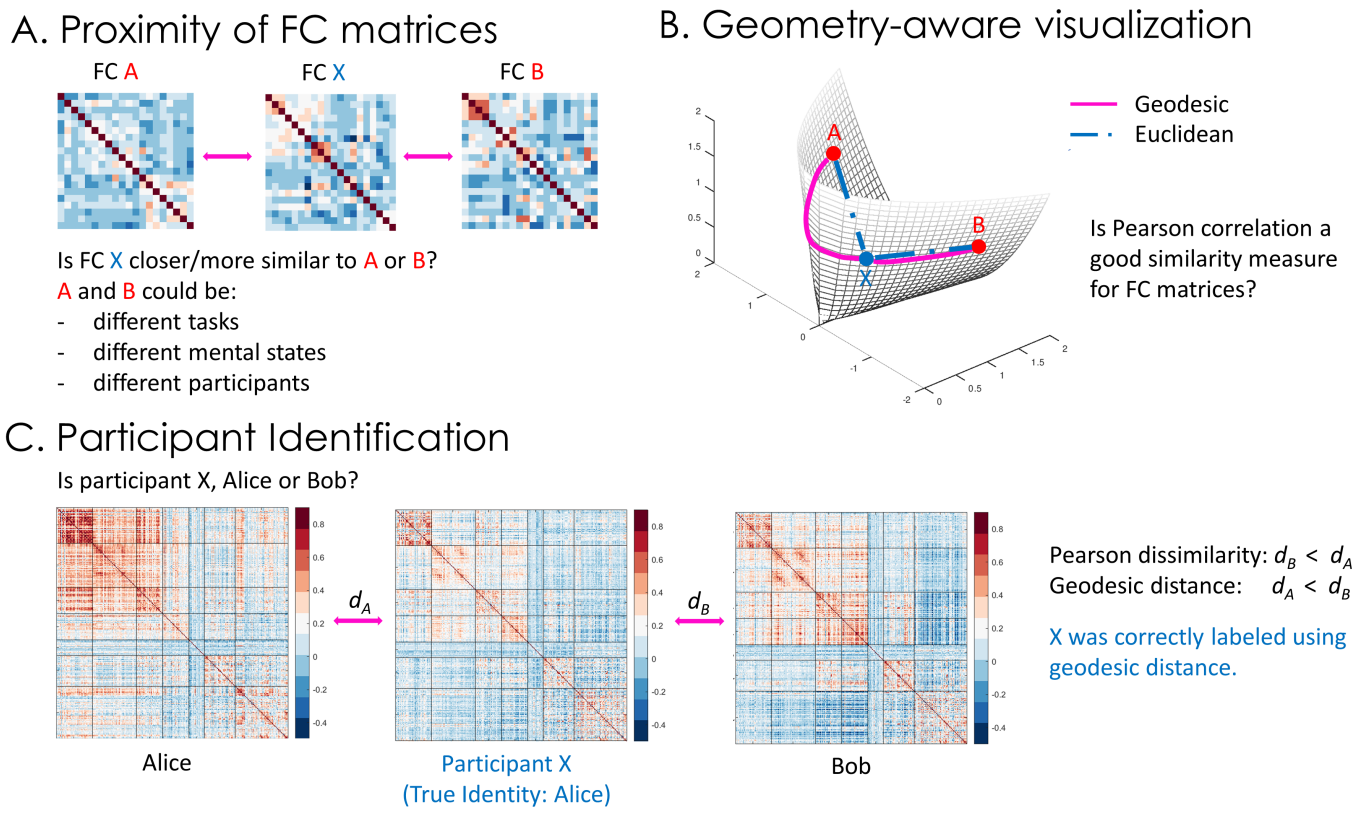


Fig. 1: Functional connectivity matrices and their underlying geometry. A. Similarity of functional connectivity (FC) matrices. Is the FC matrix X more similar to A or B? This question arises when the goal is to determine the task being being performed, the mental state, or the participant. B. Illustration of geodesic distance (magenta, solid) and Euclidean distance (blue, dashed) on the so-called positive semidefinite cone. The geodesic and Euclidean distances between two points can differ substantially. C. Is X, Alice or Bob? Equivalently, is the FC X more similar to that of Alice or Bob? Identification involves mapping an unknown participant's data to one of the participants in the database (only two in this case). In this example, X is correctly labeled as Alice using geodesic distance, but incorrectly labeled as Bob using Pearson dissimilarity.

In the present paper, we characterized the advantages of using a *geodesic* proximity measure between FC matrices. We apply the approach to the problem of participant identification: Given resting-state or task data, is it possible to determine a participant from her FC matrix [10]? We show that using the geodesic distance, a non-Euclidean distance metric that considers the manifold on which the data lies, improves participant identification compared to a similarity measure based on Pearson correlation (Figure 1C). The improvement is shown to be non-trivial and consistent across resting-state and task conditions.

We also investigate how distances between high-dimensional FC matrices can be effectively visualized in low-dimensional spaces. Such visualization reflected identification accuracy based on the full-dimensional data, and thus retained important distance information. We suggest that visualization in lower dimensions aids in understanding the geometry of task FC structure in relation to resting-state FC.

2. Methods

2.1. Human Connectome Project Data

We utilized data from $N = 100$ unrelated participants from the Human Connectome Project (HCP) of the 1200-participant release [7]. Data from resting-state and seven tasks were employed: emotion processing (EM), gambling (GB), language (LG), motor (MT), relational processing (RL), social cognition (SO), and working memory (WM). Throughout the paper, we refer to resting-state plus the tasks as *conditions*. For a description of the tasks and scan parameters, see [4]. Data were collected with a repetition time (TR) of 720 ms.

During each run, stimuli were presented in separate blocks often interleaved with fixation blocks. Some task runs also contained cues. To retain only task-related segments of the run, extraneous segments were trimmed. To account for hemodynamic lag, the first four TRs of the block were not used, and the first four TRs following the end of the block were used. Emotional processing, working memory, and motor tasks contained 3-second cues at block onset. Accordingly, to account for the cue response and the hemodynamic lag, data from 12 seconds after the cue onset to 3 seconds after the end of the block were used. Time course length for each condition before and after

trimming is provided in Table 1. Note that trimming the fixation periods is important in characterizing participant identification from task data, because fixation periods behave much like "mini resting periods" that can potentially provide information regarding the participant.

2.2. Preprocessing

Task data were part of the "minimally preprocessed" release, which included gradient unwarping, fieldmap-based EPI distortion correction, brain-boundary-based registration of EPI to structural T1-weighted scan, non-linear registration, and intensity normalization [12]. Cortical data were mapped to a surface representation and utilized here. In addition, we regressed out 12 motion-related variables (6 translation parameters and their derivatives) and low frequency signal changes using the 3dDeconvolve program of the AFNI package [6] (with the `ortvec` and `polort` options). Resting-state also followed the so-called minimal preprocessing pipeline, in addition to denoising using ICA-FIX [25], as provided with the data distribution. Cortical data were mapped to a surface representation. No further preprocessing was performed.

2.3. Regions of interest and organization into subnetworks

For simplicity, we focused on cortical regions of interest (ROIs) only. We used the local-global Schaefer cortical parcellations that divide the cortex into 300 ROIs [23] (throughout the text, we refer to it as "whole-cortex"). A summary ROI-level time series was obtained by averaging signals within the region. We then used the Yeo 7-network parcellation to group the ROIs into 7 subnetworks known as `visual`, `somatomotor`, `dorsal attention`, `ventral attention`, `limbic`, `frontoparietal`, and `default mode` [29]. The number of ROIs within each of the subnetworks is provided in Table 2. The ROIs and the grouping into 7 networks is shown in Figure S1. Some of the effects of varying the number of ROIs are described in the supplemental material (Section S1).

2.4. Functional connectivity

Functional connectivity was computed by Pearson correlating time series data between every pair of ROIs, resulting in 300×300 FC matrices. A symmetric matrix S that satisfies $y'Sy \geq 0$ (where y' is the transpose of y) for any non-zero vector y is said to be positive semidefinite and has eigenvalues greater than or equal to zero. After normalizing the time series of each ROI to unit variance, let $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,300})$ be the vector of activations of all ROIs at time t for $t = 1, 2, \ldots, T$. If we denote the mean across as $\bar{x}$, the covariance matrix is given by

$$Q = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(x_t - \bar{x})'. \tag{1}$$

Note that the (i, j) entry of Q is simply the Pearson correlation coefficient between the time series of regions i and j. For any non-zero vector y of dimension 300,

$$\begin{aligned} y'Qy &= y'\left(\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(x_t - \bar{x})'\right)y \\ &= \frac{1}{T}\sum_{t=1}^{T} y'(x_t - \bar{x})(x_t - \bar{x})'y \\ &= \frac{1}{T}\sum_{t=1}^{T}((x_t - \bar{x})'y)^2 \geq 0. \end{aligned}$$

Thus, covariance matrices are positive semidefinite.

If Q_1 and Q_2 are two FC matrices, it can be easily shown following the steps above that $\alpha Q_1 + \beta Q_2$ is also positive semidefinite for $\alpha, \beta > 0$. Thus, the set of all positive semidefinite matrices lie on a cone referred to as the *positive semidefinite cone* [5].

2.5. Geometry of functional connectivity matrices

Pearson correlation is often used to characterize the similarity of FC matrices. However, as correlation matrices lie on a non-linear space, a natural approach is to compute *geodesic distances* between FC matrices to quantify their distance. The geodesic distance between two points on the positive semidefinite cone, and thus between two FC matrices Q_1 and Q_2, is the shortest path between them along the manifold [20]. There exists only one geodesic path joining two such points.

For two functional connectivity matrices of size $n \times n$ (here, $n = 300$ ROIs), their geodesic distance can be computed as proposed in [20]:

$$d_G(Q_1, Q_2) = \sqrt{\text{trace}(\log^2(Q_1^{-\frac{1}{2}} Q_2 Q_1^{-\frac{1}{2}}))}, \tag{2}$$

where the matrix log operator is used here. Note that this definition assumes that the matrix Q_1 is invertible; when this was not the case the identity matrix, I, was added as a perturbation matrix to both Q_1 and Q_2 to ensure that all eigenvalues were greater than 0 (see Section S2). If $Q = Q_1^{-\frac{1}{2}} Q_2 Q_1^{-\frac{1}{2}}$, and λ_i for $i = 1$ to n are the n eigenvalues ≥ 0 of Q, the geodesic distance is simply

$$d_G(Q_1, Q_2) = \sqrt{\sum_{i=1}^{n}(\log(\lambda_i))^2}. \tag{3}$$

From (3), it is clear that $d_G \geq 0$. In addition, if $d_G = 0$, then $Q_1 = Q_2$. It is also easy to verify that $d_G(Q_1, Q_2) = d_G(Q_2, Q_1)$. Thus, the geodesic distance applied to matrices meets the criteria of a *metric* (see [11]).

If q_1 and q_2 are vectors obtained by stacking the columns of Q_1 and Q_2, respectively, Pearson dissimilarity between the two matrices is defined as

$$d_P(Q_1, Q_2) = \frac{1 - \text{corr}(q_1, q_2)}{2}, \tag{4}$$

where the corr function is the Pearson correlation coefficient. Pearson dissimilarity ranges between 0 and 1 and is *not* a formal metric because it does not satisfy the triangular inequality [27].

Condition	REST	EM	GB	LG	MT	RL	SO	WM
Frames per full run	1200	176	253	316	284	232	274	405
Frames per trimmed run	1200	141	156	295-305	170	138	160	312

Table 1: Number of frames per run (in samples) before and after trimming fixation periods.

Subnetwork	Visual	SomatoMotor	Dorsal Attention	Ventral Attention	Limbic	FrontoParietal	Default
Number of ROIs							
100	17	14	15	12	5	13	24
200	29	35	26	22	12	30	46
300	47	57	34	34	20	40	68
400	61	77	46	47	26	52	91

Table 2: Number of ROIs in each subnetwork

2.6. Participant identification

Identification involves mapping an unknown participant's data to one of the participants in the database. Since each task in the HCP data contains 2 runs for every participant, we used one run as training data (that is, to form the database) and the other run for testing. Identification was performed on each condition (resting-state or task) separately.

Participant identification is equivalent to N-class classification where the objective is to label an individual's FC matrix in the test data to one of the N participants in the training data. To do so, we used a 1-Nearest Neighbor approach [10]: An FC matrix in the test data is labeled with the participant identity of the FC that is most similar to it in the training data. Suppose Q_x^{test} is an unknown participant's FC matrix. Then

$$\text{label}(x) = \arg\min_{i=1}^{N} d(Q_i^{\text{train}}, Q_x^{\text{test}}), \tag{5}$$

where Q_i^{train} is the ith participant's FC matrix in the training data and $d(\cdot,\cdot)$ is a distance or similarity measure. Here we compare the use of a geodesic distance metric to a Pearson dissimilarity measure.

2.6.1. Identification accuracy

Participant identification was performed using the first run as training data and the second run as testing data. For the N participants in the testing data, accuracy was defined as

$$\text{Accuracy} = \frac{\text{Number of correctly labeled participants}}{\text{Total number of participants}}.$$

Then, the roles of the training and testing data were reversed and accuracy was computed again. The reported identification accuracy was the mean of the two accuracy values.

2.7. Bootstrapping

For participant identification statistics, one must confront the non-independence between participants in the sample. Consider the following case. If two participants' FC matrices Q_A and Q_B are close to each other, B might be mislabeled as A. However, if A was not in the training database, it is conceivable that B would have been labeled correctly. Therefore, the *entire group* must be considered as the unit of interest; it is the group that determines if identification performance will be poor or good.

A convenient procedure to assess variability in identification performance is to use bootstrap resamples, with each resample comprising random draws with replacement of the urn containing the group of participants. Thus, a bootstrap resample is a proxy for a group of participants, and variability can be quantified by resampling it a large number of times.

More precisely, suppose a dataset of size N for a run is denoted by $\mathcal{D}$. Let $0 \le f_P(\mathcal{D}) \le 1$ and $0 \le f_G(\mathcal{D}) \le 1$ be the participant identification accuracy obtained using Pearson dissimilarity and the geodesic distance, respectively. Let $\mathcal{R}_j$ be a dataset also of size N obtained by resampling $\mathcal{D}$, with replacement, N times. Thus, $\mathcal{R}_j$ is a bootstrap resample of $\mathcal{D}$ and may contain duplicate entries. The accuracy difference on this bootstrap resample is given by $\delta(\mathcal{R}_j) = f_G(\mathcal{R}_j) - f_P(\mathcal{R}_j)$. Such difference score is computed for $M = 1000$ bootstrap resamples $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_M$ and the *mean* difference score, $\bar{\delta}$, is computed. This process (based on M resamples) provides exactly one mean difference score. The question of interest is as follows: How are *mean* difference scores distributed? (Note that this parallels the question of the distribution of the sample mean in the setting of the standard Central Limit Theorem.) Since the object of interest is the *mean* difference score, the procedure to determine a specific $\bar{\delta}$ is repeated $B = 1000$ times, resulting in $\{\bar{\delta}_1, \bar{\delta}_2, \ldots, \bar{\delta}_B\}$ (that is, B mean differences).

Reported p-values were computed as follows. Because accuracy differences are percentages, we initially applied a standard Fischer-z transformation to $\{\bar{\delta}_1, \bar{\delta}_2, \ldots, \bar{\delta}_B\}$ so that their distribution would be approximately normal. To test the null hypothesis $H_0 : \bar{\delta} = 0$, a one-sample t-test was then used.

2.7.1. Evaluating shorter data segments

To understand the effect of the length (or the number of frames) of the run, we truncated runs to smaller segments. For a particular segment length, 50 segments were obtained each of which had a unique, randomly-chosen starting point in the run.

The objective was to pick several segments of the same length without favoring those that started at the beginning of the scan. For each segment, 1000 bootstrap iterations were used to obtain a mean accuracy score.

2.8. Multidimensional scaling

Naturally, visualizing distances between FC matrices is not straightforward given their high dimensionality. Here, we used *non-metric multidimensional scaling* to visualize distances in three dimensions [15]. Whereas standard multidimensional scaling computes the Euclidean distance between the high-dimensional vectors of interest, non-metric multidimensional scaling takes as input any *dissimilary matrix* of the form

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,100} \\ d_{2,1} & d_{2,2} & \dots & d_{2,100} \\ \vdots & \vdots & \ddots & \vdots \\ d_{100,1} & d_{100,2} & \dots & d_{100,100} \end{bmatrix}$$

where $d_{i,j}$ is the "distance between the FC matrices i and j ($N = 100$ represents the number of participants). Here, either the geodesic distance or Pearson dissimilarity was used. Given D, non-metric multidimensional scaling finds a set of $\mathbb{R}^3$ vectors such that the Euclidean distance between these vectors preserves, to the extent possible, the high-dimensional distances:

$$\hat{d_{i,j}} = \|x_i - x_j\|_2^2 \approx d_{i,j}$$

where the vectors x are low dimensional. Thus, if $d_{i,j} = d(Q_i, Q_j)$ is the distance between two FC matrices Q_i and Q_j, and $\hat{d_{i,j}}$ is the distance in the lower-dimensional representation, the output (set of points) is produced by minimizing the *stress* function:

$$S = \sqrt{\frac{\sum_{i<j} (d_{i,j} - \hat{d_{i,j}})^2}{\sum_{i<j} d_{i,j}^2}}.$$

The optimal distances, $\hat{d_{i,j}}$, are obtained using a gradient descent approach that minimizes the stress. The MATLAB 2018a [17] implementation of *mdscale* with 1000 gradient descent iterations was used.

Note that the objective of using non-metric multidimensional scaling was to represent in a more intuitive manner the relationships between high-dimensional functional connectivity matrices. Thus, points in the lower-dimensional representation no longer lie on the positive semidefinite cone and *closeness* should be interpreted in the Euclidean sense (two points are close if their Euclidean distance is small). The visualizations, approximate as they are, are only provided to aid understanding, and are not part of the procedure to determine identification accuracy.

2.9. Note on p-values

As discussed by many others recently, we do not view "statistical significance dichotomous thresholds (for example, $p < 0.05$) as the ultimate criterion in deciding whether a result is "real" or not ([3, 18]). In any case, understanding variability and the unlikeliness of a result provides some information. Given that we compare geodesic distance to Pearson dissimilarity across conditions and other parameters, some form of correction for multiple comparisons is opportune. Accordingly, we provide the uncorrected p-value as well as the Bonferroni-corrected α level (which we call the "reference α") so that the reader can further gauge the "strength" of the finding. Again, we do not advocate using the Bonferroni-corrected α in a dichotomous fashion, but provide it as an additional "reference point for the reader.

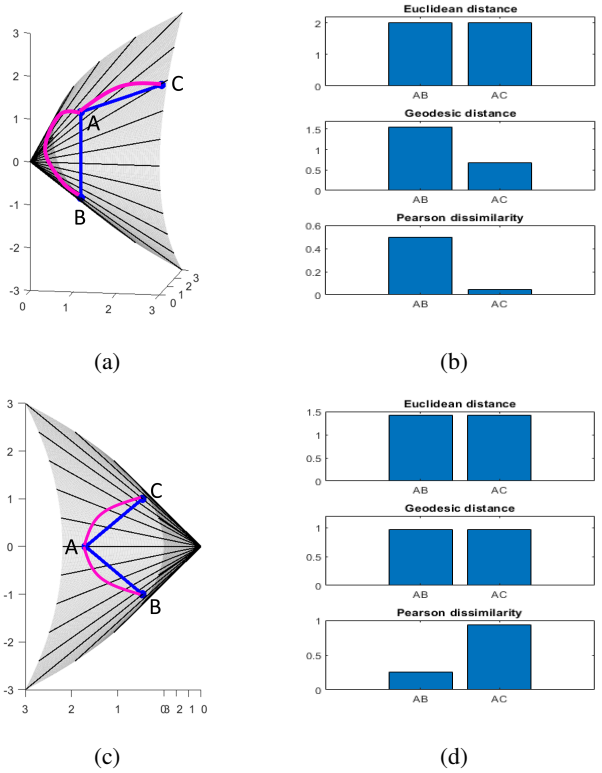


Fig. 2: Motivating functional connectivity geometry. (a) Identical Euclidean distance does not imply identical geodesic distance. (c) Identical geodesic distance can yield very different Pearson dissimilarity. (b, d) Comparison of distances/similarity AB and AC in (a) and (c), respectively.

3. Results

3.1. Motivation behind geodesic distance

We motivate the geodesic distance with simple examples from the space of 2×2 FC matrices. Since FC matrices are symmetric and positive semidefinite, they take the form

$$Q = \begin{bmatrix} x & y \\ y & z \end{bmatrix},$$

and satisfy $x \geq 0, y \geq 0$ and $xy - z^2 \geq 0$. Since the matrices have only three unique entries, all points that satisfy these equations can be plotted in three dimensions in Euclidean space, and form a *positive semidefinite cone* (Figure 1B).

In the first example, we considered three points on the cone (i.e., three 2×2 FC matrices) A, B and C such that B and C are equidistant from A in terms of the Euclidean distance (Figure 2a). If a tangent surface to the cone is drawn at A, the point C is much closer to the tangent surface than B. Thus the geodesic distance between A and C is larger than that between A and B (Fig. 2b). In this case, Pearson dissimilarity is capable of distinguishing the two distances.

To motivate why Pearson dissimilarity is problematic, consider that the Pearson correlation between two vectors is equivalent to the cosine of the angle between them after they have been "centered" individually (that is, the mean of each vector is subtracted from it) and normalized. Indeed, the computation of Pearson correlation eliminates the contribution of the signal mean, as can be readily seen in the following equation:

$$\mathrm{corr}(x, y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}},$$

where x and y are vectors. For FC matrices, such centering which is implicit in Pearson correlation alters the eigenvalues and the positive semidefiniteness of the matrix. Since the eigenvalues are the basis for computing geodesic distances, we see that Pearson correlation in fact distorts the evaluation of similarity between connectivity matrices (in relation to what is estimated with the geodesic distance).

In a second illustrative example (Figure 2c), we consider three points A, B and C on the cone such that B and C are symmetrically on either side of A. By symmetry, A is equidistant from B and C in terms of both the Euclidean distance and geodesic distance. However, Pearson dissimilarity between the two sets of points can be quite distinct. Suppose O is the origin and $\angle AOB = \angle AOC$ (where $\angle$ is the angle subtended between A and B). However, because Pearson correlation mean centers the vectors A and B, the correlation is related to mean-centered vector angles that can be quite different from the original ones (Figure 2d). In other words, if $\underline{A}, \underline{B}$, and $\underline{C}$ are vectors obtained by centering A, B, and C, in most cases $\angle \underline{A}O\underline{B} \neq \angle \underline{A}O\underline{C}$. The upshot is that measures of similarity based on Pearson correlation do not correspond to actual distances between functional connectivity matrices.

3.2. *Geodesic distance and participant identification*

Participant identification ($N = 100$) was performed on each condition (resting-state and tasks) using two measures: geodesic distance and Pearson dissimilarity (Methods 2.6). FC matrices obtained from one run were used as training data and matrices from the second run as testing data. Identification accuracy for each condition is shown in Figure 3 (accuracy based on chance would be 1%; see supplemental material for comparison with accuracy values reported previously [10]).

To assess the robustness of the results and for statistical comparisons between the two measures, identification was performed on bootstrap resamples. For each bootstrap resample, the difference between accuracy using geodesic distance and Pearson dissimilarity was computed. A one-sample two-tailed

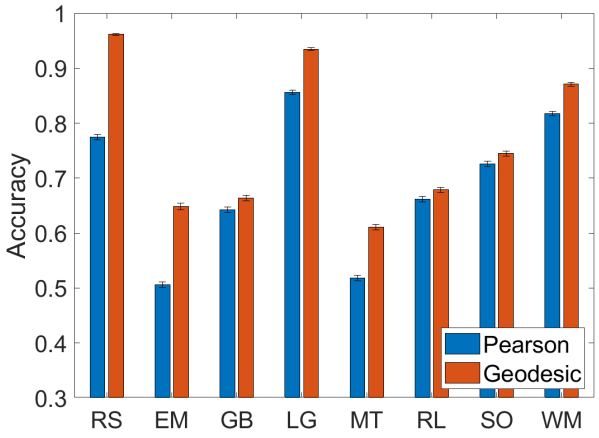


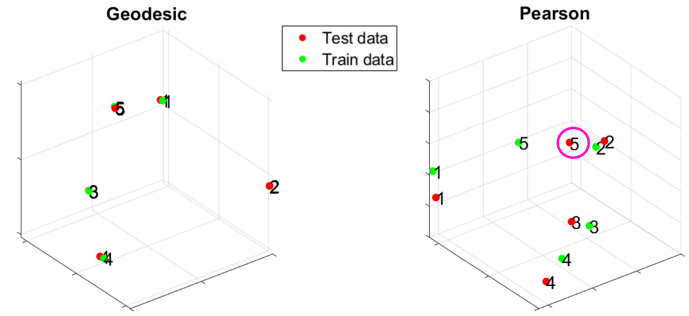
Fig. 3: Participant identification for the eight conditions using the geodesic distance and Pearson dissimilarity. Training and testing data were from the same condition. Accuracy improved using the geodesic distance on each condition. Error bars indicate standard error of the mean across bootstrap iterations. Abbreviations: EM, emotion processing; GB, gambling; LG, language; MT, motor; RL, relational processing; RS, resting-state; SO, social cognition; WM, working memory.

t-test was then used to assess the null hypothesis that the difference distribution had zero mean (Methods 2.7). For each condition, using the geodesic distance improved identification accuracy over Pearson dissimilarity ($p < 10^{-4}$ for all tasks; reference $\alpha = 0.05/8 = 0.00625$ given 8 conditions; Table S1 and Fig. S3). The mean improvement using geodesic distance was around 8%, ranging from 2% (*relational processing*) to as much as 19% (*resting-state*). For *resting-state* and the *language* conditions, the accuracy obtained using the geodesic distance was very hight and close to 95%.

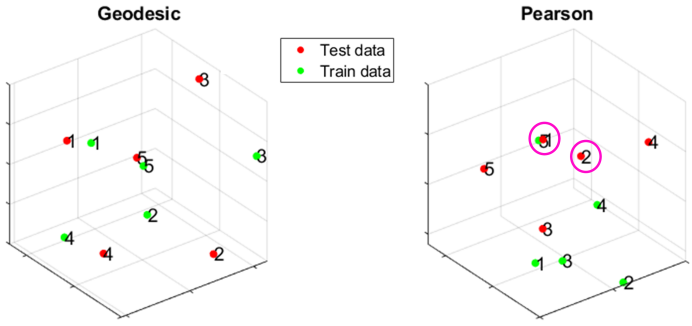
3.3. *Low-dimensional visualization of functional connectivity matrices*

Since FC matrices are high dimensional, multidimensional scaling was used to visualize the distances between them in three dimensions (Figure 4). For *resting-state*, FC matrices within-participant geodesic distances between training and testing were very small, whereas distances between different participants were considerably larger, consistent with the high identification accuracy. Visualization of Pearson dissimilarity revealed similar characteristics, but the ratios of within- to between-participant distances were not as large. In fact, using Pearson dissimilarity resulted in participant 5 being mislabeled as participant 2, for example.

For the *emotional processing* task, within-participant distances were *not* much smaller than between-participant distances even for the geodesic distance consistent with the lower accuracy on this task. However, all participants in the randomly chosen subset were still labeled correctly. Using Pearson dissimilarity, two participants were mislabeled. In general, using the geodesic distance resulted in more favorable ratios of within- to between-participant FC distances.



(a) Resting-state



(b) *Emotional processing*

Fig. 4: Visualization of geodesic distance and Pearson dissimilarity. Distance/similarity between high-dimensional functional connectivity matrices (300 × 300) was visualized in three dimensions using generalzied multi-dimensional scaling. Training data (green) and testing data (red) were selected from five random participants (numbers 1-5). Mislabeled participants are encircled in magenta. (A) *Resting-state*. (B) *Emotional processing* task. For *resting-state*, within-participant geodesic distances were very small relative to between-participant distances in the lower-dimensional representation (when numbers labeling the participants overlapped, only one of them is visible).

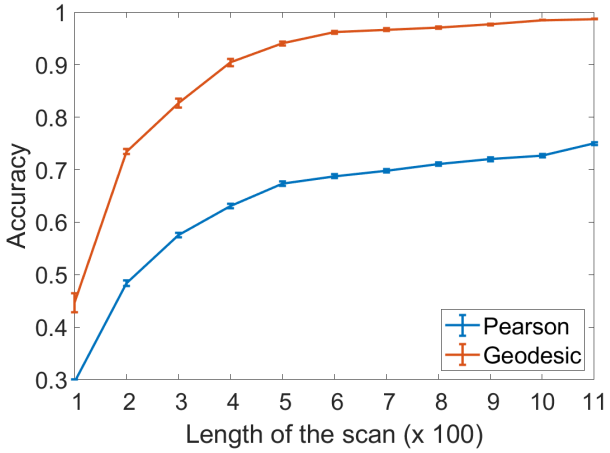


Fig. 5: Participant identification accuracy as a function of segment length for *resting-state* data. Accuracy using geodesic distance exceeded Pearson dissimilarity at each segment length (see text). Error bars indicate standard error of the mean across bootstrap iterations.

3.4. Identification accuracy and time course length: resting-state data

Since the length of the time course plays a key role in the quality of the estimate of the FC matrix [16, 30], we sought to characterize its effect on participant identification. Because *resting-state* data had the longest time course (1200 TRs), shorter segments varying from 100 to 1100 TRs (in steps of 100) were extracted. Accuracy improved with length for both measures (Figure 5). Accuracy using the geodesic distance was higher than Pearson dissimilarity for segment lengths greater than 200 TRs ($p < 10^{-4}$; reference $\alpha = 0.05/11 = 0.0045$ given 11 segment lengths; Fig. S4). For segment length of 100 TRs, accuracy using geodesic distance was still higher than Pearson dissimilarity (but $p = 0.051$). Notably, the geodesic distance, with segment lengths as short as 300 TRs, outperformed the best accuracy using Pearson dissimilarity which was obtained with the full time course (four times more data; $p < 10^{-4}$; reference $\alpha = 0.05/11 = 0.0045$ given 11 segment lengths).

3.5. Identification accuracy and time course length: task data

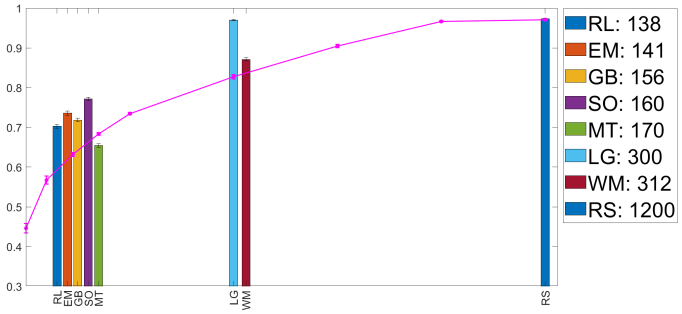
Although accuracy increased with segment length for *resting state*, length did not predict performance straightforwardly (Figure 6A). In particular, *working memory* and *language* tasks had comparable time course lengths, but identification accuracy differed by as much as 10%. To probe this issue further, runs were trimmed so that they all had the same duration (138 TRs, which was the length of the shortest task; for conditions with more data, this target length was obtained by deleting time points at the beginning and end of the data segment, thereby retaining the middle part).

With time course length equated, accuracy still varied considerably across tasks (Figure 6B). Accuracy obtained using the geodesic distance exceeded that of Pearson dissimilarity for all conditions except the *gambling* task ($p = 1$ for *gambling*, $p < 10^{-4}$ for all other tasks; reference $\alpha = 0.05/8 = 0.00625$ given 8 conditions; Fig S5). Notably, although *resting-state* had the highest identification accuracy when the entire time course was used, it had the lowest identification accuracy when length was equated across conditions.

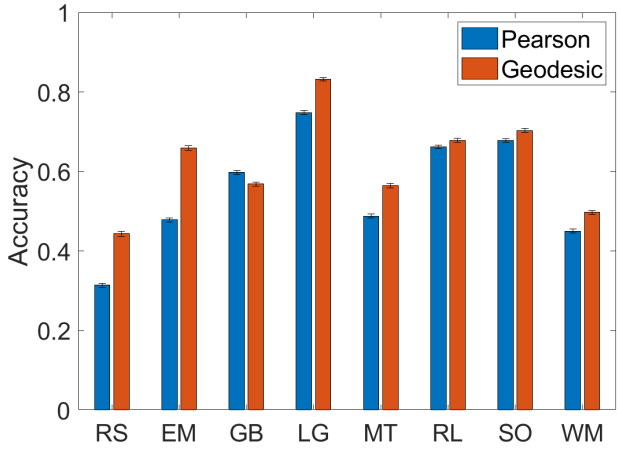
3.6. Brain subnetworks and participant identification

Particular brain subnetworks are known to be engaged more prominently, as well as exhibit enhanced functional connectivity, during particular tasks [21]. To evaluate performance based on subsets of regions, ROIs were grouped into seven subnetworks (Methods 2.3). Was the best subnetwork for identification dependent on condition? Data for all conditions were trimmed so that they had the same length (138 TRs; the `limbic` subnetwork was excluded because identification accuracy was less than 10% across conditions).

Using geodesic distance improved the accuracy across most conditions for most subnetworks (Figure 7A). In particular, for the `visual, dorsal attention, frontoparietal` and `default mode` subnetworks, accuracy was comparable to that obtained with the whole cortex. For example, the `default mode` subnetwork produced accuracy over 90% for the *language* task. The `frontoparietal` performance on *resting-state* and *emotion processing* was close to 80%. Further inspection of Figure 7A revealed additional features of con-



(a)



(b)

Fig. 6: Participant identification and time course length. (A) Accuracy based on geodesic distance for *resting-state* and task conditions (time course length in TRs in the inset). The magenta curve shows the accuracy for *resting-state* data trimmed to segment lengths shorter and longer than those of task data (lengths from left to right: 100, 125, 145, 170, 200, 300, 600, 900, and 1200 TRs). (B) Accuracy when data was trimmed such that all conditions had the same time course length (138 TRs). Error bars indicate standard error of the mean across bootstrap iterations. Abbreviations: EM, emotion processing; GB, gambling; LG, language; MT, motor; RL, relational processing; RS, resting-state; SO, social cognition; WM, working memory.

dition/subnetwork combinations. For example, the `visual` subnetwork was not very suitable for identification based on *resting-state* data. Not surprisingly, the `default mode` subnetwork performed well with *resting-state* data. Interestingly, the `frontoparietal` subnetwork performed nearly as well with *resting-state* data, too. These two subnetworks obtained even higher identification accuracy during the *language* task.

To further evaluate performance of subnetworks, identification accuracy was averaged across conditions (Figure 7B). By using the geodesic distance, accuracy improved substantially, with several subnetworks improving by over 20%. Except for the `somatomotor` subnetwork, using the geodesic distance resulted in improved performance ($p = 0.996$ for `somatomotor`, $p < 10^{-5}$ for all other subnetworks; reference $\alpha = 0.05/7 = 0.0071$ given 7 subnetworks; see Fig. S6 for bootstrap distributions). The highest mean accuracies were observed in the `visual`, `dorsal attention`, `frontoparietal`, and `default mode` networks for both geodesic and Pearson measures, indicating that some subnetworks are more suitable than others for participant identification.

Figure 8 displays geodesic identification accuracy for each condition as a function of subnetwork size. Whereas the smallest subnetwork (`limbic`) performed poorly for all conditions, accuracy did not always increase with size. For example, the `dorsal attention` and `ventral attention` subnetworks have the same size, but the former produced considerably higher accuracy on each condition ($p < 10^{-12}$ for all conditions; reference $\alpha = 0.05/8 = 0.00625$ given the 8 conditions; see Fig. S7 for bootstrap distributions). Across conditions, the `dorsal attention` improved over the same-sized `ventral attention` by over 20%. Of note, the `somatomotor` subnetwork was larger than all but the `default mode` subnetwork, but it produced relatively low identification accuracy; at the same time, the largest subnetwork (`default mode`), was associated with consistently high accuracy across conditions. Finally, no single subnetwork exhibited the highest accuracy for all conditions. In fact, performance varied across conditions, but also varied in particular ways across subnetworks for each condition. Notably, the `visual`, textttdorsal attention, `frontoparietal`, and `default mode` subnetworks performed consistently well. Similar trends were observed for the Pearson dissimilarity measure but overall accuracy levels were lower (Figure S12).

3.7. Combining subnetworks improved identification accuracy

As described, subnetworks had comparable (and sometimes higher) identification accuracy than whole-cortex performance, although subnetworks were associated with much smaller matrices, of course. Could particular subnetworks be combined to further improve identification? We tested this possibility by targeting two subnetworks that exhibited high performance overall, namely `frontoparietal` and `default mode` (see Figure 7B). The combined network included all within-network functional connections of course, but also all between-network links (for example, a functional connection between a region of the `frontoparietal` network and a region of the `default mode` network). Accuracy using geodesic distance was superior to Pearson dissimilarity (Figure 9; $p < 10^{-15}$ for all conditions; reference $\alpha = 0.05/8 = 0.00625$ given 8 conditions; Figure S8).

Using geodesic distance, the combined subnetwork also outperformed both the individual subnetworks on all conditions except the *language* task ($p = 0.24$ for the *language* task, $p < 10^{-12}$ for all other conditions; $\alpha = 0.05/16 = 0.003125$ given 8 conditions and comparisons with two subnetworks; see Figures S9-S10 for bootstrap distributions). In addition, for the geodesic distance, the combined subnetworks exhibited higher accuracy than whole-cortex FC matrices ($p < 10^{-12}$ for all conditions; $\alpha = 0.05/8 = 0.00625$ given 8 conditions; see Fig. S11 for bootstrap distributions) although the number of ROIs in the combined subnetwork (108) was nearly a third as those in the cortex (300). Clearly, the improvement in accuracy was not a simple consequence of increased size, but resulted from improved identity characterization.

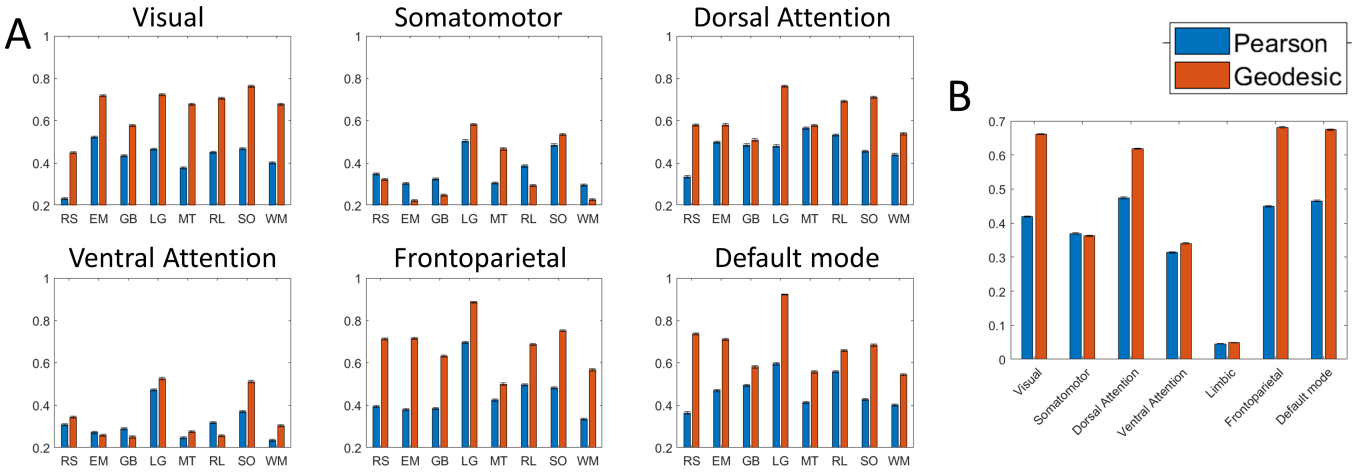


Fig. 7: (a) Participant identification accuracy using subnetworks. Runs were trimmed such that all conditions had the same time course length. Some subnetworks were more suitable than others for identifying individual differences. The use of geodesic distance showed considerable improvements in accuracy for most subnetworks. (b) Across subnetworks, average participant identification accuracy is displayed. The geodesic distance substantially improved identification accuracy. Error bars indicate standard error of mean across bootstrap iterations. Abbreviations: EM, emotion processing; GB, gambling; LG, language; MT, motor; RL, relational processing; RS, resting-state; SO, social cognition; WM, working memory.

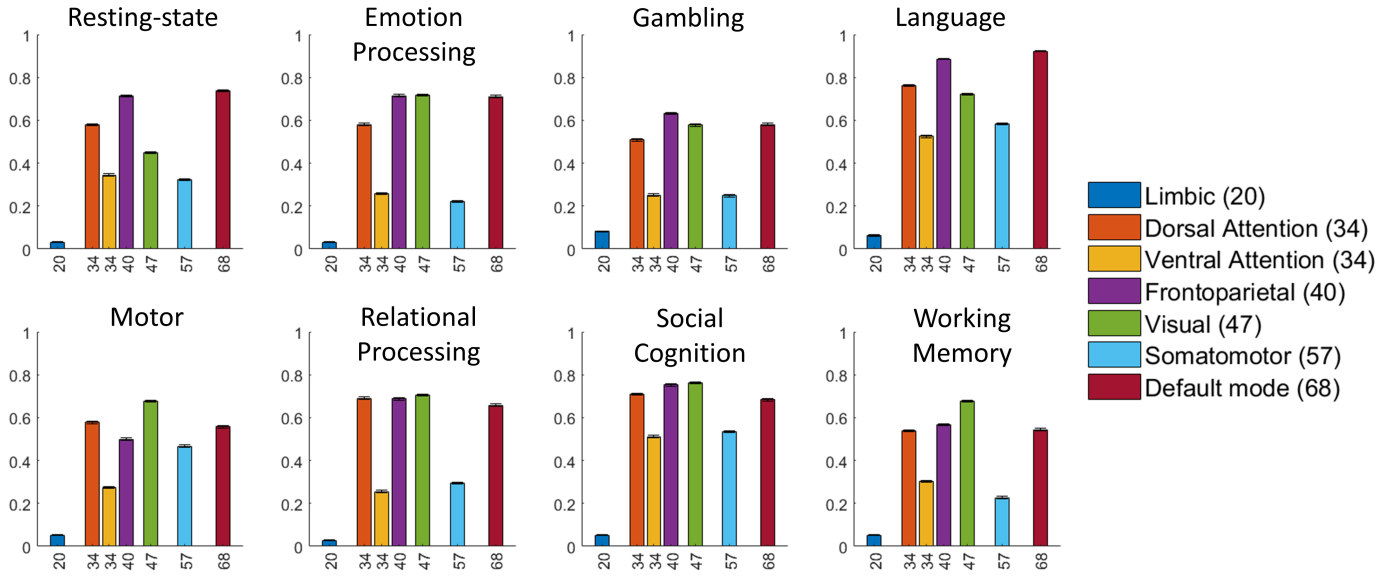


Fig. 8: Participant identification accuracy plotted against subnetwork size for each condition (geodesic distance). The size of the subnetwork (the number of ROIs) is also indicated in the inset. The error bars represent standard error of the mean across bootstrap iterations.

3.8. Information transfer between conditions

In the previous sections, training and testing data were based on the same condition. Here, we sought to understand if participants could be identified if the training and testing data were obtained from different conditions; for example, identifying a participant performing a *working memory* task when the training used *resting-state* data. Time series length was not equated across conditions because our goal was to evaluate how transferable identity-related information was between pairs of conditions. Accordingly, we did not want to potentially degrade FC information by using shorter data segments. Identification was performed on the combined `default-plus-frontoparietal` network, which as discussed performed well across conditions (Figure 9).

Results for both geodesic distance and Pearson dissimilary are displayed in Figure 10. Whereas Pearson dissimilarity was useful in identifying participants when they performed the same task (within-conditions, diagonal entries), performance deteriorated when the training and test data originated from different tasks. Notably, across-condition identification was considerably higher with the geodesic distance, and this enhancement was rather striking when the training data was from *resting-state*, and to some extent based on the *language* and *working memory* tasks. For example, testing *working memory* data based on training with *resting-state* data yielded 99% accuracy,

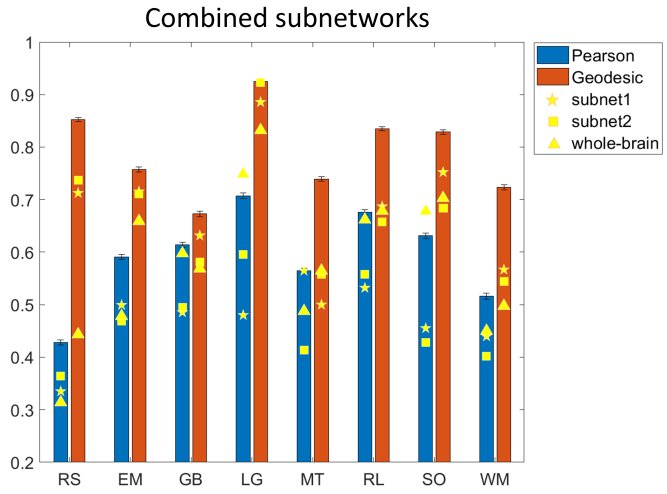


Fig. 9: Participant identification accuracy by combining subnetworks. For the geodesic distance, the `frontoparietal` (`subnet1`) and `default mode` (`subnet2`) subnetworks were combined. For the Pearson dissimilarity measure, the `dorsal attention` (`subnet1`) and `default mode` (`subnet2`) subnetworks were combined (the top two subnetworks based on mean accuracy across conditions for this measure). Abbreviations: EM, emotion processing; GB, gambling; LG, language; MT, motor; RL, relational processing; RS, resting-state; SO, social cognition; WM, working memory.

which intriguingly was even better than when training with *working memory* itself (accuracy: 94.4%, $p < 10^{-4}$). On average, training with *resting-state* yielded 87.7% accuracy when testing on *other* conditions (see the "column mean" in Figure 10). The present results indicate that the geometry of FC is especially important for across-task identification (see Discussion).

Because in this section time course length was not equated across conditions, we note that those with longer lengths aided across-task identification. Accordingly, transfer might particularly benefit from employing training sets with longer data segments. Nevertheless, future research should also evaluate transfer effects when longer data segments are available for a wider range of tasks (for example, $\geq$ 300 TRs) so as to characterize their transfer potential.

3.9. FC geometry of task and resting-state data

As some conditions yielded high identification accuracy when training and testing were based on different conditions, we sought to visualize distance/dissimilarity in a lower dimensional space via multidimensioanal scaling. Figure 11(a) displays the low-dimensional representation of the distances/dissimilarities for a set of randomly chosen participants when *resting-state* was employed for training data and working memory for testing (untrimmed data). Based on the geodesic distance, *resting-state* FC matrices were relatively close together to one another; in contrast, *working memory* FC matrices were further "spread out". Intriguingly, such geometry allowed for the separation of FCs based on participant identity. To see this, consider the panels in Figure 11(b), which show participant-level distances. In contrast, using Pearson dissimilarity, the geometry did not allow accurate participant identity. In fact, nearly all participants in this illustrative sample were misidentified.

The results in Figure 11(a) prompted us to investigate, in an exploratory fashion, distance/dissimilarity between conditions, specifically, *resting-state*, *motor*, and *language* (Figure 12. Intriguingly, the geometry of distances was quite different when geodesic distances were used compared to Pearson dissimilarity. These observations suggest that when FC matrices are used for *task classification* (not identification as done here), different algorithms may be more suited for this aim. For example, non-linear radial basis functions might function better for the geodesic case, and linear classifiers for Pearson dissimilarity. Although a fuller investigation of this issue is beyond the scope of the present paper, we believe this is a fruitful direction. Furthermore, the analysis of functional connectivity of mental states should take into account participant-related information since it plays a potentially dominant contribution in the identification of mental states [28]

4. Discussion

In this paper, we investigated participant identification based on FC matrices from fMRI data by employing geometry-aware methods. Correlation matrices are objects that lie on non-linear surfaces, and thereby benefit from non-Euclidean distance measures. Indeed, we show that using the geodesic distance improved participant identification, at times by as much as 20%. Further, low-dimensional visualization based on geodesic distance contributes to understanding how FC geometry affects identification.

4.1. Factors influencing participant identification

Scan duration determines the amount of data used to estimate FC matrices, and played a key role in identification accuracy. For *resting-state* data, accuracy improved with time course length and was close to 95% when the entire data were employed (1200 TRs), but fell to under 50% when trimmed to under 150 TRs. The steep drop is possibly due to the unconstrained nature of *resting-state* data, and reveals that longer data segments are required to more robustly identify functional connectivity patterns that are unique to individuals. Notably, using the geodesic distance resulted in higher accuracy than Pearson dissimilarity even when, say, only a fourth of the data were employed for FC estimation. Thus, a more suitable geometry is particularly appealing when data-limited scenarios are envisioned.

When time course length was trimmed to the same duration, identification accuracy still varied across scanning conditions. The *resting-state* condition resulted in the lowest accuracy. With the data trimmed to the minimum amount of data, the *language* task exhibited over 80% accuracy. Accuracy of all task conditions exceeded 50%, with four of them exceeding 60%. Thus, even with rather limited amounts of data identifying the participant was feasible. In addition, we observed considerable variability is performance across conditions, consistent with previous literature suggesting that brain states can be manipulated to emphasize individual differences in FC [9].

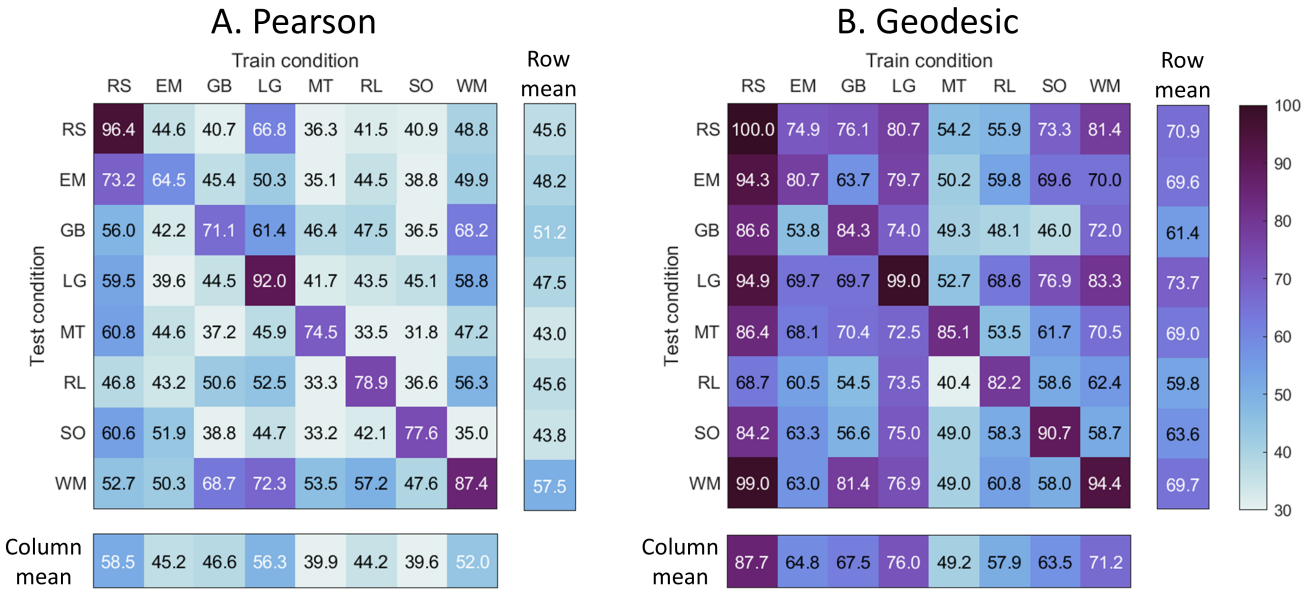


Fig. 10: Participant identification accuracy when the training and testing data were based on different conditions. The combined network containing the `frontoparietal` and `default mode` subnetworks was employed. The mean accuracy for each train and test condition is also indicated. For example, when *resting-state* is used as training data, the column mean is computed as the accuracy across all other conditions (i.e., except *resting-state* itself). The row means are computed in a similar fashion by excluding the diagonal term. Abbreviations: EM, emotion processing; GB, gambling; LG, language; MT, motor; RL, relational processing; RS, resting-state; SO, social cognition; WM, working memory.

Thus far, we have discussed findings based on whole-cortex FC matrices (300 ROIs were employed). We reasoned that particular subsets of regions potentially might be more informative than others. To evaluate this possibility identification was applied to *resting-state* and task conditions separately for each individual subnetwork of the Yeo parcellation ([29]). The FC matrices employed were therefore relatively small (the number of ROIs ranged from 20-68). Four subnetworks (`vision`, `dorsal-attention`, `control`, `default`) stood out as consistently exhibiting the highest levels of performance. The average accuracy across conditions approach 70% for the four networks. Intriguingly, accuracy for the *language* task based on the `control` and `default` subnetworks exceeded that observed with the whole cortex. Whereas subnetwork size might contribute to its ability to identify participants, it is clearly not the driving factor. For example, the `dorsal-attention` and the `ventral-attention` networks had the same number of ROIs, but the former outperformed the latter consistently (on average by over 30%).

To further explore subnetwork contributions we also combined the two that displayed the highest individual accuracy (`control` and `default`) into a single network. Remarkably, the combined network always numerically outperformed the individual subnetworks, and indeed the entire cortex. Although it was beyond the scope of the present study, it would be valuable to investigate in future studies factors contributing to the performance of individual subnetworks, and their combinations. For example, subnetworks may contribute highly to identification because their individual-specific functional connectivity information capitalizes on the contributions of these subnetworks to task performance. Alternatively, but not mutually exclusively, subnetworks that do not participate as much during a task may contain diagnostic information with respect to participant identity.

To what extent does participant identification transfer between experimental conditions? We found that training with one condition and testing with another produced good levels of identification accuracy. Certain combinations that on the surface were not obvious produced particularly impressive results; for example, training with *gambling* and testing with *working-memory*, or training with working-memory and testing with *language*. Training with *motor* produced the least transfer to other tasks, perhaps due to the low-level specificity of this task. Notably, training with *resting-state* produced very high transfer, such that testing with each task attained accuracy over 80% (with the exception of *relational processing*), and in some instances over 90% (note that accuracy with *working-memory* was 99%).

4.2. *Low-dimensional distance visualizations*

Relationships between high-dimensional FC matrices (300×300) were visualized in three Euclidean-space dimensions using multidimensional scaling. Both the Pearson dissimilarity measure and geodesic distance were used. Note that computing geodesic distances takes into account the non-linear geometry of correlation matrices. Once their distances are computed, and the space nonlinearity taken into account, they can be illustrated in Euclidean space (naturally, some distortion ensues due to dimensionality reduction).

In our explorations, low-dimensional visualizations reflected identification accuracy on the full data, and thus preserved important distance information. In particular, the higher

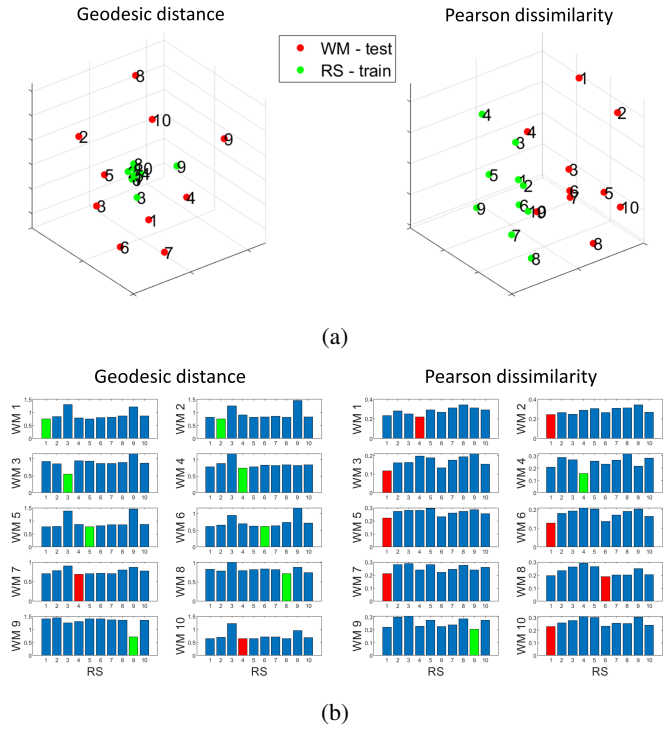


Fig. 11: Visualization of task and *resting-state* functional connectivity distances/dissimilarities in a three-dimensional space using multidimensioanl scaling. The numbers indicate participant IDs. A. Distances/dissimilarities between the functional connectivity matrices of *resting-state* (RS, used for training) and *working memory* (WM, used for testing) for a set of 10 randomly chosen participants. B. Participant-level distances/similarities between training and testing data. Correct identification is marked in green and incorrect in red. For example, when using geodesic distance, the best candidate for WM participant 1 (call it WM1) was RS participant 1 (RS1), and the best candidate for WM2 was RS2. However, incorrect classifications were also observed, such as RS4 (not RS7) being closest to WM7. For Pearson dissimilarity most classifications were incorrect, such as RS1 (not RS10) being most similar to WM10.

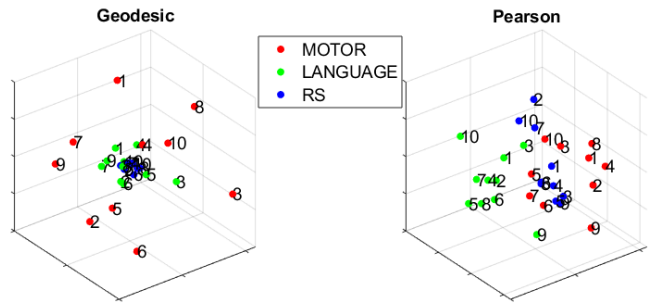


Fig. 12: Functional connectivity geometry of *resting-state* and task conditions. Training data for 10 random participants were employed (indicated by the numbers). Distances/dissimilarities in low dimensions were obtained via multidimensional scaling. Note that the geometry in low dimensions differed considerably for geodesic and Pearson, suggesting that condition categorization (not participant identification) should capitalize on such geometry for better performance.

identification accuracy using the geodesic distance resulted in relatively low within- and high between-participant distances. Visualization of FC from task data revealed insights into the geometry of task correlation matrices in relation to *resting-state*. Identification accuracy is related to the ratio of within- to between-participant distances. Surprisingly, with geodesic distances, tasks associated with higher identification accuracy exhibited smaller between-participant distances. Still, the more favorable ratio of within- to between-participant distances led to favorable identification accuracy. Thus, the underlying geometry of functional connectivity may provide further insights into our finding that high identification accuracy was attained when training and testing were based on different scanning conditions.

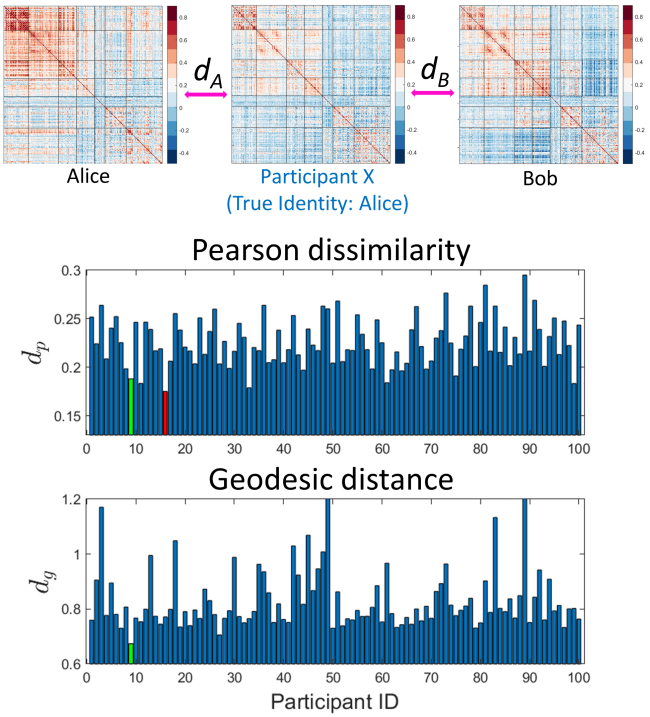


Fig. 13: Visual comparison of functional connectivity (FC) matrices can be unintuitive. Example FCs from *resting-state* data where the geodesic distance correctly labeled the test participant but Pearson dissimilarity did not. The bottom plots show the geodesic distances and Pearson dissimilarities between the test FC and each of the FCs in the training data. The green bar indicates the distance between the test FC to the correct training set FC; the red bar indicates an incorrectly labeled training set FC. For the geodesic distance, the labeled participant had indeed the smallest value; not so in the case of Pearson dissimilarity. This example also questions the common practice of informally evaluating functional connectivity similarity via simple visual inspection. At the very least, it is not immediate that participant *X* is more similar to Alice than Bob.

In the visualizations based on geodesic distance, distances between task FCs did not appear to form convex sets (if A and B are two points in a convex set, any point on the line joining them also belongs to the set), and were instead in clustered arrangements. Of note, previous work performing clustering of FCs [2, 13] have used k-means which are not well suited to finding non-convex clusters [8]. Instead, methods such as spectral clustering are capable of capturing very general structures [19], and are potentially more suitable for classifying functional connectivity.

Pearson correlation is a common approach to compare FC matrices. The present study demonstrates that non-linear measures are better suited to characterize functional connectivity

relationships. The low-dimensional visualization briefly explored here hints at the different geometries associated with the geodesic non-linear metric and the Pearson approach. Surprisingly, we noted in our investigations that simple visual inspection of the correlation matrices as commonly done in the field to highlight similarities between conditions can also be problematic, and in fact can lead to unintuitive scenarios (Fig 13).

4.3. Conclusions

Time series correlation matrices capture important aspects of brain functional organization. Here, we propose the use of a geodesic distance metric that reflects the underlying non-Euclidean geometry of functional connectivity matrices. We compared identification performance (also called fingerprinting; that is, assigning a participant label to novel functional connectivity data) obtained with standard Pearson correlation and the proposed geodesic distance. The latter not only improved identification accuracy but also provided insights into the geometry of task and resting-state conditions. Importantly, the approach advocated here is general and can be utilized to study the clustering of brain states, how tasks potentially reconfigure brain networks, and to characterize intersubject correlations.

Acknowledgements

M.V. was supported by a fellowship by the Brain and Behavior Initiative, University of Maryland, College Park. L.P. is supported by the National Institute of Mental Health (R01 MH071589 and R01 MH112517). Data were provided 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.

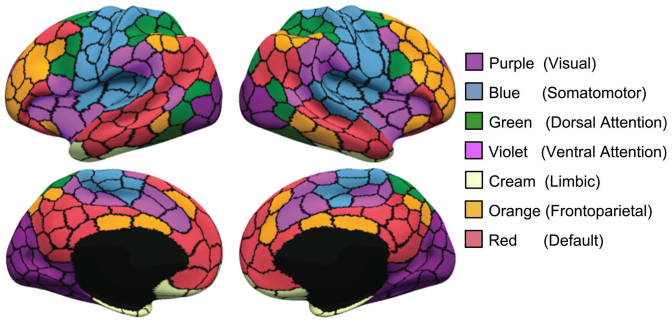


Fig. S1: Parcellation of the cortex into 300 ROIs as provided by [23]. ROIs were grouped into the 7 networks described in [29].

Supplemental material

S1. Effect of number of ROIs in the parcellation on identification

To study the effect of the parcellation scheme on participant identification accuracy using the two measures, we employed various parcellations with ROIs ranging from a 100 to 400. In general, mean participant identification accuracy increased with increase in ROIs indicating that finer resolution or detail in the FC revealed more uniqueness. Mean accuracy using the geodesic distance was consistently higher than the mean accuracy using Pearson dissimilarity. For several conditions (*resting-state*, *language*, *motor*), accuracy using geodesic distance on FCs obtained with 100 ROIs was greater than accuracy obtained using Pearson dissimilarity with 400 ROIs (Fig. S2).

S2. Computing geodesic distances for matrices without full rank

Computing the geodesic distance between two FC matrices Q_1 and Q_2 (Equation 2) requires Q_1 to be invertible, or equivalently, all the eigenvalues of Q_1 must be strictly greater than zero. When FC matrices are based on n ROIs and n is larger than number of frames in the run, the *rank* of the resulting FC matrix is not full (i.e., $< n$), and some of its some eigenvalues are equal to 0. In practice, when the number of ROIs $n < (0.9 \times$ number of frames), we applied the procedure below to ensure full rankness.

To handle such cases, we adopted a simple approach here: we added the identity matrix I to both Q_1 and Q_2, causing the eigenvalues of the correlation matrices of interest to be increased by 1. Because all eigenvalues are then greater than 0, the matrices are invertible. In such cases, the geodesic distance, $d_G(Q_1+I, Q_2+I)$, serves as a proxy for the geodesic distance between the two matrices. Note that the scenario of low-rank FC matrices arises only for whole-cortex analysis, as for the subnetwork analyses, the number of ROIs in question was always greater than the number of frames in the run.

For reference, the procedure above was employed in the following cases: whole-cortex results for all tasks; whole-cortex *resting-state* results with lengths less than 400 TRs; and whole-cortex results involving trimmed data.

S3. Replication - Constable 2015

The problem of participant identification has been addressed previously by Finn et al [10]. In their study, a group of 126 participants from the Q2 release of HCP were used. The release consisted of volumetric data for each participant in the MNI template. Standard preprocessing procedures were applied to the fMRI data and included removal of linear components related to the 12 motion parameters, regression of the mean time courses of the white matter and CSF as well as the global signal, removal of linear trend, and low-pass filtering. The data was summarized into an ROI-level time series by using the Shen parcellation that consisted of 268 ROIs. For participant identification, they also used a 1-nearest neighbor classifier with Pearson dissimilarity as the closeness measure. They reported an identification accuracy of around 90% on *resting-state* data which is much higher than the 75% we obtained using surface-based fMRI data from the latest release of HCP.

To understand the differences, we also performed participant identification of the unrelated 100 subject volumetric fMRI data provided by HCP. However, the current release provides aggressively preprocessed resting-state data using ICA-FIX [25]. Using Pearson dissimilarity and the Shen parcellation, identification accuracy was 83% (as opposed to 75% with the Yeo parcellation on surface fMRI data). Using the geodesic distance, accuracy again improved to 95% (96% with the Yeo parcellation). Although we didn't obtain an accuracy of 90% as in Finn. et al for this group of 100 subjects, these results indicate that the use of the geodesic distance improves the accuracy substantially irrespective of the type of preprocessing on the data.

[1] Aggarwal, C.C., Hinneburg, A., Keim, D.A., 2001. On the surprising behavior of distance metrics in high dimensional space, in: International conference on database theory, Springer. pp. 420–434.

[2] Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D., 2014. Tracking Whole-Brain Connectivity Dynamics in the Resting State. Cerebral Cortex 24, 663–676. doi:10.1093/cercor/bhs352.

[3] Amrhein, V., Greenland, S., McShane, B., 2019. Scientists rise up against statistical significance.

[4] Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J.M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A.Z., Essen, D.C.V., Consortium, f.t.W.M.H., 2013. Function in the human connectome: Task-fMRI and individual differences in behavior. NeuroImage 80, 169–189. doi:10.1016/j.neuroimage.2013.05.033.

[5] Boyd, S., Vandenberghe, L., 2004. Convex optimization. Cambridge university press.

[6] Cox, R.W., 1996. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical research 29, 162–173.

[7] Essen, D.C.V., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, f.t.W.M.H., 2013. The WU-Minn Human Connectome Project: An overview. NeuroImage 80, 62–79. doi:10.1016/j.neuroimage.2013.05.041.

[8] Estivill-Castro, V., 2002. Why so many clustering algorithms: a position paper. ACM SIGKDD Explorations Newsletter 4, 65–75. doi:10.1145/568574.568575.

[9] Finn, E.S., Scheinost, D., Finn, D.M., Shen, X., Papademetris, X., Constable, R.T., 2017. Can brain state be manipulated to emphasize individual differences in functional connectivity? NeuroImage 160, 140–151. doi:10.1016/j.neuroimage.2017.03.064.

[10] Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome

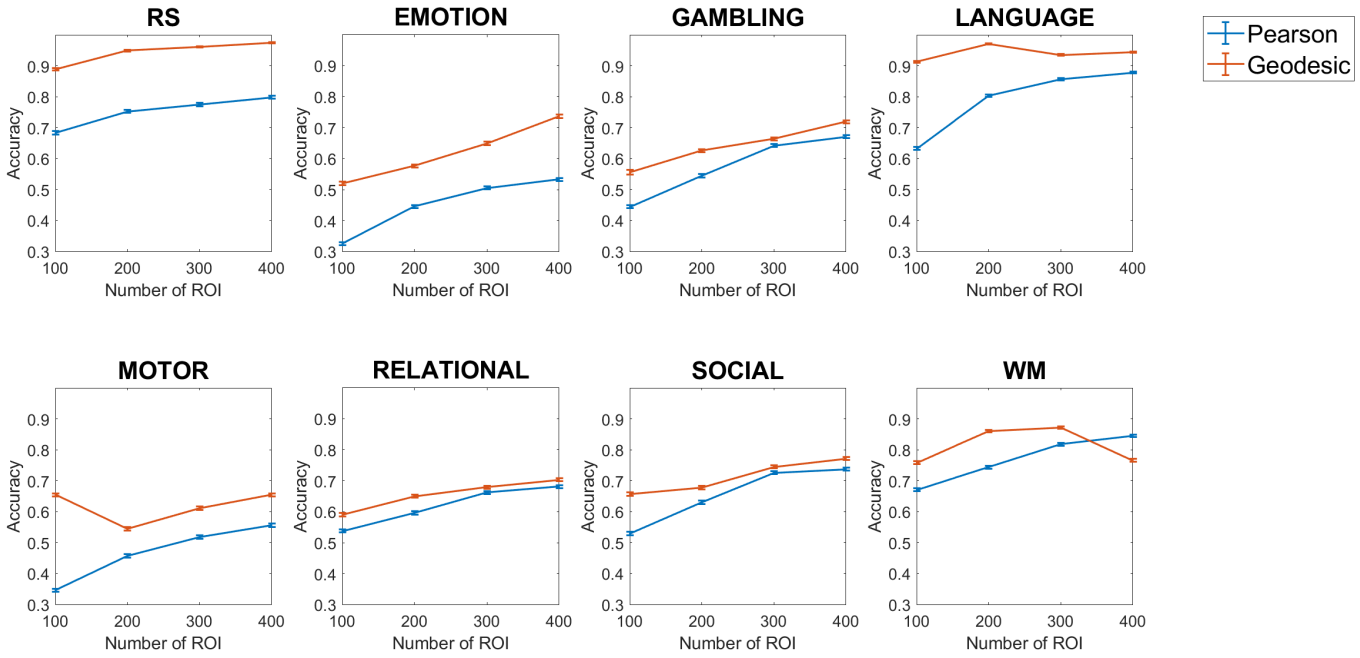


Fig. S2: Participant identification accuracy as a function of the number of ROIs. Here, training and testing data are from the same task. Error bars indicate standard error of the mean across the bootstrap iterations.

Scan	REST	EM	GB	LG	MT	RL	SO	WM
p-value	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-4}$	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-32}$

Table S1: *Whole-cortex FCs:* Comparing geodesic distance and Pearson dissimilarity for participant identification. p-values based on a bootstrapping procedure. The distributions of the means are shown in Fig. S3.

fingerprinting: identifying individuals using patterns of brain connectivity. Nature neuroscience 18, 1664.

[11] Förstner, W., Moonen, B., 2003. A metric for covariance matrices, in: Geodesy-The Challenge of the 3rd Millennium. Springer, pp. 299–309.

[12] Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Essen, D.C.V., Jenkinson, M., Consortium, f.t.W.M.H., 2013. The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage 80, 105–124. doi:10.1016/j.neuroimage.2013.04.127.

[13] Gonzalez-Castillo, J., Hoy, C.W., Handwerker, D.A., Robinson, M.E., Buchanan, L.C., Saad, Z.S., Bandettini, P.A., 2015. Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. Proceedings of the National Academy of Sciences 112, 8762–8767. doi:10.1073/pnas.1501242112.

[14] Hansen, E.C., Battaglia, D., Spiegler, A., Deco, G., Jirsa, V.K., 2015. Functional connectivity dynamics: Modeling the switching behavior of the resting state. NeuroImage 105, 525–535. doi:10.1016/j.neuroimage.2014.11.001.

[15] Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29, 1–27.

[16] Leonardi, N., Ville, D.V.D., 2015. On spurious and real fluctuations of dynamic functional connectivity during rest. NeuroImage 104, 430–436. doi:10.1016/j.neuroimage.2014.09.007.

[17] MATLAB, 2018. version 9.4.0.813654 (R2018a). The MathWorks Inc., Natick, Massachusetts.

[18] McShane, B.B., Gal, D., Gelman, A., Robert, C., Tackett, J.L., 2019. Abandon Statistical Significance. The American Statistician 73, 235–245. doi:10.1080/00031305.2018.1527253.

[19] Ng, A.Y., Jordan, M.I., Weiss, Y., 2002. On spectral clustering: Analysis and an algorithm, in: Advances in neural information processing systems, pp. 849–856.

[20] Pennec, X., Fillard, P., Ayache, N., 2006. A riemannian framework for tensor computing. International Journal of computer vision 66, 41–66.

[21] Pessoa, L., 2014. Understanding brain networks and brain organization. Physics of life reviews 11, 400–435.

[22] Ponsoda, V., Martínez, K., Pineda-Pardo, J.A., Abad, F.J., Olea, J., Román, F.J., Barbey, A.K., Colom, R., 2017. Structural brain connectivity and cognitive ability differences: A multivariate distance matrix regression analysis. Human brain mapping 38, 803–816.

[23] Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T., 2017. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral Cortex 28, 3095–3114.

[24] Schultz, D.H., Cole, M.W., 2016. Higher intelligence is associated with less task-related brain network reconfiguration. Journal of Neuroscience 36, 8551–8561.

[25] Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Woolrich, M.W., Xu, J., Yacoub, E., Uurbil, K., Essen, D.C.V., Glasser, M.F., Consortium, W.M.H., 2013. Resting-state fMRI in the Human Connectome Project. NeuroImage 80, 144–168. doi:10.1016/j.neuroimage.2013.05.039.

[26] Sporns, O., 2010. Networks of the Brain. MIT press.

[27] Van Dongen, S., Enright, A.J., 2012. Metric distances derived from cosine similarity and pearson and spearman correlations. arXiv preprint arXiv:1208.3145 .

[28] Xie, H., Calhoun, V.D., Gonzalez-Castillo, J., Damaraju, E., Miller, R., Bandettini, P.A., Mitra, S., 2018. Whole-brain connectivity dynamics reflect both task-specific and individual-specific modulation: A multitask study. NeuroImage 180. doi:10.1016/j.neuroimage.2017.05.050.

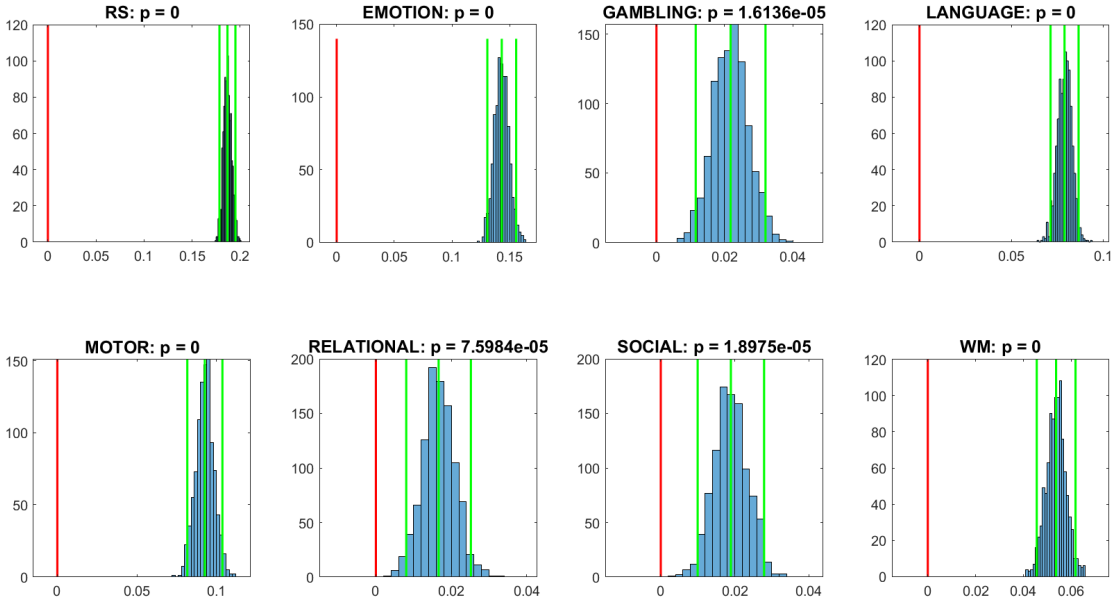


Fig. S3: *Whole-cortex FCs:* Comparing accuracy based on geodesic distance and Pearson dissimilarity for each task. Here, full time course lengths were used for each task. For each outer bootstrap iteration, the difference between the mean participant identification accuracy using the geodesic distance and the Pearson dissimilarity were computed to obtain difference distributions. The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

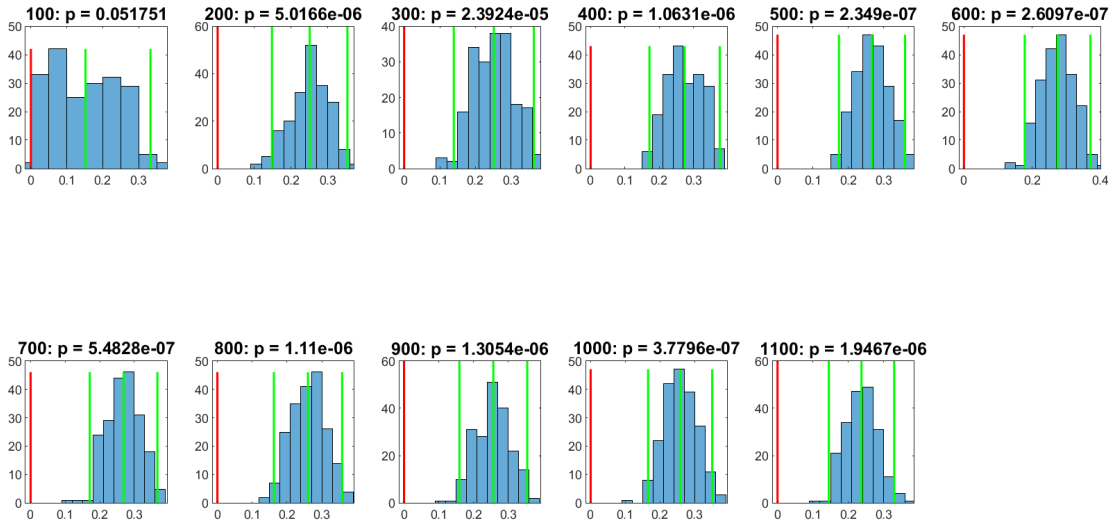


Fig. S4: *Whole-cortex FCs:* Comparing accuracy based on geodesic distance and Pearson dissimilarity for each segment length. For each bootstrap iteration, the difference between the mean participant identification accuracy using the geodesic distance and the Pearson dissimilarity across the 8 tasks were computed to obtain difference distributions. The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

[29] Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zllei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology 106, 1125–1165. doi:10.1152/jn.00338.2011.

[30] Zalesky, A., Breakspear, M., 2015. Towards a statistical test for functional

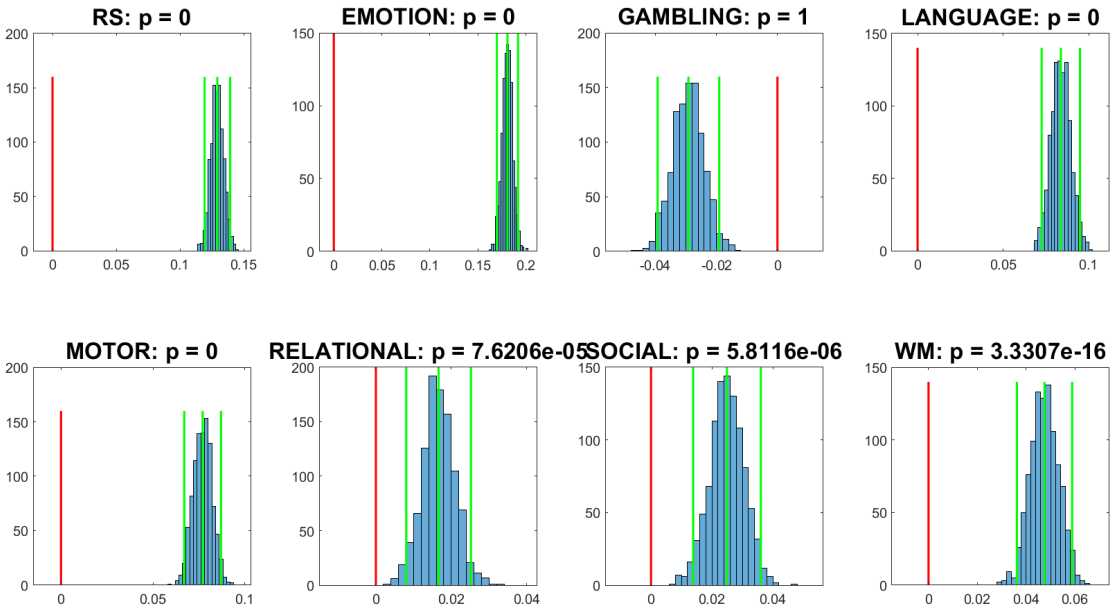


Fig. S5: *Whole-cortex FCs:* Comparing accuracy based on geodesic distance and Pearson dissimilarity for each task. Here, tasks were trimmed such that they all had the same time course length (of 138). For each bootstrap iteration, the difference between the mean participant identification accuracy using the geodesic distance and the Pearson dissimilarity across the 8 tasks were computed to obtain difference distributions. The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

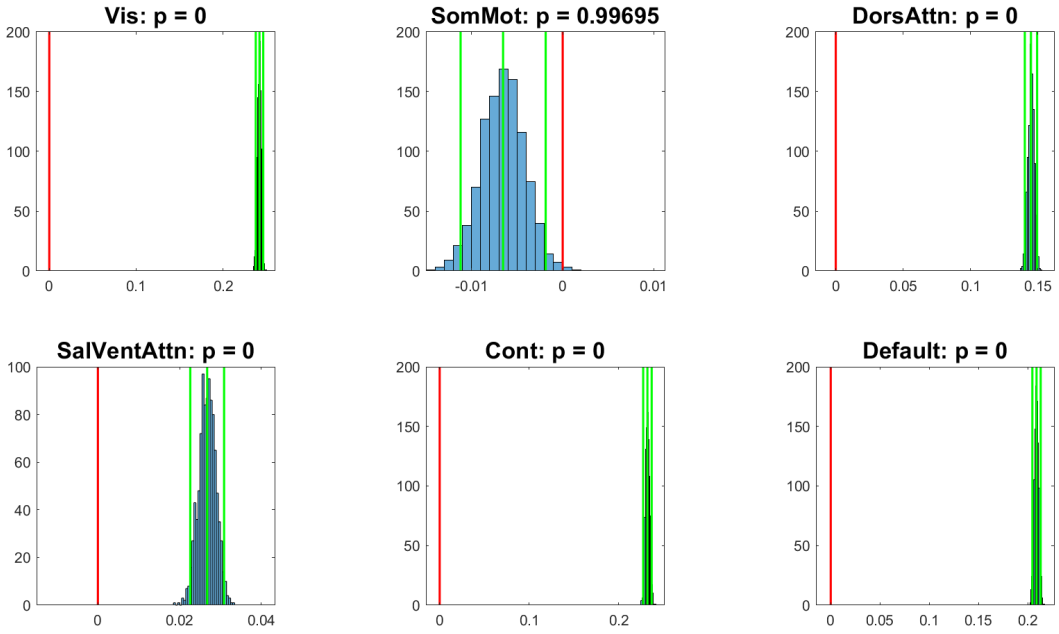


Fig. S6: *Subnetwork FCs:* Comparing mean accuracy based on geodesic distance and Pearson dissimilarity in each subnetwork. Here, tasks were trimmed such that they all had the same time course length (of 138). For each bootstrap iteration, the difference between the mean participant identification accuracy using the geodesic distance and the Pearson dissimilarity across the 8 tasks were computed to obtain difference distributions. The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

connectivity dynamics. NeuroImage 114, 466–470. doi:10.1016/j.neuroimage.2015.03.047.

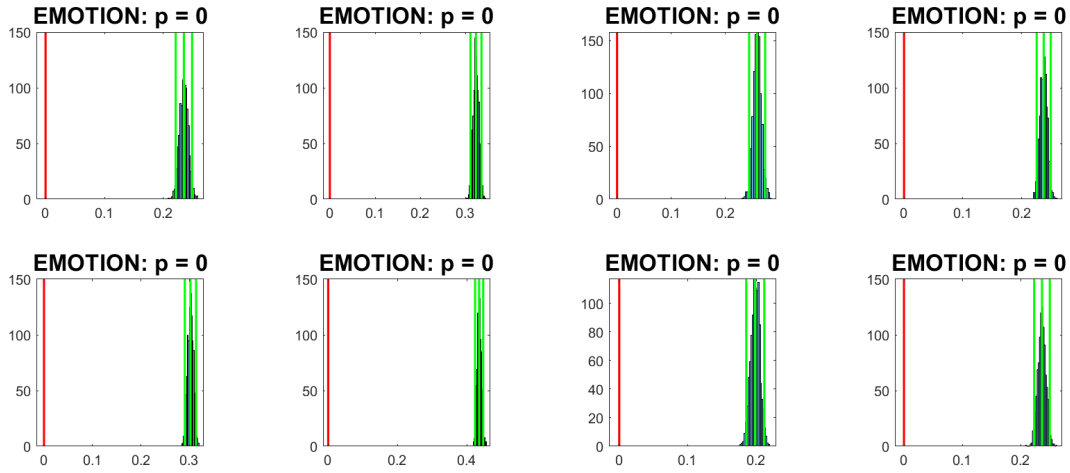


Fig. S7: *Subnetwork FCs:* Comparing mean accuracy in the `dorsal attention` and `ventral attention` subnetworks for each condition using the geodesic distance. The subnetworks were of identical size for the 300 ROIs parcellation. Here, tasks were trimmed such that they all had the same time course length (of 138). The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

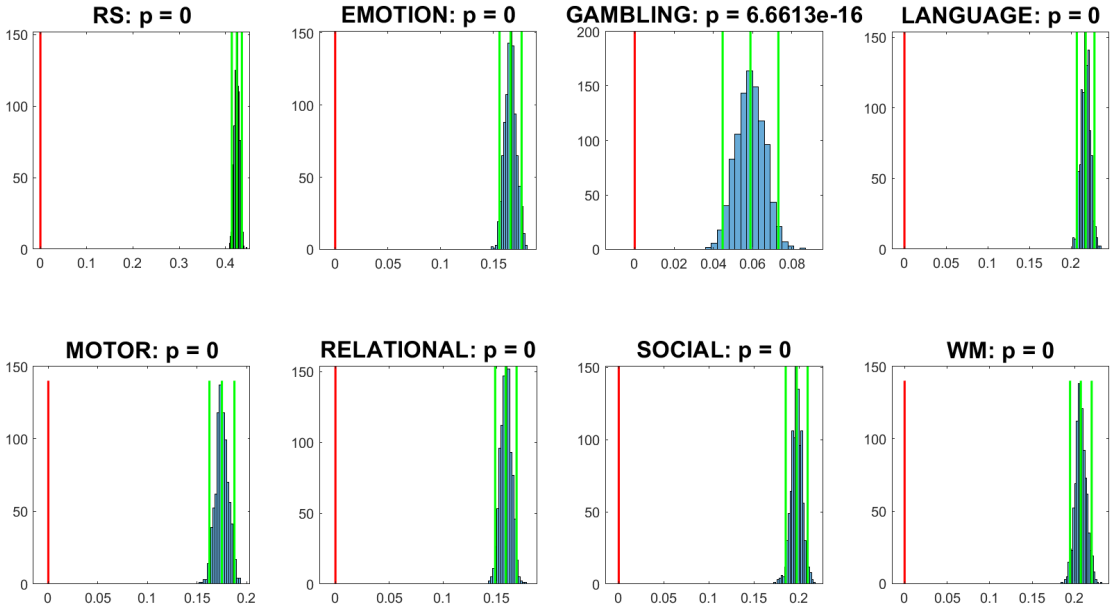


Fig. S8: *Combined subnetwork FCs:* Comparing accuracy based on geodesic distance and Pearson dissimilarity for each condition. Here, tasks were trimmed such that they all had the same time course length (of 138). For each outer bootstrap iteration, the difference between the mean participant identification accuracy using the geodesic distance and the Pearson dissimilarity were computed to obtain difference distributions. The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

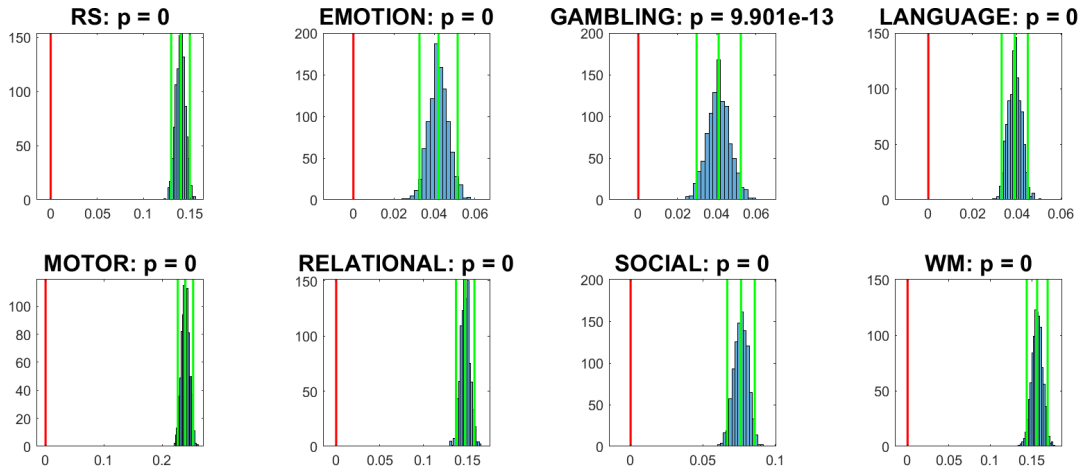


Fig. S9: *Geodesic distance:* Comparing identification accuracy using the combined subnetwork with the `control` subnetwork (part of the combined subnetwork) for each condition. The error bars represent standard error of the mean across 1000 bootstrap iterations. The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

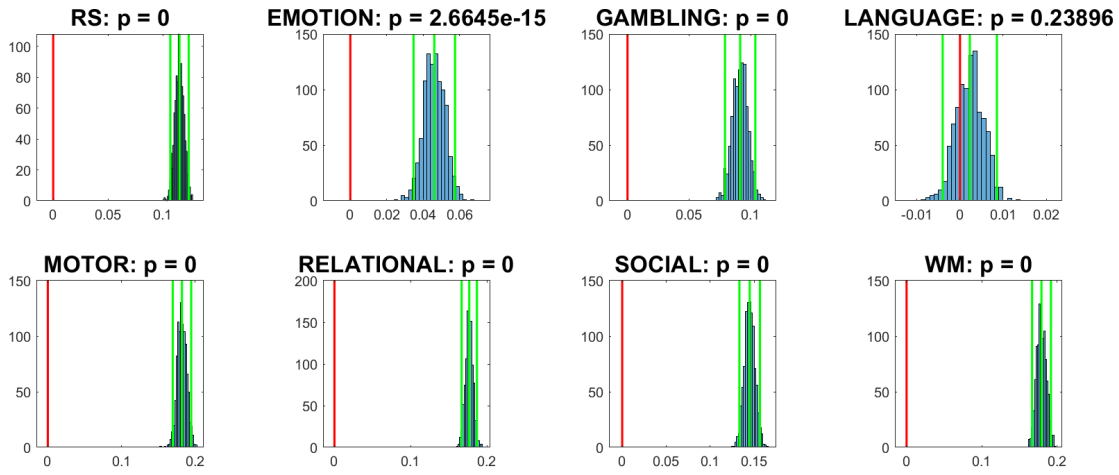


Fig. S10: *Geodesic distance:* Comparing identification accuracy using the combined subnetwork with the `default` subnetwork (part of the combined subnetwork) for each condition. The error bars represent standard error of the mean across 1000 bootstrap iterations. The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

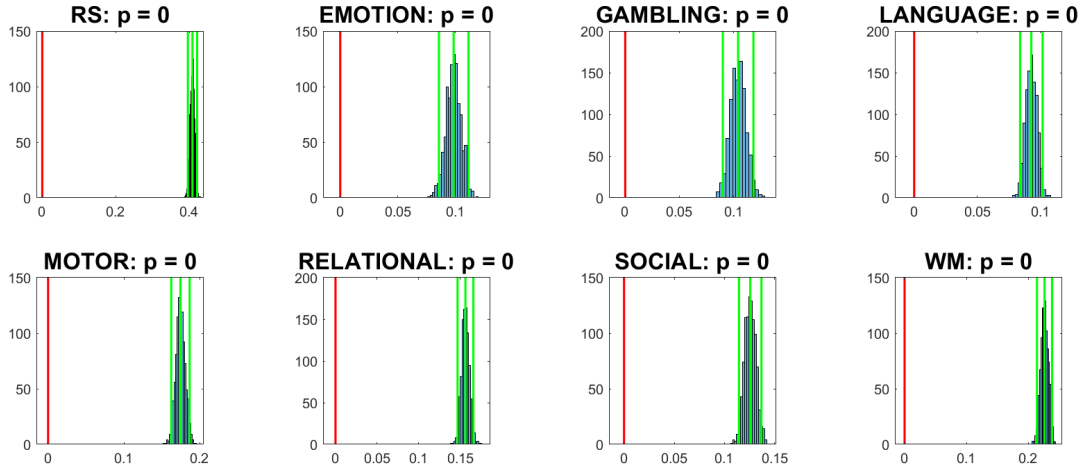


Fig. S11: *Geodesic distance:* Comparing identification accuracy using the combined subnetwork with whole-cortex FCs for each condition. The error bars represent standard error of the mean across 1000 bootstrap iterations. The three green lines indicate the mean and the 95% confidence interval. The red line indicates zero difference.

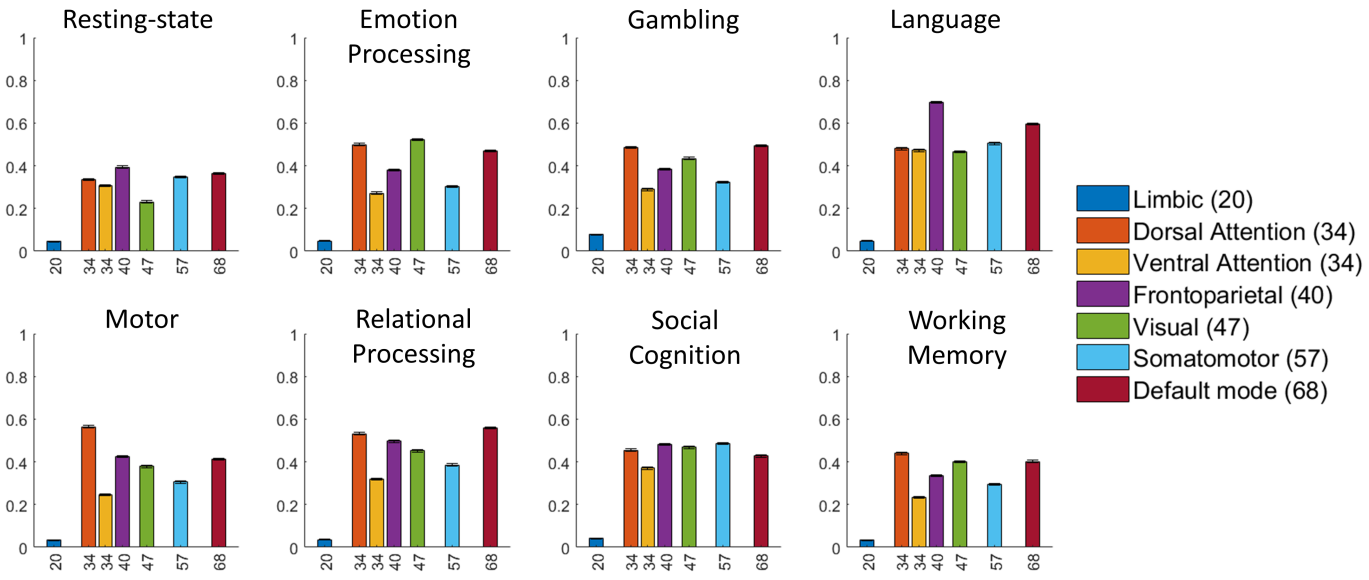


Fig. S12: *Pearson dissimilarity:* Participant identification accuracy plotted against subnetwork size for each task. The error bars represent standard error of the mean across 1000 bootstrap iterations.