Proceedings of the Master's Programme
COGNITIVE NEUROSCIENCE

Underlying Features of Inter-Individual Variability in Brain Function Using a Siamese Neural Network

Marleen Voorn[1]

Supervisors: Linda Geerligs[1] and Umut Güçlü[1]

[1]*Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands*

Corresponding author: Marleen Voorn
E-mail: marleenvoorn.95@gmail.com

Large-scale networks are comprised of distributed brain areas that show correlated fluctuations in their spontaneous resting state functional magnetic resonance imaging activity, and these patterns of functional connectivity show high individual variability. However, it is still unknown what the most important dimensions are along which individual functional connectivity patterns can vary. In this study, we investigated the discriminating inter-individual features of brain function. We used a deep neural network to learn to discriminate subjects' connectivity patterns from each other (n = 996). We found that we can reduce the high-dimensional connectivity profiles to only three components and still identify subjects' functional connectivity patterns with 96% accuracy. The three identified components were mapped back onto the large-scale brain networks, and were classified as 1) sensorimotor networks component, 2) within-networks component, 3) higher-order networks component. Furthermore, we found correlations between the higher-order networks component and cognitive measures as well as with relevant brain regions' area size. This implies that we identified three robust components of functional connectivity that can be linked to cognition and amount of grey matter.

Human neuroimaging studies often investigate general patterns of brain activity that are shared across individuals. Spontaneous fluctuations as measured with resting state functional magnetic resonance imaging (fMRI – rfMRI) show temporal correlations between various brain regions, showing complex patterns of functional connectivity (FC) (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995). This group averaging has resulted in basic understanding of functional brain organization: by combining data from thousands of individuals, neuroscientists have been able to reliably describe the organization of the human brain into large-scale functional networks using fMRI scans. Large-scale networks are comprised of distributed brain areas that show correlated fluctuations in their spontaneous rfMRI activity.

Even though this provided insight in group-level tendencies of functional connectivity that generalize across individuals, it also obscured information about subject-specific features (Gordon et al., 2017): it may be possible that individual functional resting connectivity reflects a unique processing

organization (Seitzman et al., 2019). Moreover, the heterogeneity across subjects is ignored: group averaging limits the detail and specificity of the resting state functional connectivity.

It has been shown that functional connectivity in subjects at rest shows high individual variability (Finn et al., 2015). Therefore, if we use an approach that moves beyond the population level to inferences about single subjects, descriptors of unique functional brain organization can be explored.

To investigate whether there are features of functional connectivity that are distinct enough to identify an individual, several studies have demonstrated that individuals can accurately be identified based on their functional connectivity patterns (Finn et al., 2015; Gordon et al., 2017; Gratton, Sun, & Petersen, 2017). This implies that each individual's functional architecture is unique. Identification across subjects is accurate regardless whether fMRI data is collected during a task or during rest. This indicates that 1) functional networks are fundamentally stable, and 2) rfMRI patterns are informative about a given person's stable traits regardless of that person's thoughts or behavioural task during scanning (Gratton et al., 2017). When rfMRI data of different sessions between subjects were compared, they had high similarity within subject and much lower similarity across subjects (Gordon et al., 2017). Thus, resting state functional connectivity profiles were found to be unique as well as stable over time, and could therefore act in a similar way as a fingerprint (Finn et al., 2015).

This variability is not only found in single regions, but also in large-scale networks (Seitzman et al., 2019). Individual resting state variability was strongest in higher-order association cortices or 'control' systems, rather than primary sensory cortices or 'processing' systems (Finn et al., 2015; Seitzman et al., 2019). The most successful networks for identification of individuals appear to be 1) the medial frontal network, which is part of the default mode network (DMN), and 2) the frontoparietal (FP) network (Finn et al 2015).

If we could link this unique functional organization on the level of large-scale brain networks to behavioural phenotypes, it would provide insight in the neural mechanisms that underlie cognition. It has been demonstrated that resting state functional connectivity is a good predictor for task outcomes (Finn et al., 2015; Rosenberg et al., 2016; Shen et al., 2017; Tavor et al., 2016). Furthermore, the networks that show the highest inter-individual variability in functional connectivity (i.e., the networks that are most discriminative between individuals) were also most relevant to individual differences in behaviour (Finn et al., 2015). This suggests that the functional hierarchy that is observed during rest is the same during action and cognition (Tavor et al., 2016).

However, it is still unknown what the most important dimensions are along which individual functional connectivity patterns can vary. It is difficult to quantify how all types of network variants may be distributed (Gordon et al., 2017) and how this is organized to the level of individuals. To overcome this multiple comparison problem, an approach that reduces the feature space is required. Deep neural networks (DNNs) are of growing importance in the field of neuroscience and this approach is particularly suitable for this matter: Artificial neural networks are inspired by concepts of neuroscience, but neglect many details of biological neural networks. DNNs have great computational efficiency, which enables them to perform complex tasks such as image recognition or translation (Kietzmann, McClure, & Kriegeskorte, 2018). Moreover, biological functional neural networks can be recreated by an artificial neural network, such as an "attention network" (Xu et al., 2015). Thus, concepts from neuroscience have inspired engineers to build DNNs. This pipeline can also be reversed: neuroscientists have been inspired by results from DNNs (Storrs & Kriegeskorte, 2019).

The main aim of this study was to unravel the underlying features of inter-individual variability in functional brain connectivity. To this end, a DNN was trained to distinguish rfMRI patterns between subjects. The network reduced the high-dimensional functional connectivity patterns to their most relevant features. To extract and summarize the most relevant features from functional network connectivity, a Siamese neural network (SNN) was used. This is a specific DNN that once it is trained, can accurately dissociate between similar input (e.g., two different photos of the same face) and dissimilar input (Koch, 2015). By analysing the features that the network has

learned, we can gain a deeper understanding of individual differences in brain function. Furthermore, to understand the nature and neural mechanisms of individual rfMRI patterns, it is of importance to determine what information in the resting-state signal is driving the predictors for cognition (Tavor et al., 2016).

Methods

Human Connectome Project data

Subject information

The data set used in this study is from the Human Connectome Project (HCP), the WU-Minn HCP Data - 1200 subjects (Van Essen et al., 2013). This data set has a large sample size in comparison to other neuroimaging studies. The HCP data release is an open resource database, facilitating replication and extension by other researchers (Finn et al., 2015). The included HCP resting-state fMRI runs were acquired at 3T in sessions of around 15-20 minutes on two separate days. Both days (HCP file names: rfMRI_REST1 and rfMRI_REST2) included a LR as well as a RL phase encoding sessions. An EPI acceleration factor of x8 was used, resulting in spatial resolution of 2 × 2 × 2 mm and temporal resolution (TR) of 0.72 s (Smith et al., 2013).

The full release contained 1113 healthy subjects with 3T MR Session Data, but the data analysis was constrained to subjects for whom all four resting state fMRI sessions acquired on two separate days were available (n = 996; age 22-35). Each session consisted of 1200 TRs (with a TR of 0.72 s), resulting in 4800 TRs per subject for further analysis.

Pre-processing

The data analysis was based on rfMRI data that was pre-processed with ICA-FIX (HCP file names: Resting State fMRI 1 FIX-Denoised [Extended] and Resting State fMRI 2 FIX-Denoised [Extended]). ICA was used to remove nuisance and motions signals(Glasser et al., 2013).

First, regression analysis of the resting state fMRI activity was performed for all four sessions individually to filter out the brain signals with low frequency. A high-pass filter of 0.008 Hz was applied, because slow signal changes originate predominately from non-physiological sources such as MR scanner drift, and were therefore treated as covariates of no interest (Fransson, 2005).

Next, the cortical parcellation scheme from (Gordon et al., 2016) was used, because it maximized the similarity of functional connectivity within each parcel. Moreover, the functional connectivity homogeneity within each parcel is high (Betzel et al., 2019). This scheme consisted of 333 re-gions of interest (ROIs) distributed over 12 large-scale networks, which are: Default, Visual, FrontoParietal, DorsalAttention, VentralAttention, Salience, CinguloOpercular, Cingulo-Parietal, SomatoMotorHand, SomatoMotor-Mouth, Auditory, and RetrosplenialTemporal (Gordon et al., 2016).

Last, ROIs with less than 20 active voxels were excluded. Active voxels were defined as voxels where a significant change as compared to the baseline in the blood-oxygen-level-dependent (BOLD) signal was detected. This resulted in 332 ROIs for further analysis.

Connectivity profiles

Calculation of connectivity matrices was performed in MATLAB. Pearson correlation coefficients between pairs of ROI time courses over all four sessions were calculated. This resulted in a 332 × 332 symmetric connectivity matrix per subject. Thus, a connectivity profile included correlations for all ROIs with the other ROIs over the complete time course.

Siamese Neural Network

Model architecture

The model we used to distinguish rfMRI connectivity patterns between subjects is a Siamese Neural Network with two symmetric twin networks (Fig. 1), implemented with PyTorch (Python). Each twin network consisted of five fully connected layers with Nl hidden units per layer l. The inputs x_1 and x_2 are the flattened upper triangular of connectivity matrices, resulting in an input size of $N \cdot (N-1) / 2$ where N is the number of included ROIs. Thus, the dimension of one input sample was 54946 × 1. Because of the

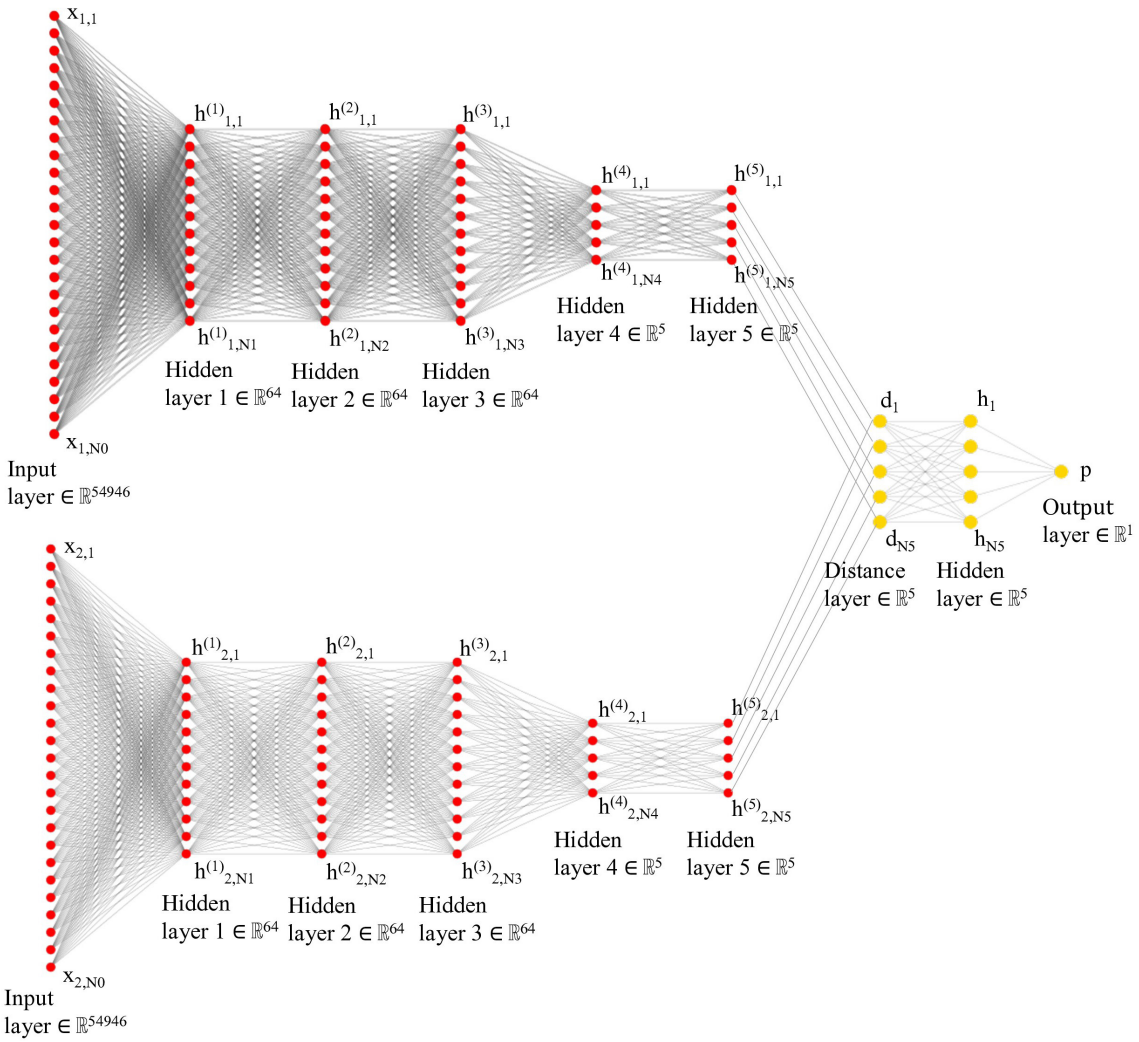

Figure 1. The model used: a Siamese Neural Network with five hidden fully connected layers. Inputs were flattened upper triangular parts of two different connectivity matrices and fed into each Input layer (red nodes, first column) of each twin network. Outputs of each twin network were the activations of Hidden layer five (red nodes, last column). Twin networks were joined with a distance function in the Distance layer (yellow nodes, first column) and resulted in binary classification with logistic prediction p (final yellow node). Symmetric twin networks (red nodes) share the same weight matrices at each layer. N0 = 54946, N1–3 = 64, N4–5 = 5.

symmetry, both twins of the network share the same weight and bias values. Weight and bias values are the learnable parameters by the model, where weight values control the strength of connections between the input and hidden nodes and bias values are applied with the input. Therefore, twins compute the same metric when x_1 and x_2 were presented to twin network 1 and 2, or when x_1 and x_2 were swapped and presented to the twin networks (Koch, 2015).

In this model, the activation function used in all hidden layers was the rectified linear unit (ReLU) function (Equation 1):

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \quad (1)$$

This is a linear function where all negative inputs are returned as zero and all positive inputs are returned unchanged. ReLU is often used as default activation function for neural networks because it is easy to train. We also tried to train the model using the *tanh* activation function: $\frac{e^x - e^{-x}}{e^x + e^{-x}}$, but this resulted in lower accuracy of the model.

Each hidden vector in a fully connected layer l is denoted as $\mathbf{h}_{1,l}$ for the first twin, and

$\mathbf{h}_{2,l}$ for the second twin. Thus, after the last fully connected layer (l=5), twin network one and twin network two respectively output feature vector $\mathbf{h}_{1,5}$ and $\mathbf{h}_{2,5}$ based on input x_1 and x_2. This last fully connected layer outputs the highest-level feature representation on each side. The dimension of these representations is determined by N_5, which was 5 × 1. This was the minimum number of hidden nodes in the final hidden layer five that still gave high accuracy.

Last, the twins are joined by a distance function. The distance is determined as the absolute difference between $\mathbf{h}_{1,5}$ and $\mathbf{h}_{2,5}$, combined by a final ReLU followed by a sigmoid activation (σ) that outputs a value for p (Equation 2):

$$p = \sigma(\alpha|\mathbf{h}_{1,l} - \mathbf{h}_{2,l}|) \quad (2)$$

where α are additional parameters learned by the model, weighting the importance of the distance of all five elements between $\mathbf{h}_{1,5}$ and $\mathbf{h}_{2,5}$ (Koch, 2015). The sigmoid function was chosen as final activation because it maps onto interval ‹0,1›, which can be interpreted as the probability that x1 and x2 belong to the same subject (p ≥ .5) or to different subjects (p < .5).

Learning

The aim was to train the network in such a way that it can discriminate between pairs of connectivity profiles it has never seen before as belonging to the same subject, or to different subjects.

Preparation of fMRI volumes as input to the DNN. It was required to create multiple connectivity profiles per subject to have a sufficient input set for the SNN to be able to learn the underlying discriminating features of brain connectivity. A total of 4,780,800 fMRI volumes (1200 volumes/run × 4 runs/subject × 996 subjects) were available across all subjects. Multiple connectivity profiles for one subject were created by including different (random) parts of the total time course when the Pearson correlation coefficient (reflecting connectivity between ROIs) was calculated, resulting in multiple connectivity profiles based on one subject's resting state fMRI data. With this, it was possible to create input pairs x1

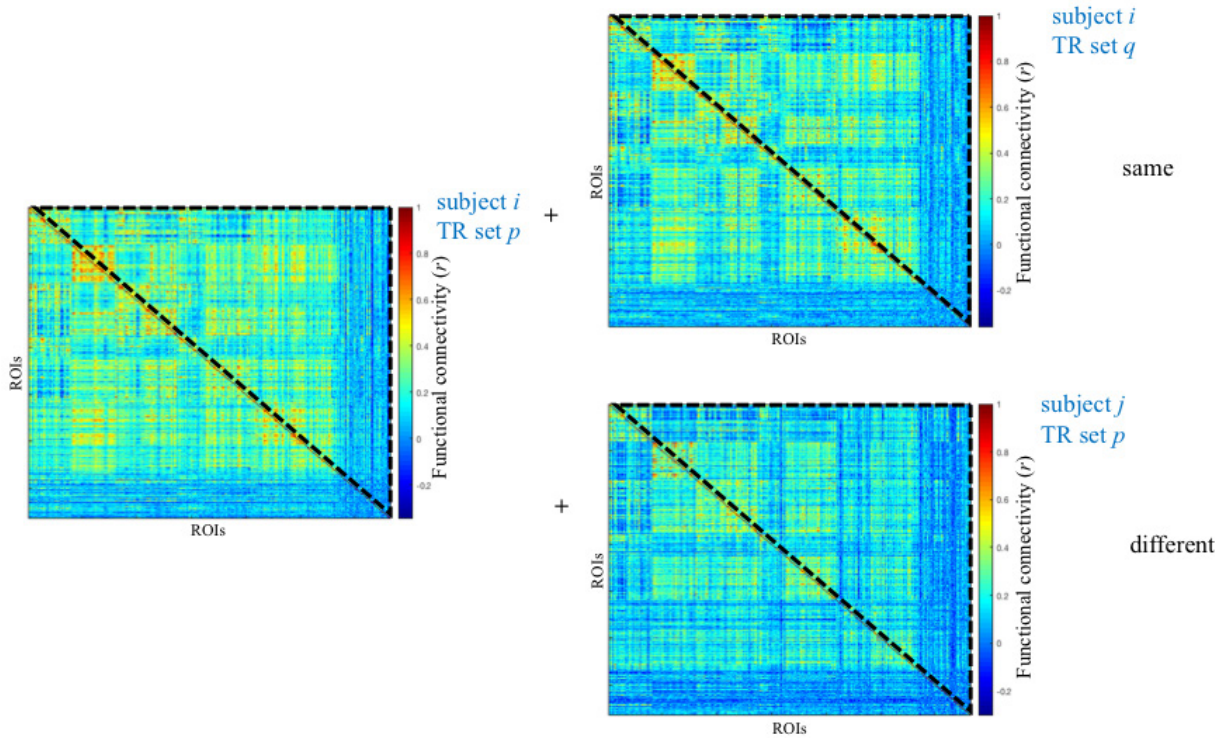

Figure 2. The general strategy: train the model to discriminate between same/different pairs of connectivity profiles. Pairs of connectivity profiles were constructed with a random TR set (i.e., 240 random rfMRI volumes), originating from the same subject (left + top right) or a different subject (left + bottom right).

and x2 originating from the same subject or different subjects (Fig. 2). Only the upper triangular parts of the matrices (without the diagonal) were flattened and used as inputs (Fig. 2, dashed lines).

Because a random set of fMRI volumes was included to create connectivity profiles, time correlations were broken, making the data independent and identically distributed (i.i.d.). This implies that information about one datapoint (rfMRI activity in voxels at a random timepoint) does not give any information about another datapoint (rfMRI activity in voxels at another random timepoint).

Learning schedule. The training set consisted of connectivity profiles constructed with 240 randomly sampled volumes out of a total of 2400, corresponding to the resting state fMRI from two sessions on day one. The validation set consisted of connectivity profiles constructed with 240 randomly sampled volumes out of a total of 2400, corresponding to the resting state fMRI from two sessions on day two.

Weights and biases were randomly initialized and were trained over 200 iterations, using Adam optimizer. The learning rate was set to 10−4 and the minibatch size was 256. Each training iteration, a random minibatch was fixed and pairs of connectivity matrices $x_1^{(i)}$ and $x_2^{(i)}$ were constructed, where i denotes one sample of the minibatch. 50% of the pairs in one minibatch belonged to the same subject $y(x_1^{(i)}, x_2^{(i)}) = 1$ or different subjects $y(x_1^{(i)}, x_2^{(i)}) = 0$.

After every 50 iterations, the model was evaluated for 50 iterations with a validation set. Every evaluation was identical to allow for a good comparison of the model's performance during learning.

Loss function. As a loss function, the Binary Cross-Entropy function was used (Equation 3). This function was chosen because it takes in the probabilities of the binary classes 0 and 1, and outputs a low loss value if the (negative log of the) probability of the true class is high, and conversely, a high loss value if the (negative log of the) probability of the true class is low.

$$L(\mathbf{x}_1, \mathbf{x}_2) = -\frac{1}{N}\sum_{i=1}^{N}\Big[\mathrm{y}(x_1^{(i)}, x_2^{(i)})\log p(x_1^{(i)}, x_2^{(i)}) + (1 - \mathrm{y}(x_1^{(i)}, x_2^{(i)}))\log(1 - p(x_1^{(i)}, x_2^{(i)}))\Big] \quad (3)$$

Decorrelation of final hidden activations. In order to visualize and interpret the learned features as distinct components, it was necessary that the final activations were uncorrelated. To ensure that the final feature vector hm,5 of each twin network m was uncorrelated, a regularization term was imposed on the loss function (Equation 4-5; [Cogswel, Ahmed, Girshick, Zitnick, & Batra, 2016]). To this end, the covariances were minimized, which corresponds to penalizing the norm of covariance matrix C.

$$\text{Final loss function} = L(x_1, x_2) + \lambda \cdot LDeCov \quad (4)$$

with

$$LDeCov^m = ||C||^2_F - ||diag(C)|| \quad (5)$$

where $||.||_F$ is the Frobenius norm. Here, matrix *C* contains the covariances between all pairs of activations from $\mathbf{h}_{m,5}$, divided over N_5. This last division step was performed to ensure a stable *LDeCov* term, independent of the number of final hidden units chosen in the model. This resulted in $LDeCov^1$ for $\mathbf{h}_{1,5}$ for twin network one, and $LDeCov^2$ for $\mathbf{h}_{2,5}$ for twin network two (Equation 5). Last, the mean of the two terms was taken, multiplied with λ, and added to the loss function (Equation 4). Grid search for λ in the range of 0.1 to 1.0 was performed to find the highest value for λ. This resulted in the optimal value λ=1.0.

Visualization of final hidden activations

The activations of the final hidden layer (Fig. 1, l=5) provided five uncorrelated hidden representations of the input connectivity profiles. Note that these final representations in both twin networks would be the same when input x1 and x2 were identical, since both networks share the same learned weight and bias values. To map each learned feature by the model onto a feature of functional connectivity, the following analysis was performed (for complete overview, see Figure 3).

Per subject *i*, the connectivity matrix based on rfMRI data from the complete time course was taken as input $x_m^{(i)}$ for one twin network *m*. The activations of the last fully connected layer represented the highest-level features $\mathbf{h}_{m,5}^{(i,complete)}$. To observe the contribution of each connection (i.e., each matrix' entry), its

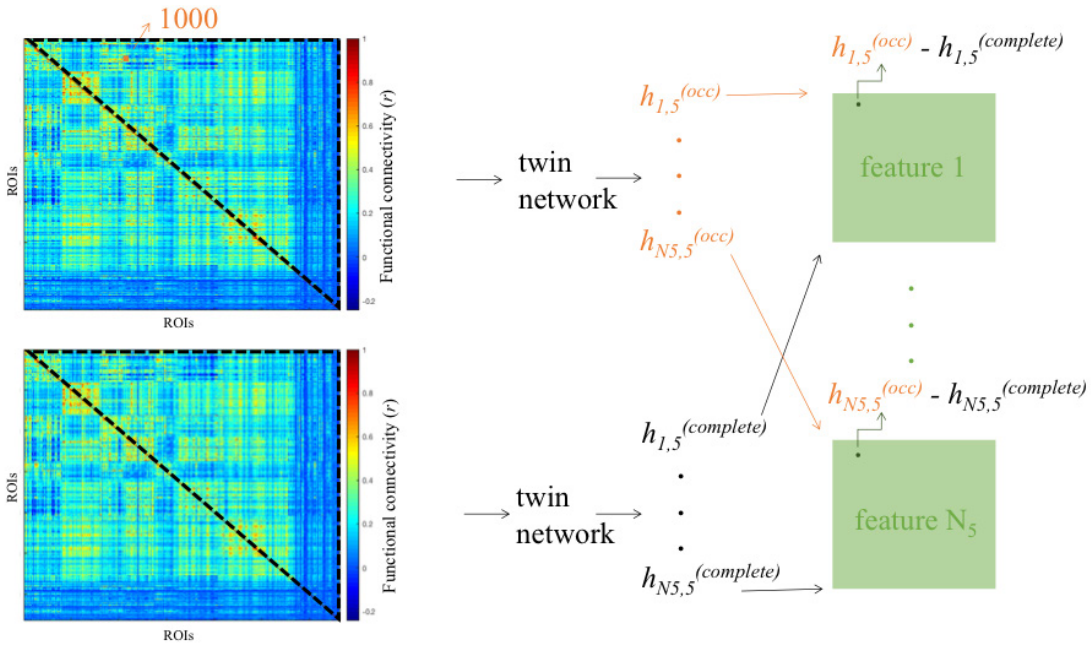

Figure 3. The visualization strategy: Each value of a connectivity profile (connection strength between two ROIs) was set to a high value (1000). As inputs, only the upper flattened triangular parts of matrix were used (indicated with dashed lines), and final hidden activations were observed (indicated with orange). The change in final hidden activations represented the contribution of a specific connection to the activations (indicated in dark green). The changes were stored in five separate matrices, each representing one final hidden activation (indicated in light green). Per matrix, each value represented the contribution of the connection strength between two ROIs to one final hidden activation.

value was set to 1000, and this new connectivity profile was taken as input $x_m^{(i)}$. Similarly, activations of the final hidden layer were observed as $\mathbf{h}_{m,5}^{(i,occ)}$. The difference between hm,5(i,occ) and hm,5(i,complete) represented the contribution of each connection (i.e., functional connectivity between ROIs) to each of the five activations of the last fully connected layer for subject *i*.

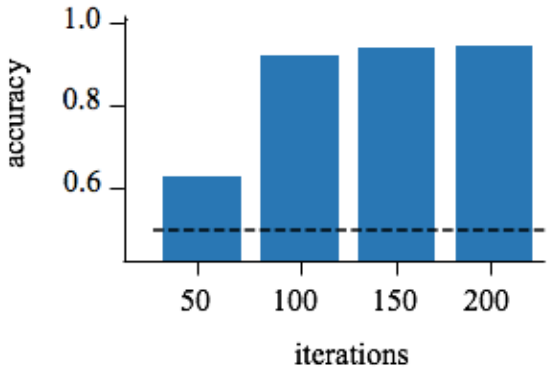

Figure 4. The model's accuracy of the validation set consisting of rfMRI volumes from day two, trained with rfMRI volumes from day one.

The extreme value of 1000 was chosen to ensure an observable change in final activation values. Changing each connection to zero would not give the right contrast, since zero would be a plausible value for a correlation between a connection (i.e., the original value of a connectivity matrix' entry).

This process was repeated for all 996 subjects and then averaged, resulting in five (one per hidden node output) 332 × 332 matrices over all subjects.

Included tests to find relationship with final hidden activations

For all 996 subjects their rfMRI volumes were fed into a trained twin network. Per subject, this resulted in final activation values (Fig. 1, red nodes, last column), representing a 'score' for each final feature learned by the model.

Cognitive performance. First, to explore a link between the final hidden activations and cognitive performance, Pearson correlation coefficients between final activation values and multiple cognitive tests were calculated.

Table 1. Correlations between the three final hidden activations of hidden layer five of a twin network.

output number	0	1	2
0	1	.083	.087
1	.083	1	.056
2	.087	.056	1

This included the following tests from the NIH Toolbox and Penn tests (Elam, 2020), none of them were age adjusted. (Verbal) episodic memory, which involves the acquisition, storage and effortful recall of new information, executive function, which specifically measures cognitive flexibility, fluid intelligence via non-verbal reasoning, language decoding, which measures the reading decoding skill, working memory, which involves information processing and storage and last, spatial orientation via line orientation.

Surface thickness and area size of brain regions. Next, to explore a link between the final hidden activations and brain regions, Pearson correlation coefficients between final activation values and surface thickness and area size of multiple brain regions were calculated. Surface thickness and area size from subjects were obtained as described in Van Essen et al. (2012). Included brain regions were the following cortical regions with clear anatomical landmarks (based on [Hagmann et al., 2008]; Figure 1[3a]): lateral occipital cortex, cuneus, superior parietal cortex, precuneus, inferior parietal cortex, supramarginal gyrus, transverse temporal gyrus, superior temporal cortex, middle temporal gyrus, inferior temporal gyrus, posterior cingulate cortex and superior frontal cortex.

Results

Evaluation of SNN performance

The SNN was trained for 200 iterations in total. After one training block of 50 iterations with rfMRI volumes from day one, there was a validation block of 50 iterations with rfMRI volumes from day two. The accuracy was measured as a percentage of correct classified pairs of a minibatch (256 input pairs, 50% same and 50% different subjects), averaged over 50 iterations (Fig. 4). After one training block the SNN's accuracy was already above

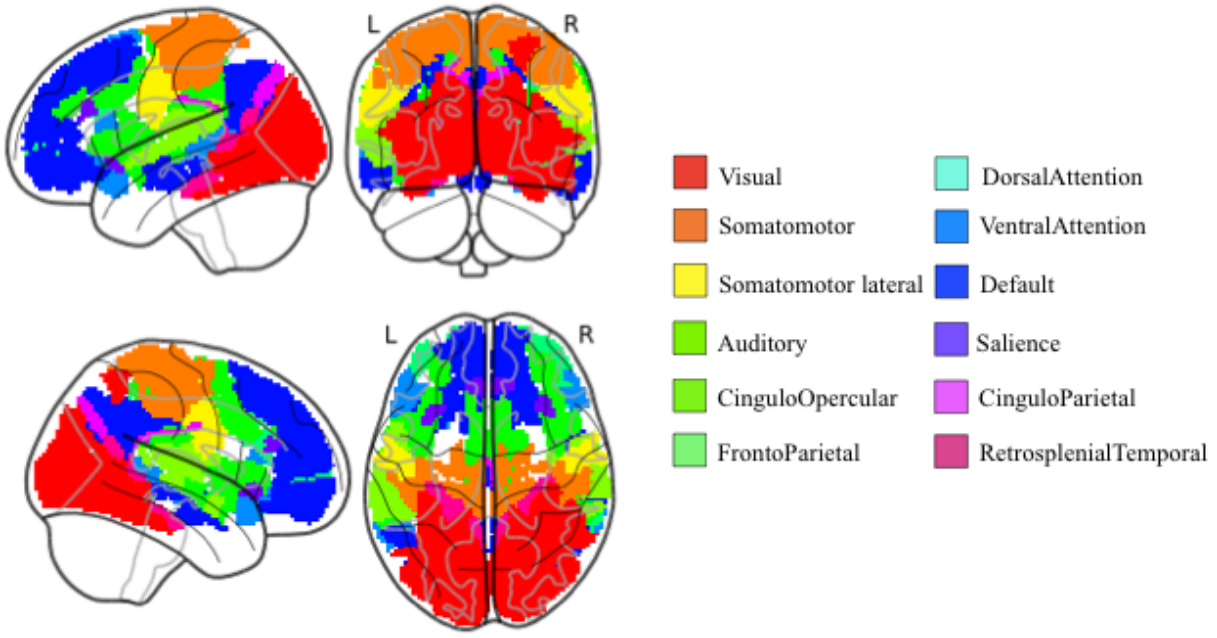

Figure 5. Locations of the large-scale brain networks.

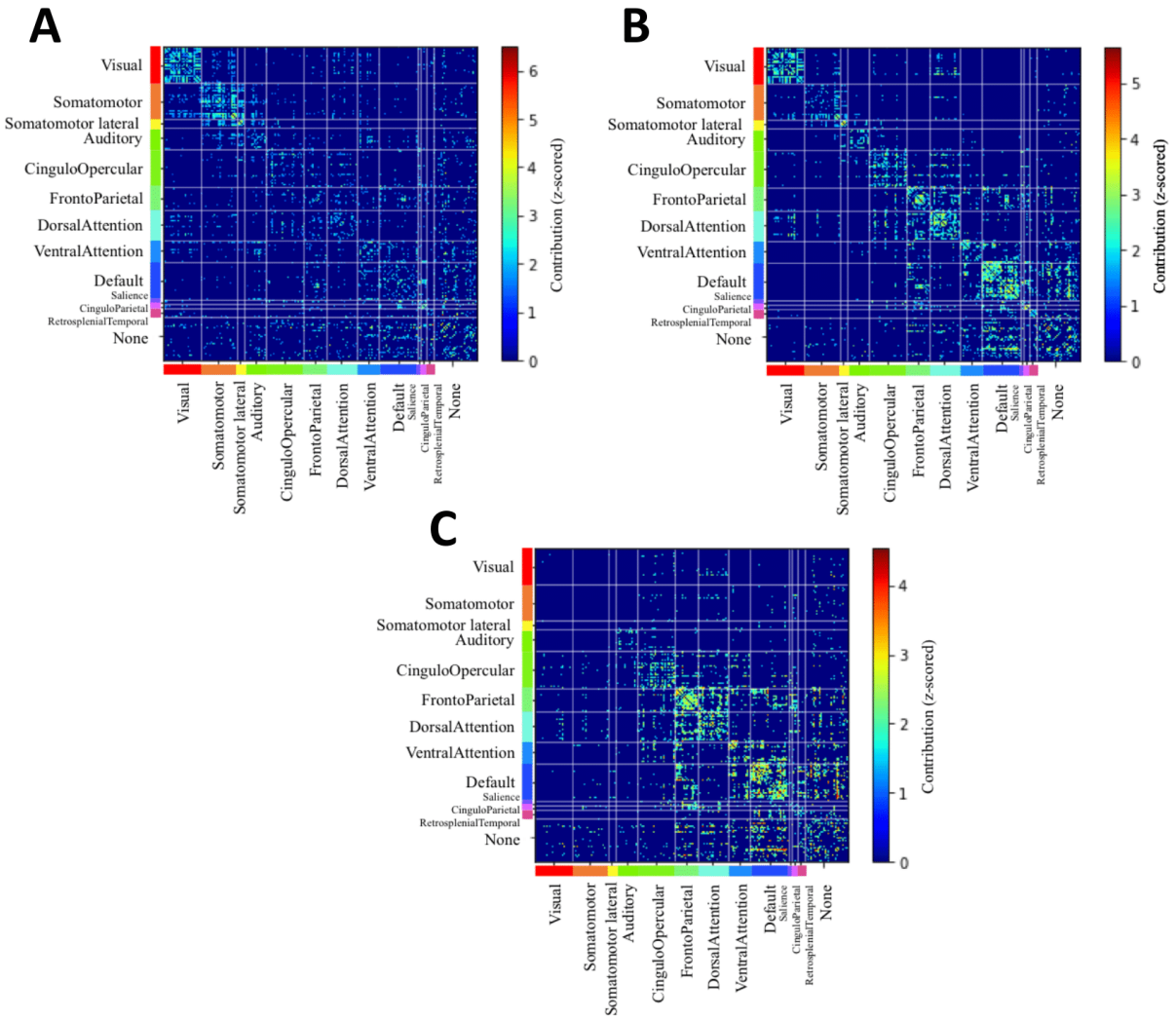

Figure 6. Contribution of each connection between ROIs to the model's final hidden activation values. A. Visualisation of output zero. B. Visualisation of output two. C. Visualisation of output three.

chance level (indicated with dashed line, >50% correct classified). After the second training block, the model already reached excellent accuracy (>90% correct classified). The training phase was considered done when the accuracy did not increase anymore, which was after the fourth training block. This resulted in an accuracy of .96, or 96% correct classified input pairs.

The final hidden layer five of each twin network m had final hidden activations $h_{m,5}$ of size $N_5 = 5$ (Fig. 1, red nodes, last column). However, two of the five final hidden activations were zero for all subjects' connectivity profiles. This resulted from the ReLU activation function that could output a zero (Equation 1). Only three of the five final hidden activations had a non-zero value, therefore, only three outputs are discussed below. To determine whether the final hidden activations had correlations with each other, connectivity profiles from all 996 subjects were fed into one twin network (Fig. 1, red nodes). The three final outputs were uncorrelated (Table 1).

Interpretation of final hidden activations

The used parcellation scheme comprised 332 ROIs (Gordon et al., 2016), distributed over 12 large-scale brain networks (Fig. 5).

The occlusion analysis as previously described (Methods: Visualization of final hidden activations) resulted in a 332 × 332 matrix for each final hidden activation. The brain networks on both axes were sorted (Gratton et al., 2018): first the processing

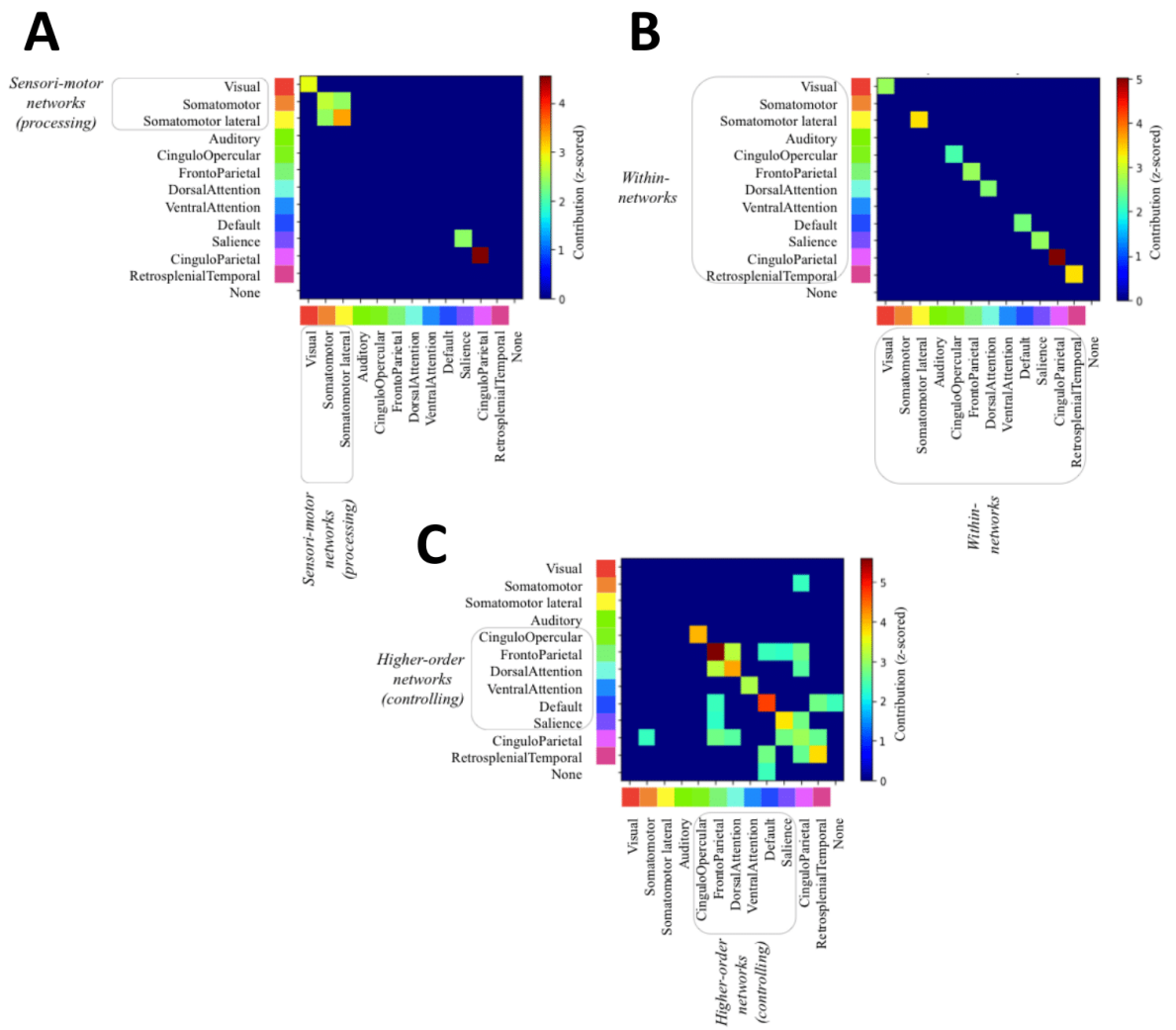

Figure 7. Contribution of each connection between ROIs to the model's final hidden activation values, averaged over all ROIs in a large-scale brain network. A. Visualisation of output zero. B. Visualization of output two. C. Visualization of output three.

networks (Visual, Somatomotor, Somatomotor lateral, Auditory), next the control networks (CinguloOpercular, FrontoParietal, DorsalAttention, VentralAttention, Default), and last the rest (CinguloParietal, RetrosplenialTemporal, None). Values were z-scored with a threshold of z > 1.5 (Fig. 6).

The highest contribution (i.e. greatest change in output value when the connection's value was set to 1000) to output zero (Fig. 6A) were mostly found in connections between ROIs belonging to the Visual, Somatomotor (SM) and Somatomotor lateral (SM-lat) networks. This component describes connections within the visual network and connections within the somatomotor network, but not between the two networks. Furthermore, connections within other specific networks were observed, however, they are less dominant. The most contributing connections between ROIs for output two were not found in a specific set of networks, but roughly all large-scale brain networks had similar contribution to output two (Fig. 6B). Most contributing connections between ROIs of a network were mostly found with ROIs within the same network. The connections between ROIs with the highest contribution to output three (Fig. 6C) were mostly found in the CinguloOpercular (CO), FrontoParietal (FP), DorsalAttention (DA), VentralAttention (VA), Default (DM) and Salience network, with the strongest connections within FP and DMN. Also, less dominant connections between different networks were observed.

Averaging over all ROIs within a large-scale brain network resulted in a 12 × 12 matrix for each final hidden activation, where

all large-scale brain networks were plotted against each other. Values were z-scored with a threshold z > 2.2 (Fig. 7). The three final hidden activation represented distinct components: 1) sensori-motor networks 2) within-networks and 3) higher-order networks.

Evaluation of final hidden activations with cognitive performance

To explore a link between the found components and cognitive performance, the final hidden activations were correlated with scores on multiple cognitive tests from the NIH Toolbox and Penn tests (p-values were Bonferroni corrected for all comparisons, see Methods: Cognitive performance for included tests).

The higher-order networks component (Fig. 7C) correlated with the fluid intelligence score ($r = .16$, $p < .001$) and with the language decoding score ($r = .17$, $p < .001$). Furthermore, correlation was found between the within-networks component (Fig. 7B) and the working memory score ($r = .18$, $p < .001$). The sensori-motor networks component (Fig. 7A) did not show any significant correlations with the included cognitive tests.

Evaluation of final hidden activations with brain regions

To explore a link between the found components and brain regions, the final hidden activations were correlated with surface thickness and area size of multiple brain regions (p-values were Bonferroni corrected for all comparisons, see Methods: Surface thickness and area size of brain regions for included brain regions). Correlations between the higher-order networks component and precuneus surface area ($r = .29$, $p < .001$, Fig. 8, green arrow), superior parietal cortex area ($r = .28$, $p < .001$, Fig. 8, orange arrow), superior frontal cortex area ($r = .28$, $p < .001$, Fig. 8, pink arrow), inferior temporal gyrus area ($r = .27$, $p < .001$, Fig. 8, blue arrow), and middle temporal gyrus area ($r = .27$, $p < .001$, Fig. 8, yellow arrow) were found. The sensori-motor networks component (Fig. 7A) and the within-networks component (Fig. 7B) did not show any significant correlations with the included brain regions.

Discussion

The main aim of this study was to unravel the underlying features of inter-individual variability in functional brain connectivity. We

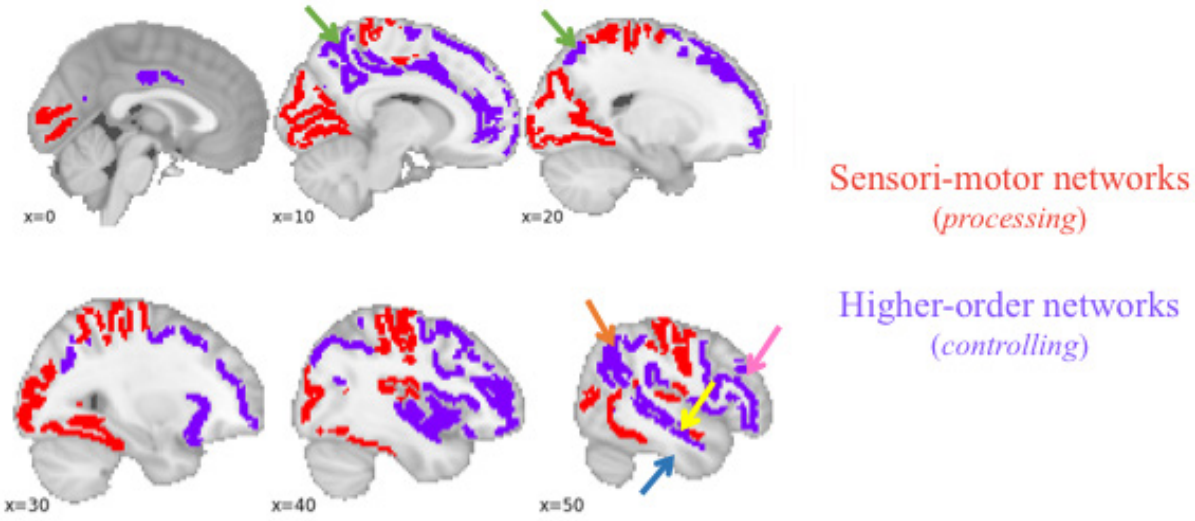

Figure 8. The locations of the ROIs belonging to the sensori-motor networks component (Visual, SM, SM-lat) are shown in red, the locations of the ROIs belonging to the higher-order networks component (CO, FP, DA, VA, DM, and Salience) are shown in purple. Significant correlations between the higher-order networks component and brain regions are indicated with coloured arrows. Indicated brain regions overlap with locations of higher-order networks.

have identified three components of inter-individual variability in functional connectivity, that together can be used to verify the identity of an individual with > 95% accuracy. Each component represented a different set of large-scale brain networks. First, the sensori-motor networks component ('processing' systems [Gratton et al., 2017; Gratton et al., 2018]), consisting of the Visual, SM and SM-lat networks (Fig. 7A). Second, the higher-order networks component ('control' systems [Gratton et al., 2017; Gratton et al., 2018]), consisting of the CO, FP, DA, VA, DM, and Salience networks (Figure 7C). Third, the within-networks component, consisting of inter-networks connections rather than a specific set of networks (Figure 7B).

We demonstrated that it is possible to develop a model that identifies individuals from a large group solely based on functional connectivity. With large amounts of data (resting state fMRI data from 996 subjects acquired on two separate days [Van Essen et al., 2013]), the model reached an accuracy of 96% correct classification. This indicates that intraindividual functional organization is idiosyncratic (Finn et al., 2015), and can be used as a measure to discriminate between individuals. This also indicates that inter-individual variability in functional organization arises from stable factors rather than more transient factors such as ongoing cognition, and can provide meaningful information about individual stable traits (Gratton et al., 2017).

Furthermore, we demonstrated that we can condense a functional connectivity matrix to only three dimensions and still classify individuals with high accuracy. More specifically, the input space of 54946 (representing resting state correlations between predefined ROIs, using the parcellation scheme from Gordon et al. [2016]) was reduced by the model to five final outputs, whereof three were considered relevant. These three relevant final outputs were forced to be uncorrelated. This indicates that inter-individual variability in functional connectivity can be reduced to only three meaningful distinct components that each represent a discriminative feature of inter-individual variability in functional connectivity. This also indicates that discriminating features of inter-individual variability in functional connectivity are easy to interpret because each component maps to a different set of large-scale brain networks.

Each component includes the large-scale brain networks that show the highest contribution to that component (see Figure 7 for included networks per component), which means that functional connectivity in these networks is driving the model to classify subjects' functional connectivity profiles correctly. Therefore, the 'score' for each component (i.e., the activations of final hidden layer five, see Fig. 1, red nodes, last column) could tell something about the nature and underlying mechanisms of the included large-scale brain networks of the component.

First, the three found components could provide meaningful information about an individual's cognitive measures. It is known that inter-individual variability in functional connectivity is relevant to individual differences in behaviour (Finn et al., 2015; Rosenberg et al., 2016; Shen et al., 2017; Tavor et al., 2016). We observed correlations between the higher-order networks component and language decoding and fluid intelligence. This is consistent with previous findings where functional properties of the FP network appeared to be most related to intelligence (Finn et al., 2015; Kanai & Rees, 2011). Furthermore, we observed a correlation between the within-networks component and working memory. The within-networks component represents connections within large-scale brain regions, with connections within the CinguloParietal network yielding the highest contribution to the 'score' for that component. This network comprises the anterior cingulate cortex (aCC), posterior cingulate cortex (pCC), lateral parietal cortex (LPC) and part of the superior frontal cortex (SFC) (see Toro et al., [2008], Fig. 5), which is in line with previous research where it is shown that working memory is associated with the prefrontal cortex and more posterior regions (Cohen et al., 1997).

Next, we observed correlations between the higher-order networks component and multiple area sizes of cortical brain regions (Fig. 8, indicated with arrows). The locations of the areas with significant correlations overlap with the locations of the ROIs of the large-scale brain networks included in the higher-order networks component. This implies

that there is some relationship between the area size of brain regions contributing to the higher-order networks component and the 'score' for the higher-order networks component. This suggests that anatomical features such as grey matter might relate to the inter-individual discriminating features of functional connectivity and that amount of grey matter is associated with the strength of connection within these regions.

In this study, three distinct components that serve as underlying discriminating feature of inter-individual variability in functional connectivity are discussed. Further work should focus on the generalization of the model, and investigate how robust the identified components as underlying features of inter-individual variability in functional connectivity are.

First, we used patterns of brain activity in resting state, but patterns in other states such as a task state could also be included. Next, we used fMRI activity acquired on two consecutive days, but fMRI activity obtained within a longer time span could also be included. Next, the large subject-level specific variability in functional connectivity demands for the use of a large data set such as the HCP data set, therefore, development of a future model could use another large data set. Last, our used model had five final hidden activations, where only three of them had a non-zero value. Decreasing the number of final hidden nodes caused the model's performance to drop. We do not understand why this happened and how to resolve this. Ideally, the model would use all the final hidden nodes when discriminating between subjects' functional connectivity profiles.

To conclude, we showed that the input space of a complete FC profile can be reduced to only three components that the model uses to successfully distinguish individuals' FC profiles from each other. This indicates that functional organization is individual specific, and arises from stable factors. Furthermore, we showed that each discriminating feature of functional connectivity represented a meaningful distinct component that maps to a different set of large-scale brain networks.

References

Betzel, R., Bertolero, M.A., Gordon, E., Gratton, C., Dosenbach, N., & Bassett, D. (2019). The community structure of functional brain networks exhibits scale-specific patterns of inter- and intra-subject variability. *NeuroImage, 202*, 115990.

Biswal, B., Zerrin Yetkin, F., Haughton, V., & Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine, 34*, 537-541.

Cogswel, M., Ahmed, F., Girshick, R., Zitnick, L., & Batra, D. (2016). Reducing overfitting in deep neural networks by decorrelating representations. *ICLR*.

Cohen, J., Perlstein, W., Braver, T., Nystrom, L., Noll, D., Jonides, J., & Smith, E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386*, 604-608.

Elam, J. (2020). Hcp data dictionary public (11 June, 2020).

Finn, F., Shen, X., Scheinost, D., Rosenberg, M., Huang, J., Chun, M., ... & Constable, R. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience, 18*, 1664–1671.

Fransson, P. (2005). Spontaneous low-frequency bold signal fluctuations: An fmri investigation of the resting-state default mode of brain function hypothesis. *Human Brain Mapping, 26*, 15-29.

Glasser, M., Sotiropoulos, S., Wilson, J., Coalson, T., Fischl, B., Andersson, J., ... & Jenkinson, M. (2013). The minimal processing pipelines for the human connectome project. *NeuroImage, 80*, 105-124.

Gordon, E., Laumann, T., Adeyemo, B., Huckins, J., Kelley, W., & Petersen, S. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex, 26*(1), 288-303.

Gordon, E., Laumann, T., Gilmore, A., Newbold, D., Greene, D., Berg, J., ... & Dosenbach, U. (2017). Precision functional mapping of individual human brains. *Neuron, 95*(4), 727-807.

Gratton, C., Laumann, T., Nielsen, A., Greene, D., Gordon, E., Gilmore, A., ... & Petersen,

S. (2017). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron, 98*(2), 439-452.

Gratton, C., Sun, H., & Petersen, S. (2018). Control networks and hubs. *Psychophysiology, 55*.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C., Wedeen, V., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology, 6*(7).

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience, 12*, 231-242.

Kietzmann, T., McClure, P., & Kriegeskorte, N. (2018). Deep neural networks in computational neuroscience. *BioRxiv*.

Koch, G. (2015). Siamese neural networks for one-shot image recognition.

Rosenberg, M., Finn, E., Scheinost, D., Papademetris, X., Shen, X., Constable, R., & Chun, M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience, 19*, 165-171.

Seitzman, B., Gratton, C., Laumann, T., Gordon, E., Adeyemo, B., Dworetsky, A., ... & Petersen, S. (2019). Trait-like variants in human functional brain networks. *Proceedings of the National Academy of Sciences, 116*(45), 22851-22861.

Shen, X., Finn, E., Scheinost, D., Rosenberg, M., Chun, M., Papademetris, X., & Constable, R. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols, 12*, 506-518.

Smith, S., Vidaurre, D., Beckmann, C., Glasser, M., Jenkinson, M., Miller, K., ... & Van Essen, D. (2013). Functional connectomics from resting-state fmri. *Trends in Cognitive Neuroscience, 17*(12), 666-682.

Storrs, K., & Kriegeskorte, N. (2019). Deep learning for cognitive neuroscience. *ArXiv*.

Tavor, I., Parker Jones, O., Mars, R., Smith, S., Behrens, T., & Jbabdi, S. (2016). Task free mri predicts individual differences in brain activity during task performance. *Science, 352*(6282), 216-220.

Toro, R., Fox, P., & Paus, T. (2008). Functional coactivation map of the human brain. *Cerebral Cortex, 18*(11), 2553-2559.

Van Essen, D., Glasser, M., Dierker, D., Harwell, J., & Coalson, T. (2012). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral Cortex, 22*(10), 2241-2262.

Van Essen, D., Smith, S., Barch, D., Behrens, T., Yacoub, E., & Ugurbil, K. (2013). The wu-minn connectome project: An overview. *NeuroImage, 80*, 62-79.

Xu, K., Lei Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., ... & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. *Proceedings of the 32nd International Conference on Machine Learning, 37*.