nature research

Corresponding author(s): Maxwell Bertolero

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	All data are available through the Human Connectome Project
Data analysis	All analysis code is written in python and is freely available at www.github.com/mb3152/hcp_performance

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

• The datasets analyzed during the current study are available in the Human Connectome Project database: https://www.humanconnectome.org

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	All analyses were quantitative measures of individual differences in human brain functional connectivity measured with fMRI
Research sample	All data analyzed is included in the Human Connectome Project S500 Release
Sampling strategy	No sample size calculation was executed, as the sample size is very large and cross validation techniques were employed
Data collection	All data was independently collected by the Human Connectome Project. We analyzed all subjects that were part of the S500 release. Extensive details can be found at: https://www.humanconnectome.org/storage/app/media/documentation/s500/hcps500meg2releasereferencemanual.pdf
Timing	All data was collected between 2009 and 2014
Data exclusions	All subjects that were part of the S500 data release and completed the fMRI sessions were included. This results in 476 unique subjects, 475, 458, 472, and 474 in the working memory, relational, language/math and social task, respectively.
Non-participation	For each fMRI task analyzed, only subjects that completed the task and resting-state scans were analyzed. This results in 25, 42, 28, and 26 subjects that were not included in the working memory, relational, language/math and social task, respectively.
Randomization	Subjects were not placed in different groups

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above; also, extensive details regarding the demographics and screening protocol can be found here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3724347/
Recruitment	All subjects were recruited independently by the Human Connectome project. Details can be found here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3724347/

Magnetic resonance imaging

Experimental design

Design type	All analyses were executed on the preprocessed time series. No event related designs or block designs were used.

Design specifications

The following descriptions for each task have been adapted for brevity from the Human Connectome Project Manual.
Working Memory. The category specific representation task and the working memory task are combined into a single task paradigm. Participants were presented with blocks of trials that consisted of pictures of places, tools, faces and body parts (non-mutilated parts of bodies with no "nudity"). Within each run, the 4 different stimulus types were presented in separate blocks. Also, within each run, 1⁄2 of the blocks use a 2-back working memory task and 1⁄2 use a 0-back working memory task (as a working memory comparison). A 2.5 second cue indicates the task type (and target for 0-back) at the start of the block. Each of the two runs contains 8 task blocks (10 trials of 2.5 seconds each, for 25 seconds) and 4 fixation blocks (15 seconds). On each trial, the stimulus is presented for 2 seconds, followed by a 500 ms inter-task interval (ITI).
Gambling. Participants play a card guessing game where they are asked to guess the number on a mystery card (represented by a "?") in order to win or lose money. Participants are told that potential card numbers range from 1-9 and to indicate if they think the mystery card number is more or less than 5 by pressing one of two buttons on the response box. Feedback is the number on the card (generated by the program as a function of whether the trial was a reward, loss or neutral trial) and either: 1) a green up arrow with "$1" for reward trials, 2) a red down arrow next to -$0.50 for loss trials; or 3) the number 5 and a gray double headed arrow for neutral trials. The "?" is presented for up to 1500 ms (if the participant responds before 1500 ms, a fixation cross is displayed for the remaining time), following by feedback for 1000 ms. There is a 1000 ms ITI with a "+" presented on the screen. The task is presented in blocks of 8 trials that are either mostly reward (6 reward trials pseudo randomly interleaved with either 1 neutral and 1 loss trial, 2 neutral trials, or 2 loss trials) or mostly loss (6 loss trials pseudo- randomly interleaved with either 1 neutral and 1 reward trial, 2 neutral trials, or 2 reward trials). In each of the two runs, there are 2 mostly reward and 2 mostly loss blocks, interleaved with 4 fixation blocks (15 seconds each).
Motor. Participants are presented with visual cues that ask them to either tap their left or right fingers, or squeeze their left or right toes, or move their tongue to map motor areas. Each block of a movement type lasted 12 seconds (10 movements), and is preceded by a 3 second cue. In each of the two runs, there are 13 blocks, with 2 of tongue movements, 4 of hand movements (2 right and 2 left), and 4 of foot movements (2 right and 2 left). In addition, there are 3 15-second fixation blocks per run. This task contains the following events, each of which is computed against the fixation baseline.
Language & Math. The task consists of two runs that each interleave 4 blocks of a story task and 4 blocks of a math task. The lengths of the blocks vary (average of approximately 30 seconds), but the task was designed so that the math task blocks match the length of the story task blocks, with some additional math trials at the end of the task to complete the 3.8 minute run as needed. The story blocks present participants with brief auditory stories (5-9 sentences) adapted from Aesop's fables, followed by a 2-alternative forced- choice question that asks participants about the topic of the story. For example: "after a story about an eagle that saves a man who had done him a favor, participants were asked, "Was that about revenge or reciprocity?" The math task also presents trials aurally and requires subjects to complete addition and subtraction problems. The trials present subjects with a series of arithmetic operations (e.g., "fourteen plus twelve"), followed by "equals" and then two choices (e.g., "twenty- nine or twenty- six"). Participants push a button to select either the first or the second answer. The math task is adaptive to try to maintain a similar level of difficulty across participants.
Social (Theory of Mind). Participants were presented with short video clips (20 seconds) of objects (squares, circles, triangles) that either interacted in some way, or moved randomly on the screen. After each video clip, participants judge whether the objects had a mental interaction (an interaction that appears as if the shapes are taking into account each other's feelings and thoughts), Not Sure, or No interaction (i.e., there is no obvious interaction between the shapes and the movement appears random). Each of the two task runs has 5 video blocks (2 Mental and 3 Random in one run, 3 Mental and 2 Random in the other run) and 5 fixation blocks (15 seconds each).
Relational. The stimuli are 6 different shapes filled with 1 of 6 different textures. In the relational processing condition, participants are presented with 2 pairs of objects, with one pair at the top of the screen and the other pair at the bottom of the screen. They are told that they should first decide what dimension differs across the top pair of objects (differed in shape or differed in texture) and then they should decide whether the bottom pair of objects also differ along that same dimension (e.g., if the top pair differs in shape, does the bottom pair also differ in shape). In the control matching condition, participants are shown two objects at the top of the screen and one object at the bottom of the screen, and a word in the middle of the screen (either "shape" or "texture"). They are told to decide whether the bottom object matches either of the top two objects on that dimension (e.g., if the word is "shape", is the bottom object the same shape as either of the top two objects. For both conditions, the subject responds yes or no using one button or another. For the relational condition, the stimuli are presented for 3500 ms, with a 500 ms ITI, and there are four trials per block. In the matching condition, stimuli are presented for 2800 ms, with a 400 ms ITI, and there are 5 trials per block. Each type of block (relational or matching) lasts a total of 18 seconds. In each of the two runs of this task, there are 3 relational blocks, 3 matching blocks and 3 16-second fixation blocks.

Behavioral performance measures

All performance measures were chosen a priori. In the working memory tasks, we used the mean accuracy across all n-back conditions (face, body, place, tool). In the relational task, we used mean accuracy across both the matching and the relational conditions. For the language task, we took the maximum difficulty level that the subject achieved across both the math and language conditions. We did not use accuracy, because the task varies in difficulty based on how well the subject is doing, making accuracy an inaccurate measure of performance for these tasks. For the social task, given that almost all subjects correctly identified the social interactions as social interactions, we used the percentage of correctly identified random interactions.

Acquisition

Imaging type(s)

Functional

Field strength

3T

Sequence & imaging parameters

Sequence: Gradient-echo EPI
TR: 720 ms

TE: 33.1 ms
flip angle: 52 deg
FOV: 208x180 mm (RO x PE)
Matrix: 104x90 (RO x PE)
Slice thickness: 2.0 mm; 72 slices; 2.0 mm isotropic voxels
Multiband factor: 8
Echo spacing: 0.58 ms
BW: 2290 Hz/Px

Task , Runs, Frames per run, Run Duration (min:sec)
REST (Resting-state), 4, 1200, 14:33
Working Memory, 2, 405, 5:01
Gambling, 2, 253, 3:12
Motor, 2, 284, 3:34
Language, 2, 316, 3:57
Social Cognition, 2, 274, 3:27
Relational Processing, 2, 232, 2:56
Emotion Processing, 2,176, 2:16

Area of acquisition: Whole Brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software: Analysis of Functional NeuroImages (AFNI)

Normalization: Registration of the T1 to atlas space includes an initial volumetric registration to MNI152 space using FSL's linear FLIRT tool, followed by the nonlinear FNIRT algorithm.

Normalization template: MNI152

Noise and artifact removal: Data and Preprocessing We used fMRI data from the Human Connectome Project33 S500 release. For the task-based fMRI data, Analysis of Functional NeuroImages (AFNI) was used to preprocess the images50. The AFNI command 3dTproject was used, passing the mean signal from the cerebral spinal fluid mask, the white matter mask, the whole brain signal, and the motion parameters to the "-ort" options, which removes these signals via linear regression. Within AFNI, the "-automask" option was used to generate the masks. The "-passband 0.009 0.08" option, which removes frequencies outside of 0.009 and 0.08, was used. Finally, the "-blur 6" option, which smooths the images (inside the mask only) with a 6 mm FWHM filter after the time series filtering. Given the short length of the Emotion task (176 frames; Resting-State:1200, Social: 274, Relational:232, Motor:284, Language:316, Working Memory:405, Gambling:253) it was not included in our analyses. For the fMRI data collected at rest, we used the images that were previously preprocessed by the Human Connectome Project with ICA- FIX. We also used the AFNI command 3dBandpass to further preprocess these images. We used it to remove the mean whole brain signal and frequencies outside 0.009 and 0.08 (explicitly, "-ort whole_brain_signal.1D -band 0.009 0.08 -automask"). We did not regress out stimulus or task effects from the time series of each node, because how nodes' low frequency oscillations respond to stimulus or task effects is meaningful. Moreover, other investigators have noted that task effect regression has minimal effects.

Volume censoring: As subject motion during fMRI can impact functional connectivity estimates and has been shown to bias brain-task performance relationships, performance prediction analyses were executed with scrubbing executed on frames with frame-wise displacement greater than 0.2 millimeters, including the frame before and after the movement. Frame-wise displacement measures movement of the head from one volume to the next, and was computed as the sum of the absolute values of the differentiated rigid body realignment estimates (translation and rotation in x, y, and z directions) at every time point with rotation values evaluated with a radius of 50 mm(52). Frames were removed after all preprocessing was executed. Subjects with more than 75 percent of frames removed were not analyzed. Moreover, we executed all analyses after regressing out mean frame-wise displacement from the task performance values (Extended Data Figure 1).

Statistical modeling & inference

Model type and settings: Functional Connectivity

Effect(s) tested: Pearson r correlations between all ROIs / nodes

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s)	The Power atlas (34) was used to define the 264 nodes in our graph because it was the only atlas that met all of the following requirements: (1) Given that the homogeneity of nodes in this atlas is high and they do not share physical boundaries, it will not overestimate the local connectivity of regions, (2) it is the only atlas that is defined based both on functional connectivity and studies of task activations making it optimal for our current analyses, (3) it accurately divides nodes into communities observed with other approaches (e.g., at the voxel level), and this division has been used in many studies(8,12,34,53). A canonical division of nodes into communities aides in the interpretation and generalizability of our results. It can be found at: http://www.nil.wustl.edu/labs/petersen/Resources_files/Consensus264.xls. Moreover, we used this division to calculate within and between community edge weight changes across subjects. (4) It has anatomical coverage of cortical, subcortical, and cerebellar regions.
Statistic type for inference (See Eklund et al. 2016)	Each Pearson r correlation was calculated between ROIS
Correction	No multiple comparisons were calculated for the functional connectivity measurements; however, graphs were thresholded to retain only the strongest 5-15 percent of connections / correlations.

Models & analysis

n/a	Involved in the study
☐	☒ Functional and/or effective connectivity
☐	☒ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	Pearson r correlation
Graph analysis	Subject level weighted thresholded graphs were used, with the weights normalized to sum to an identical value across subjects. Modularity, the participation coefficient, and the within community strength were all calculated for each individual.
Multivariate modeling and predictive analysis	A predictive multilayer perceptron model (three layers (enough for non-linear relationships), eight neurons (one per feature) in each layer) was used to predict subjects' task performance. Known as deep neural networks, these predictive models are constructed by tuning the weights between neurons and layers to achieve the most accurate relationship between the features (input) and the value the model is trying to predict (output). The predictive model's features (n=8) captured how well subjects' nodes' diversity and locality, network connectivity (i.e., edge weights in the network), and modularity (Q) are optimized for the performance of a task. For example, for the feature that captures how optimized the diversity of subjects' nodes' are for task performance, for each node, the Pearson r across subjects between that node's participation coefficients (which measures diversity) and task performance values was calculated. The feature, then, for a given subject, is the Pearson r across nodes between those r values for each node and each node's participation coefficient in that subject, representing how optimized the diversity of that subject's nodes' are for performance in the task. The same procedure is executed for locality (using the within community strengths) and edge weights. Finally, the Q values of the network are included in the model. The predictive model was fit for each of the four cognitive tasks that subjects performed in the Human Connectome Project for which performance was measured (Working Memory, Relational, Language and Math, Social tasks; see Methods for task performance measures). Each predictive model was fit to the subjects' networks constructed during the performance of each task as well as the resting state (four features from each). The inclusion of the resting-state and the cognitive task state allowed the model to capture the subjects' so-called intrinsic network states as well as the subjects' task driven network states. Using a leave-one- out cross-validation procedure, the features were constructed and the model was fit with data from all subjects except one. The predictive model was then used to predict the left-out subject's task performance. To test the accuracy of the model, the Pearson r between the observed and predictive performance of each subject was calculated. The model was executed with scikit-learn in python. A random state was initiated for each run. The default parameters were used: sklearn.neural_network.MLPRegressor(hidden_layer_sizes=(8,12,8,12), activation='relu', solver='adam', alpha=0.0001, batch_size='auto', learning_rate='constant', learning_rate_init=0.001, power_t=0.5, max_iter=200, shuffle=True, tol=0.0001, verbose=False, warm_start=False, momentum=0.9, nesterovs_momentum=True, early_stopping=False, validation_fraction=0.1, beta_1=0.9, beta_2=0.999, epsilon=1e-08) Statistical Methods The number of subjects was determined by the number of subjects released by the Human Connectome Project at the start of the analyses. As this dataset represented the largest dataset of its kind at that time and the number of subjects is greater than many similar analyses47, no power analysis was computed. Total N=Working Memory: 475, Gambling: 473, Relational:458, Motor:475, Language & Math:472, Social:474, Resting State: 476. However, as we only analyzed subjects with both Resting-State and the task scans, N=Working Memory:473, Relational:457, Language & Math:471, Social:473. This results in a unique N=476 across tasks, in that 476 different subjects had a resting state scan and at least one task scan. As scrubbing (which removes frames with large head motion) can cause too many frames to be removed from the time series, subjects with less than 75 percent of remaining frames were not included in the analyses that implemented scrubbing; thus, for analyses using scrubbed data, N=Working Memory:351,

Relational:335, Language & Math:348, Social:358.

All confidence intervals (CI) are reported with alpha=0.05. For Pearson r correlation coefficients CIs, the interval of r values is given by Fisher transforming r to z, computing the interval, and then Fisher reverse transforming the z intervals back to r intervals. For t-tests, the confidence interval represents the largest and smallest differences in means across the two distributions. For all t-tests, distributions were confirmed as normal (p<1e-3) or exhibiting no significant evidence as not normal (k2>0.0) using D'Agostino and Pearson's omnibus test k2. All p values are two sided tests.

All p values that are part of a family of tests are Bonferroni corrected for multiple comparisons. For example, we test if two tasks' hub and network structures are similarly optimal for the same subject measures, testing across a large number of subject measures. In this case, we applied a Bonferroni correction to the p-values to determine whether the effect remained true for particular subject measures. Here, the number of tests is equal to the number of subject measures, 47. Individual subject networks were built independently for each task and task performance is different for each task. Thus, these tests are not strictly in the same family. However, to be conservative, we still Bonferroni corrected these p-values. In these cases, the family size is either 4 or 7, depending on the number of tasks analyzed. Unless otherwise stated, all p values are Bonferroni corrected.

Many statistical tests are calculated here without reported p values. For example, Pearson r values are used to calculate functional connectivity. Here, only the r values are of interest—more precisely, individual differences in the r values across subjects, and how these differences relate to individual differences in cognition. This treatment of multiple comparisons in the context of functional connectivity and individual differences in cognition is common and recommended47,63. We extend this notion to other analyses here as well. For example, we use the Pearson correlation coefficient r to compare how well different nodes' participation coefficients across subjects explain variance in network modularity or task performance (the diversity facilitated modularity and performance coefficients). In these cases, we relate these r-values to other measures, and are only concerned with how these r-values explain another distributions' variance (here, we find a positive correlation between these r-values and a node's mean participation coefficient across subjects). We are not concerned with the statistical significance any particular r-value as estimated by the p-value. We care about the distribution of r-values, not the distribution of p-values, and we do not make any claims about any single r-value. Thus, the p-values are neither reported nor corrected for multiple comparisons. This is precisely how functional connectivity is treated statistically.