Supporting Information for

How much data do we need? Lower bounds of brain activation states to predict human cognitive ability

Maren H. Wehrheim[a,b], Joshua Faskowitz[c], Olaf Sporns[c], Christian J. Fiebach[a,d], Matthias Kaschube[b,e,1], Kirsten Hilger[a,f,1,*]

[a] Department of Psychology, Goethe University Frankfurt, D-60323 Frankfurt am Main, Germany
[b] Department of Computer Science, Goethe University Frankfurt, D-60325 Frankfurt am Main, Germany
[c] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405,
[d] Brain Imaging Center, Goethe University, D-60528 Frankfurt am Main, Germany
[e] Frankfurt Institute for Advanced Studies, D-60438 Frankfurt am Main, Germany
[f] Department of Psychology I, Julius Maximilian University, D-97070 Würzburg, Germany

[1] these authors share senior authorship

[*]Corresponding author:
Kirsten Hilger
Email**:** kirsten.hilger@uni-würzburg.de

This PDF file includes:

Figures S1 to S5
Table S1 to S3
SI References

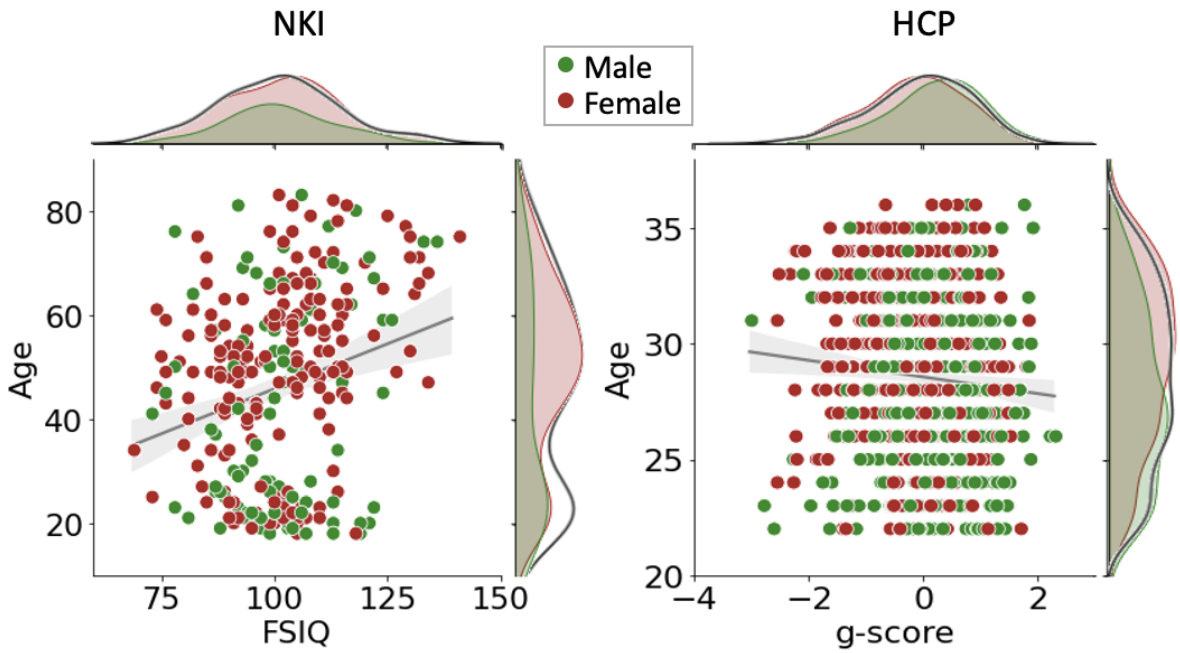

Fig. S1. Distribution of intelligence test scores (FSIQ in the main sample, NKI, 1, and latent *g*-factor in the replication sample, HCP, 2) and age in both data sets. In the main sample (NKI), intelligence test scores were weakly correlated with age (r = .24; p < .001) with a mean (standard deviation) of 101.44 (13.24) and 46.68 (18.54) for intelligence and age, respectively. Within the male and female subgroup, the test scores and age were correlated with r = .18 (p = .07) and r = .29 (p < .001), respectively. Intelligence tests scores had a mean of 101.46 (101.43) and standard deviation of 12.49 (13.63) for the male (female) subgroup. Age was less similarly distributed between the gender subgroups with a mean of 42.11 (49.13) and a standard deviation of 20.20 (17.09) for the male (female) subgroup. Right: Age and *g*-score were weakly negatively correlated in the replication sample (HCP: r = -.09; p = .014). For the *g*-scores and age we observed a mean (standard deviation) of 0 (.89) and 28.55 (3.72), respectively. Within the male and female subgroup, intelligence and age were correlated with r = .01 (p = .78) and r = -.09 (p = .06), respectively. The *g*-score showed a mean of .19 (-.17) and standard deviation of .86 (.87) for the male (female) subgroup. For age we observed a mean of 27.63 (29.36) and standard deviation of 3.63 (3.61) for the male (female) subgroup.

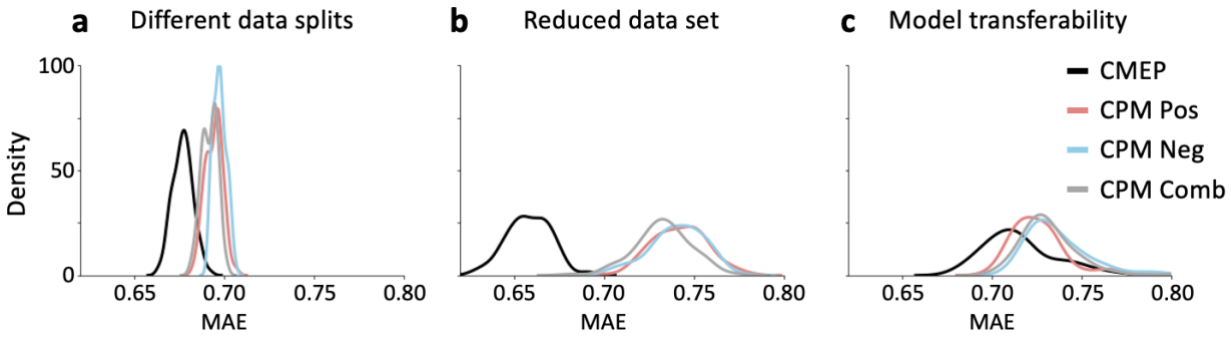

Fig. S2. Superior prediction robustness of covariance maximizing eigenvector-based predictive modelling (CMEP) relative to connectome-based predictive modelling (CPM) in the replication sample. Prediction performances (mean squared error, MAE, between observed and predicted intelligence scores) from static (time-averaged) connectivity were compared via three validity analyses between CMEP and CPM (3, 4). All analyses were conducted for CMEP (black, all brain connections), and three CPM prediction pipelines based on positive connections (light red), negative connections (light blue), and a combination of both (light gray, all connections). (a) Robustness across different data set splits. Data were randomly (100 times) split into 10 folds for cross validation. (b) Robustness across different sample sizes. Within stratified 10-fold cross-validation, the training sample was randomly (100 times) reduced to 10% of the original test-sample size. (c) Transferability of the models to a new data set. Models were trained on the replication sample (HCP) and tested on the primary sample (NKI). Both samples were parcellated into the 114 nodes schemata (5) and all intelligence scores were first standardized and then after prediction mapped back to the original scale for better comparability. The training data were randomly bootstrapped (100 times) to account for different compositions of the training data set.

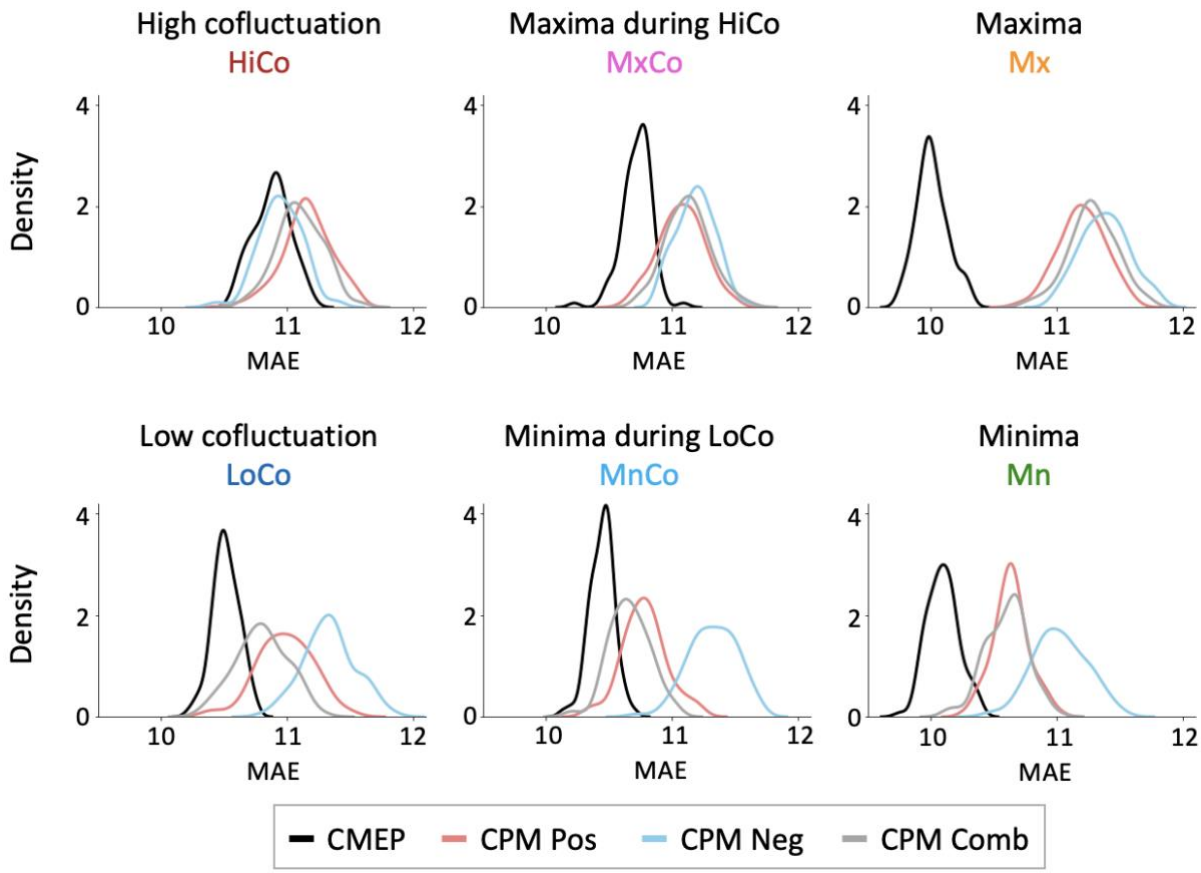

Fig. S3. Prediction robustness of Covariance Maximizing Eigenvector-Based Predictive Modelling (CMEP) in contrast to connectome-based predictive modelling (CPM) for all six different connectivity states. Prediction results are evaluated with the mean absolute error (MAE) between predicted and observed intelligence scores (FSIQ; WASI, 6). Prediction features were derived from one of six different connectivity states (see Fig. 1). Robustness is operationalized as the empirical distribution of prediction performances resulting from 100 different cross-validations splits (10-fold). CMEP (black) is implemented as described in the Methods section and illustrated in Fig. 2. In CPM (3, 4) positive and negative connectivity strengths are calculated as the sum over all functional connections that are significantly positively (light red) or negatively (light blue) correlated with intelligence above a given threshold (here: $p < .001$). These positive and negative connectivity strengths serve separately as features to fit a linear regression model to predict intelligence. Note that as CMEP does not differentiate between positive and negative functional brain connectivity, we additionally fitted CPM with positive *and* negative connectivity strengths (CPM Comb, light gray).

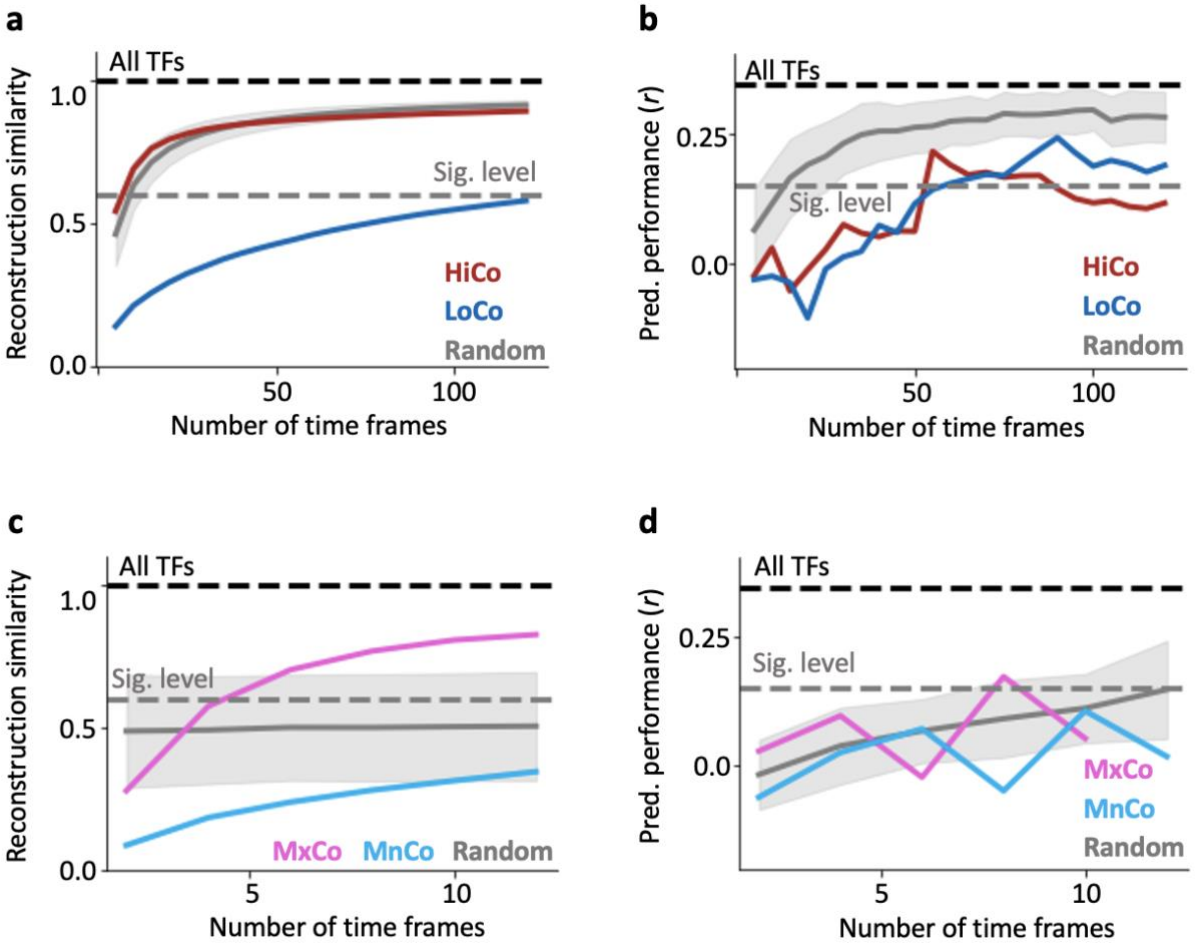

Fig. S4. Reconstruction similarity and prediction performance for different numbers of time frames. Reconstruction similarity (Pearson correlation *r* between reconstructed connectivity and static functional connectivity; dashed lines) and mean prediction performance (Pearson correlations *r* between predicted and observed intelligence scores; FSIQ; WASI, 6; solid lines) across 100 different 10-fold cross validation splits. (a, b): Functional connectivity was reconstructed from time frames with decreasing strength of cofluctuations (as indicated by RSS values) starting with the five time frames of highest cofluctuation (red) and from time frames with increasing strength of cofluctuations starting with the five time frames of lowest cofluctuation (blue). (c, d): Functional connectivity was reconstructed from only the maxima/minima within the highest/lowest cofluctuation time series (pink/light blue). For further comparability a null model (gray) was generated from 100 random time frames that were uniformly selected. The translucent band around the mean prediction performance of the random selections indicates the standard deviation of prediction results.

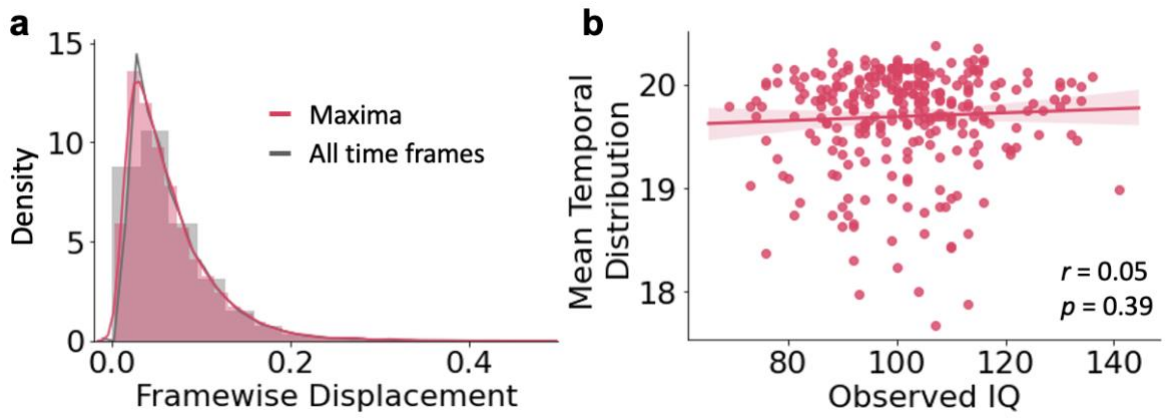

Fig. S5. Control analyses results. (**a**) Relative independence of the 44 highest maxima connectivity states from time frames of high head motion. Empirical distribution of in-scanner head motion (operationalized as mean framewise displacement) for the complete time series (gray) and the 44 highest maxima connectivity states only (pink). (**b**) Associations between intelligence (FSIQ; WASI, 6) and the mean temporal distribution of 44 highest maxima connectivity states. Each dot represents one subject and the best-fit regression line is highlighted with a translucent band corresponding to the 95%-confidence interval. *r*, Pearson correlation coefficient, *p*, 2-tailed *p*-value.

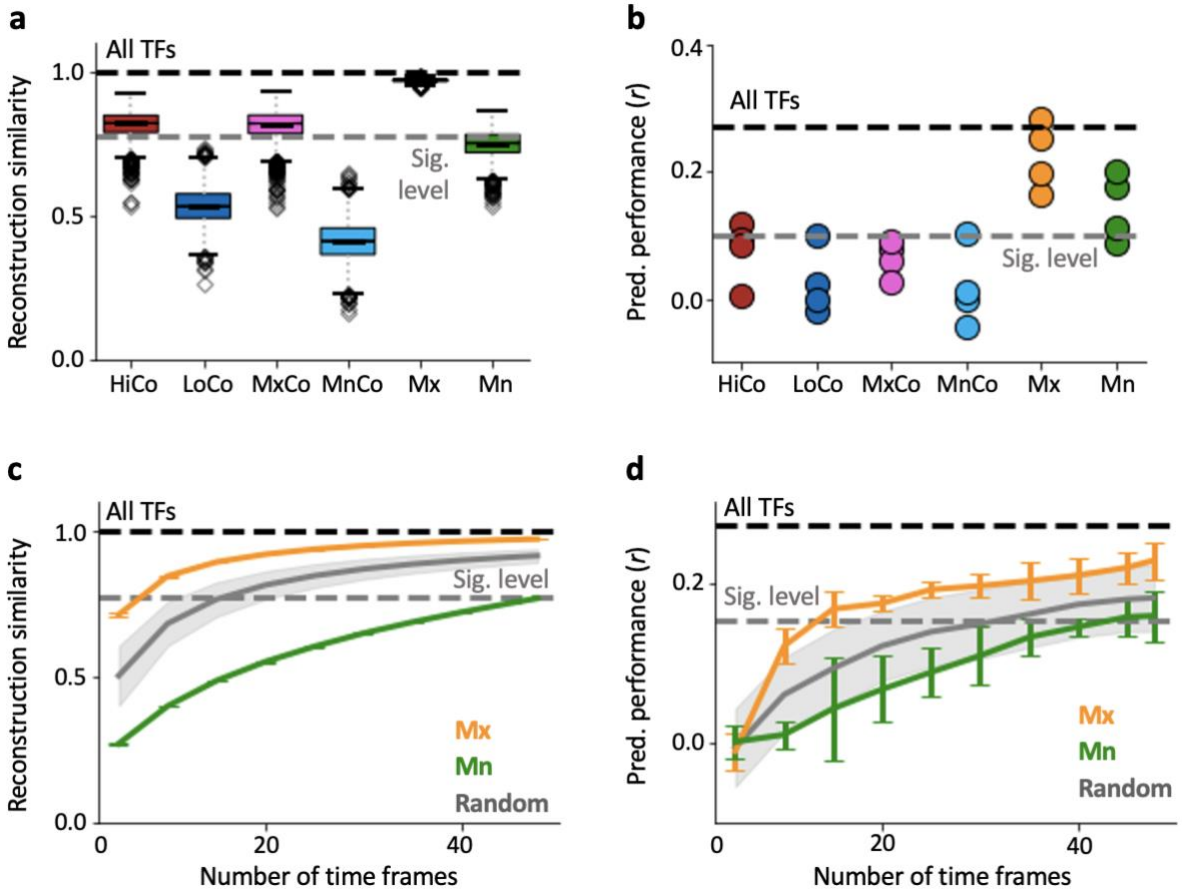

Fig. S6. The performance to predict intelligence depends on the number of temporally independent time frames rather than on reconstruction similarity also in the replication sample (HCP). (**a**) Reconstruction similarity of six different connectivity states operationalized as Pearson correlation between static functional connectivity (constructed from all time frames; TFs) and connectivity matrices reconstructed from six different selections of TFs. Boxplots depict the mean and quartiles of the subject-specific reconstruction similarity for all different connectivity states and across all four scans. The whiskers show the 1.5 x interquartile ranges. Outliers are represented by diamonds. (**b**) Prediction of intelligence (*g*-score) for the six different connectivity types from using the CMEP prediction framework (see Fig. 2). Each dot represents one scan session. (**c**) Reconstruction similarity and (**d**) performance to predict intelligence as a function of the number of (randomly selected) time frames comprising cofluctuation maxima or cofluctuation minima (orange or green dots in Fig. 1e). Gray lines represent reconstruction similarity (**c**) and predictive performance (**d**) from randomly selected time frames (see Methods for further details about the null model). Orange and green lines represent results from general maxima and general minima connectivity states averaged across scans. The whiskers represent the standard deviation across scans. Note that in (**c**) and (**d**) only the two cases (general maxima and general minima) are illustrated that allow for significant prediction of intelligence, i.e., 44 highest maxima, Mx; 44 lowest minima, Mn. The upper bounds (gray dashed lines) represent reconstruction similarity (**a, c**) or prediction performance (**b, d**) of all TFs. The lower grey dashed line reflects the 5% significance level of the within-subject similarity of static functional connectivity (**a, c**) or intelligence prediction performance (**b, d**). HiCo, high cofluctuations; LoCo, low cofluctuations; MxCo, maxima during HiCo; MnCo, minima during LoCo; Mx, Maxima; Mn, Minima (see also Fig. 1e).

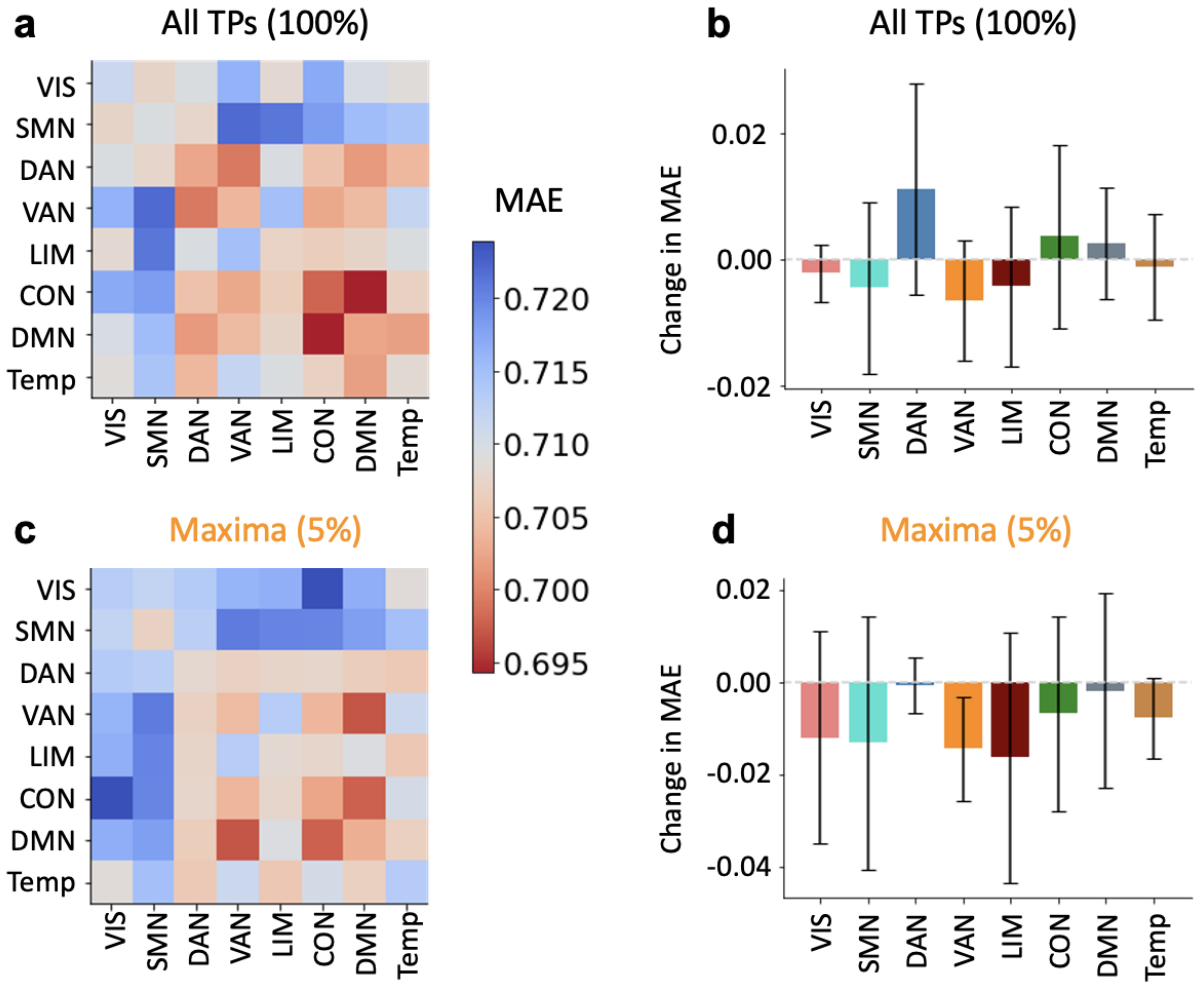

Fig. S7. Multiple functional brain systems contribute to the prediction of intelligence also in the replication sample (HCP). Intelligence (*g*-score) was predicted with CMEP from (**a, b**) static functional connectivity (all time frames; TFs) and (**c, d**) from the 44 highest maxima of the global cofluctuation (Fig. 1e). In (**a, c**) prediction performance (mean absolute error; MAE) of connectivity within or between seven functional brain networks (3) was analyzed by selecting only the specific within or between network connections, while (**b, d**) illustrates the change in prediction performance (MAE) after removing all connections a respective network was involved in. All results are depicted as the mean across the four scan sessions and whiskers indicate the standard deviation across the sessions. Significance was determined by a non-parametric permutation test with 1,000 iterations. * if $p < .05$ uncorrected for multiple comparisons and ** if $p < .05$ Bonferroni corrected for multiple comparisons (28 comparisons, $p < .0018$ in **a** and **c** and seven comparisons, $p < .007$ in **b** and **d**). VIS, visual network; SMN, somatomotor network; DAN, dorsal attention network, VAN, ventral attention network; LIM, limbic network; CON, control network; DMN, default mode network.

Tab. S1. Prediction results for 10-fold cross validation instead of leave-one-out (LOO) and when controlling intelligence scores for potential age effects.

	10-fold cross validation		**Age-adjusted intelligence scores**	
	Static functional connectivity (All TFs)	Maxima (Mx)	Static functional connectivity (All TFs)	Maxima (Mx)
r	.347**	.27**	.27*	.16
MSE	155.633**	164.87 *	157.17**	169.72
RMSE	12.475**	12.84*	12.54**	13.03
MAE	9.735**	10.02**	9.98*	10.17

Note. Results are listed for static functional connectivity (All TFs) and general maxima connectivity states (Mx). Model performance metrics reflecting the fit (*r*) or error (MSE, RMSE, MAE) between predicted and observed intelligence scores: Pearson correlation coefficient (*r*), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). Significance is determined by a non-parametric permutation test with 1,000 iterations and indicated as ** if $p < .001$, * if $p < .05$.

Tab. S2. Prediction of intelligence from functional brain connectivity for the replication sample using the Schaefer 100 nodes partition (7).

	TFs	Reconstruction similarity	*r*	MSE	RMSE	MAE
Static functional connectivity	884	n/a	.27**	0.77**	0.88**	0.70**
Highest cofluctuations (HiCo)	44	.82	.08	0.91	0.95	0.76
Lowest cofluctuations (LoCo)	44	.53	.03	0.92	0.96	0.77
Maxima during HiCo (MxCo)	7-14	.82	.07	0.91	0.95	0.76
Minima during LoCo (MnCo)	6-17	.41	.02	0.91	0.95	0.75
Highest maxima (Mx)	44	.97	.23**	0.81**	0.90**	0.72**
Lowest minima (Mn)	44	.75	.14*	0.84*	0.92*	0.74*

Note. Covariance maximizing eigenvector-based predictive modeling (CMEP; see Methods) was used in combination with a nested cross-validation scheme (see Methods, Fig. 1 and Fig. 2) to predict individual intelligence scores (latent *g*-factor derived from 12 cognitive scores) from static connectivity (all fMRI time frames; TFs), highest and lowest cofluctuation states (HiCo/LoCo; 44 TFs), maxima/minima *during* highest/lowest cofluctuation states (MxCo/MnCo; < 17 TFs), and the 44 highest maxima and lowest minima across the whole RSS time series (Mx, Mn; see Methods and Fig. 1). Reconstruction similarity values represent Pearson correlations between the static connectivity matrix (row 2) and the reconstructed connectivity matrix from the respective selection of time frames. Model performance metrics reflect the error between predicted and observed intelligence scores averaged across all four scans: Pearson correlation coefficient (*r*), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). Significance was determined by a non-parametric permutation test with 1,000 iterations for each scan and indicated as ** if $p < .001$, * if $p < .05$.

Tab. S3. Prediction of intelligence from functional brain connectivity for the replication sample using the Yeo 114 nodes partition instead of Schaefer 100 (3, 7) across all scans.

	TFs	Reconstruction similarity	*r*	MSE	RMSE	MAE
Static functional connectivity	884	n/a	.23**	0.80**	0.89**	0.71**
High cofluctuations (HiCo)	44	.81	.11	0.88	0.94	0.74*
Low cofluctuations (LoCo)	44	.53	.02	0.92	0.96	0.77
Maxima during HiCo (MxCo)	7-14	.81	.07	0.90	0.95	0.76
Minima during LoCo (MnCo)	6-17	.40	.00	0.92	0.96	0.77
Highest Maxima (Mx)	44	.97	.21**	0.81**	0.90**	0.72**
Lowest Minima (Mn)	44	.74	.16*	0.84*	0.92*	0.73*

Note. Covariance Maximizing Eigenvector-Based Predictive Modeling (CMEP; see Methods) was used in combination with a nested cross-validation scheme (see Methods, Fig. 1 and Fig 2) to predict individual intelligence scores (latent *g*-factor derived from 12 cognitive scores) from static connectivity (all fMRI time frames; TFs), highest and lowest cofluctuation states (HiCo/LoCo; 44 TFs), maxima/minima *during* highest/lowest cofluctuation states (MxCo/MnCo; < 17 TFs), and the 44 highest maxima and lowest minima across the whole RSS time series (Mx, Mn; see Methods and Fig. 1). Reconstruction similarity values represent Pearson correlations between the static connectivity matrix (row 2) and the reconstructed connectivity matrix from the respective selection of time frames. Model performance metrics reflect the error between predicted and observed intelligence scores averaged across all four scans: Pearson correlation coefficient (*r*), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). Significance was determined by a non-parametric permutation test with 1,000 iterations for each scan and indicated as ** if $p < .001$, * if $p < .05$ across all scans.

SI References

1. K. B. Nooner *et al.*, The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Front Neurosci* **6**, 152 (2012).

2. D. C. Van Essen *et al.*, The WU-Minn human connectome project: an overview. *Neuroimage* **80**, 62-79 (2013).

3. E. S. Finn, *et al.*, Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci* **18**, 1664-1671 (2015).

4. X. Shen *et al.,* Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat Protoc* **12**, 506-518 (2017).

5. B. T. Yeo *et al.,* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 1125-1165 (2011).

6. D. Wechsler, *Wechsler Abbreviated Scale of Intelligence* (Harcourt Brace and Company, 1999).

7. A. Schaefer *et al.,* Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb cortex* **28**, 3095-3114 (2018).