OXFORD

Predicting executive functioning from functional brain connectivity: network specificity and age effects

Marisa K. Heckner [iD] [1,2,*], Edna C. Cieslik[1,2], Kaustubh R. Patil[1,2], Martin Gell[1,3], Simon B. Eickhoff[1,2], Felix Hoffstädter[1,2], Robert Langner[1,2]

[1]Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Centre Jülich, 52425 Jülich, Germany,
[2]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany,
[3]Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen University, 52074 Aachen, Germany

*Corresponding author: Marisa Heckner, Institute of Neuroscience and Medicine (INM7), Research Centre Jülich, 52428 Jülich, Germany.
Email: m.heckner@fz-juelich.de

Healthy aging is associated with altered executive functioning (EF). Earlier studies found age-related differences in EF performance to be partially accounted for by changes in resting-state functional connectivity (RSFC) within brain networks associated with EF. However, it remains unclear which role RSFC in EF-associated networks plays as a marker for individual differences in EF performance. Here, we investigated to what degree individual abilities across 3 different EF tasks can be predicted from RSFC within EF-related, perceptuo-motor, whole-brain, and random networks separately in young and old adults. Specifically, we were interested if (i) young and old adults differ in predictability depending on network or EF demand level (high vs. low), (ii) an EF-related network outperforms EF-unspecific networks when predicting EF abilities, and (iii) this pattern changes with demand level. Both our uni- and multivariate analysis frameworks analyzing interactions between age × demand level × networks revealed overall low prediction accuracies and a general lack of specificity regarding neurobiological networks for predicting EF abilities. This questions the idea of finding markers for individual EF performance in RSFC patterns and calls for future research replicating the current approach in different task states, brain modalities, different, larger samples, and with more comprehensive behavioral measures.

Key words: aging; cognitive abilities; fMRI; machine learning; out-of-sample prediction.

Introduction

Executive functioning (EF) is important for many everyday behaviors, such as controlling one's attention, actions or emotions to attain higher-order or long-term goals (inhibitory control); holding content in mind and working with it as in incorporating new information in plans or considering alternatives (working memory); or changing one's perspective or mindset (cognitive flexibility; for review see Diamond 2013). EF thus denotes a multidimensional construct, which subsumes a set of different abilities that involve goal-oriented top-down control (i.e. inhibitory control, working memory, and cognitive flexibility). Therefore, its neural implementation is likely distributed across the brain, relying on a core network, the so-called multiple-demand network (intraparietal sulcus, inferior frontal sulcus, dorsolateral prefrontal cortex, anterior insula/frontal operculum, pre-supplementary motor area, and anterior cingulate cortex), and other, more specific brain regions depending on task demands (Teuber 1972; Duncan and Owen 2000; Duncan 2010; Miyake and Friedman 2012; Camilleri et al. 2018).

The human brain can be seen as a complex, nonrandom network—also called the human connectome, which can be divided into different subnetworks, subserving specific mental functions and enabling segregation and integration of the latter. One common way to identify brain networks is to study their intrinsic functional connectivity. Using this approach, several resting-state brain networks have been consistently identified, corresponding well with task-related co-activation patterns (Damoiseaux et al. 2006; Smith et al. 2009; Heckner et al. 2021). Prominent brain

networks are, for example, the default-mode network (DMN), which has been found to be most active during rest and to decrease its activity during externally oriented tasks (Raichle 2015); the executive control network (ECN), which is associated with tasks that require top-down cognitive control (Seeley et al. 2007); and the salience network, which is involved in integrating sensory with visceral, autonomic, and hedonic signals (Seeley et al. 2007). Previous research has shown that DMN deactivation decreases with age, which has further been linked to poorer performance in EF-related tasks (Persson et al. 2007; Prakash et al. 2012; Brown et al. 2019). Other research suggested that activation of the EFN and deactivation of the DMN are more strongly associated with EF performance than with chronological age (Satterthwaite et al. 2013). The default–executive coupling hypothesis of aging (DECHA) proposes that connectivity between the ECN and the DMN increases with age and is associated with poorer performance in tasks taxing cognitive flexibility (Turner and Spreng 2015; Kupis et al. 2021) and processing speed (Ng et al. 2016). Hence, healthy aging is accompanied by a decline in the performance of cognitively challenging tasks as well as altered brain activity and connectivity patterns during task and resting state (Park and Reuter-Lorenz 2008). Earlier studies found age-related differences in EF performance to be partially accounted for by changes in resting-state functional connectivity (RSFC) within brain networks associated with EF (Steffener et al. 2009; Langner et al. 2015; Hausman et al. 2020). However, it remains unclear which role RSFC within EF-related networks plays as a marker for individual EF performance. Multivariate (vs. univariate) analyses

increase the validity as they result in a pattern of connectivity (vs. individual connections) that can be linked to behavioral trait markers. This also increases the sensitivity for finding interindividual differences (Marek et al. 2022; Pat et al. 2022). Furthermore, a recent meta-analysis across 25 functional magnetic resonance imaging (fMRI) studies revealed low test–retest reliability of edge-level RSFC (Noble et al. 2019). Multivariate models, in particular multivariate prediction models have been shown to result in a substantially higher test–retest reliability compared with single imaging features, such as individual connections in RSFC (Taxali et al. 2021). Using prediction algorithms, earlier studies have been able to predict the EF abilities of previously unseen individuals from RSFC (Reineberg et al. 2015; He et al. 2021).

Taking the advantages of multivariate predictive modeling into account, the current study aimed to gain a better understanding of the neural implementation of EF, its change throughout the lifespan, and its potential as a marker for individual differences in EF performance. Therefore, we defined an EF network (EFN) based on the results of 3 large-scale neuroimaging meta-analyses capturing diverse EF facets (Rottschy et al. 2012; Langner et al. 2018; Worringer et al. 2019). In contrast to previously meta-analytically defined EF networks (Camilleri et al. 2018), we wanted to identify a widespread network which comprises all potentially relevant brain regions and is thus better suited for finding individual differences than a network based on consensus. Then, we examined to what degree individual abilities in 3 major EF subcomponents (i.e. inhibitory control, cognitive flexibility, and working memory) can be predicted from RSFC within this network in young and old adults. For each EF subcomponent, we chose a high-demand task condition, assumed to strongly probe EF abilities, as well as a low-demand control condition, assumed to hardly tax EF abilities, to test whether performance in EF-related (vs. EF-unrelated) task conditions can be better predicted from EF-related (vs. EF-unspecific) brain networks. As EF-unspecific control networks, we included a meta-analytically defined perceptuo-motor network (Heckner et al. 2021) and the whole-brain connectome (Power et al. 2011). In addition, we computed 10 random networks with 50 nodes to control for the number of features per network. For prediction, we focused on the linear regression algorithm partial least squares regression (PLSR) but additionally applied a nonlinear algorithm (random forest) as well as a data-driven feature-selection approach (Finn et al. 2015; Shen et al. 2017) for conceptual replication and testing robustness. Prediction results were submitted to a 2 (age group) × 4 (network) × 2 (task demand level) mixed-measures analysis of variance (ANOVA) to further probe network and age specificity as well as potential interactions. In particular, this study investigated (i) if young and old adults differ in the predictability of their EF abilities depending on network type or EF demand level, (ii) if an EF-related network is better at predicting EF abilities than are various EF-unspecific networks, and (iii) if this pattern changes depending on EF demand level. Including a low-demand control condition that comprised the same aspects as the high-demand condition except the EF component allowed us to investigate task specificity.

Methods
Sample
Resting-state fMRI data of 116 healthy young (age range = 20–40 years, mean age = 26.67, SD = 5.80, 64 females) and 111 old (age range = 60–80 years, mean age = 68.19, SD = 5.66, 72 females) adults were obtained from the publicly available enhanced Nathan Kline Institute—Rockland Sample (eNKI-RS; Nooner et al. 2012). These age bins were chosen to maximize the variance of age for studying age-related differences in the association between RSFC and behavioral target variables. We excluded participants with acute and/or severe psychiatric or neurological disorders in the past or when currently taking medication presumably affecting brain activity. The re-analysis of the data was approved by the Heinrich Heine University Düsseldorf ethics committee. All participants underwent the same protocol. The specific sample used is available upon request.

fMRI Data acquisition and preprocessing
Whole-brain fMRI data were obtained with a Siemens TimTrio 3T scanner using BOLD (blood oxygen level-dependent) contrast [gradient-echo EPI (echo planar imaging) pulse sequence, repetition = 1.4 s, echo time = 30 ms, flip angle = 65°, voxel size = $2.0 \times 2.0 \times 2.0$ mm^3, 64 slices]. In total, 404 volumes were acquired. Participants were instructed to keep their eyes open and maintain fixation on a central dot. Physiological and movement artifacts were removed from the EPI time-series data by using FIX (FMRIB's ICA-based Xnoiseifier, version 1.061 as implemented in FSL 5.0.9; Griffanti et al. 2014; Salimi-Khorshidi et al. 2014), which decomposes the data into independent components and identifies noise components using a large number of distinct spatial and temporal features via pattern classification. Unique variance related to the identified artefactual components was then regressed from the data. Data were further preprocessed using SPM12 (Wellcome Trust Centre Neuroimaging, London) and in-house MATLAB scripts. After removing the first 4 dummy scans of each time-series, the remaining EPI volumes were corrected for head movement by a 2-pass affine registration procedure. First, images were aligned to the initial volume and, subsequently, to the mean of all volumes. The mean EPI image was then co-registered to the gray matter probability map provided by SPM12 using normalized mutual information and keeping all EPI volumes aligned. Next, the mean EPI image of each subject was spatially normalized to MNI-152 space using the "unified segmentation" approach (Ashburner and Friston 2000). The resulting deformation parameters were then applied to all other EPI volumes.

Networks
We defined an EF-related network (EFN) by computing the maximum conjunction across 3 pertinent meta-analyses investigating working memory (Rottschy et al. 2012), cognitive action regulation (Langner et al. 2018), and multi-tasking (Worringer et al. 2019). Extracting the peak coordinates from this conjunction map resulted in a network comprised of 50 nodes. Similarly, we defined an EF-unrelated brain network linked to visual, auditory, or motor processes (Heckner et al. 2021), which comprised 59 nodes. As a whole-brain control, we employed Power et al.'s (2011) graph of putative functional areas, which includes 264 nodes. All networks are displayed in Fig. 1. Since the whole-brain approach amounted to 34,716 edges—many more than the EFN with only 1,225 — we wanted to control for the number of features. Therefore, we created 10 networks comprised of 50 nodes each, preserving the spatial properties of the EFN when pseudo-randomly sampling from a conservative whole-brain gray matter mask. In particular, we restricted the minimum, mean and maximum Euclidean distances between the sampled nodes to within one standard deviation from the minimum, mean and maximum values of the EFN, respectively (https://github.com/MartinGell/random_nets).
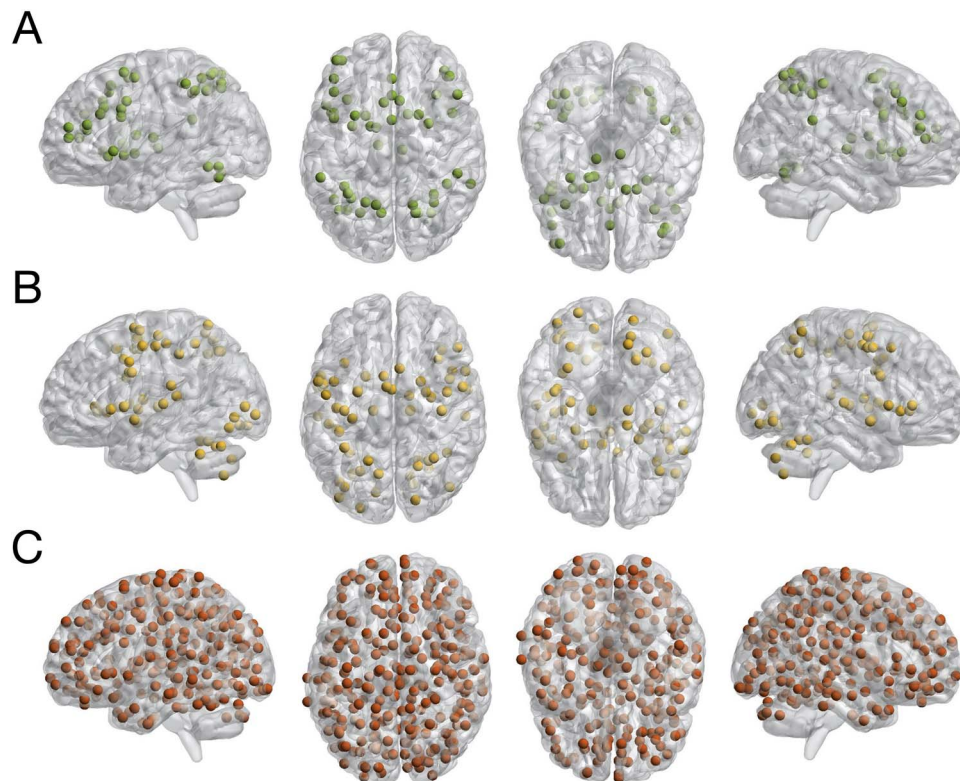
Fig. 1. Nodes of meta-analytically defined executive function A) and perceptuo-motor B) networks as well as Power et al.'s graph of putative functional areas C).

Within-network resting-state functional connectivity

For each network, RSFC was computed by first extracting the BOLD signal time-courses of all voxels within each network node expressed as the first eigenvariate for each subject. Network nodes covered a sphere with a 6-mm radius around the coordinate's peak. Peaks were extracted using the SPM Anatomy Toolbox version 3 (Eickhoff et al. 2005, 2007) in MATLAB; we made sure that the resulting spheres would not overlap or exceed the cortex. In addition, a gray matter mask including subcortical regions was used to ascertain only voxels located in gray-matter were analyzed (https://zenodo.org/record/6463123#.YlltJsjMJ3h). To reduce spurious correlations, variance explained by the mean white matter, cerebrospinal fluid, or global signal was removed from the time-series. Subsequently, data were band-pass filtered with cut-off frequencies of 0.01 and 0.1 Hz. Pairwise functional connectivity was computed as Fisher's Z-transformed linear (Pearson) correlation between the first eigenvariate of the time-series of each network's nodes. There were no significant correlations between the target variables (i.e. task scores) and sex in either subgroup. Correlations with movement and age were significant in the older subgroup. We refrained from additionally correcting RSFC values for movement (i.e. removing movement-related variance that is partially shared with age) to not unduly affect the influence of age on the association between RSFC and behavioral target variables (Miller and Chapman 2001).

To control for possible age-specific effects of global signal regression as well as movement, we additionally computed RSFC without global signal regression and with correction for 6 head movement parameters (x, y, z translations and α, β, γ rotations) derived from realignment (RPs), squared RPs as well as their derivatives. The results from RSFC without global signal regression did not change (Table S1 and S2), and the results from RSFC with movement correction did not change in regard to the main effects of network type and EF demand level as well as for all 2-way interactions and the 3-way interaction. Only the main effect of age did change (see Table S3 and S4) such that with movement correction, prediction accuracy was now better for younger than older adults. Importantly, however, this main effect was qualified by the crossed age × demand level interaction and should, therefore, be interpreted only with great caution or not at all. Secondly, as mentioned before, age and movement are likely to share variance. Therefore, partialling out movement may also remove age-related variance from the BOLD signal time-series, possibly mitigating or blurring the effects of age on the association between RSFC and the behavioral target variables.

Performance measures

EF ability scores were obtained for a highly demanding (i.e. taxing EF abilities) and less demanding (i.e. low-EF control) condition from each of 3 classical EF tasks probing working memory, inhibitory control, or cognitive flexibility, respectively. Because of the task conditions available (high- vs. low-demand), we had to dichotomize the EF demand level. Performance raw scores were z-transformed and outliers > |3| standard deviations were removed. Working memory abilities were assessed using the Short Letter-N-Back Test, which is part of Penn's computerized neurocognitive battery (CNB; Gur et al. 2010). In this test, participants are supposed to press a button as fast as possible if the letter on the screen is the same as the one shown *n* trials before. We analyzed the reaction times (RTs) of correct responses in the 1-back and 0-back conditions as high- and low-demand performance indicators, respectively. To assess inhibition abilities, we analyzed RT of the incongruent (high-demand) and congruent

(low-demand) conditions of the Color-Word Interference (CWI) Test, which is part of the Delis–Kaplan Executive Function System (D-KEFS; Delis et al. 2004). In this Stroop-like test, participants are asked to name the font color of a written word while inhibiting the automatic response to the meaning of the word, which names a different (incongruent) or the same (congruent) color. For assessing cognitive flexibility, we analyzed RT of the number and letter switching and sequencing conditions of the Trail Making Test (TMT), which is also a part of the D-KEFS. In this test, participants are required to connect consecutive targets of one type (e.g. numbers in ascending order; low-demand condition) or, in the high-demand condition, continuously switch between 2 target types (i.e. alternately connect ascending numbers with letters in alphabetical order). All tasks applied were previously evaluated and shown to have moderate to high reliability (Delis et al. 2001; Homack et al. 2005; Gur et al. 2010).

Prediction

Individual *z*-transformed performance scores were predicted from within-network RSFC using partial least squares regression, (PLSR; Krishnan et al. 2011). PLSR is similar to a supervised principal component regression (PCR; based on eigen-decomposition) and is thus advantageous when dimension reduction is beneficial for analysis. Here, we used within-network RSFC as features for prediction. With RSFC, a network comprised of 50 nodes enters 1,225 edges and thus considerably increases the feature space. In comparison with PCR, dimension reduction in PLSR is supervised (i.e. uses information about the target variables), yielding the advantage that the resulting latent variables are all related to the target variables. After dimensionality reduction, a linear regressor is applied to the transformed data.

For prediction, a 10-fold cross-validation was performed for which the data were split into 10 sets, 9 of which were used for training while the 10th was held back as a test set and subsequently used for prediction of the unseen data. This was done with each set being the test set once. In total, 100 repetitions of this 10-fold cross-validation were computed to ensure robustness.

To test whether the association between brain features and individual EF abilities may be nonlinear, we replicated our prediction analysis using a nonlinear algorithm (random forest) (Breiman 2001). Finally, we repeated our prediction analysis using a data-driven feature-selection approach (Finn et al. 2015; Shen et al. 2017) to test whether RSFC-based performance predictions differ between features obtained from meta-analytically pre-defined functional networks and features chosen from a whole-brain connectome (i.e. Power nodes) in a data-driven manner. In brief, this data-driven connectome-based predictive modeling (CPM) identifies "predictive features" by correlating each RSFC edge with the behavioral target variable from the training set. Then, the most relevant edges are chosen for prediction (i.e. edges that were significantly correlated with the target variables at $P < 0.05$). We followed the protocol by Shen et al. (2017) using positive and negative features but adjusted from leave-one-out to a 100×10-fold cross-validation scheme.

Prediction accuracy was assessed by computing the root mean squared error (RMSE), the mean absolute error (MAE), and the correlation between real and predicted values (Pearson's *r*). Accuracy measures were averaged across the 10 folds and 100 repetitions. Prediction accuracy, as represented by RMSE values for each of the 100 repetitions, was then submitted to a 2 (age group) × 4 (network) × 2 (demand level) mixed-measures ANOVA ($P < 0.00005$, Bonferroni-adjusted for the 100×10-fold

cross-validation scheme) to assess age differences in prediction accuracy, network specificity, and the impact of the EF demand level. To corroborate the main effects and to test for differences between different prediction algorithms, machine-learning-adjusted paired *t*-tests were computed (Nadeau and Bengio 2003).

Results
Behavioral analyses

We first performed *t*-tests to investigate if the age groups significantly differed in their behavioral performance (*z*-scored) in the EF-related tasks and conditions used. In all tasks and conditions, younger adults were significantly faster than older adults ($P < 0.005$). Means, standard deviations, and *t*-statistics are displayed in Table 1.

Second, we computed Pearson correlations between age and task performance for both age groups separately. Even though age had a small but significant positive correlation with RT in the CWI congruent condition for older adults ($r = 0.19$, $P = 0.048$), this association was not significant for younger adults ($r = 0.11$, $P = 0.240$). For the incongruent condition, age showed a similar significant positive correlation with RT in older adults ($r = 0.21$, $P = 0.026$) but again not in younger adults ($r = 0.066$, $P = 0.480$). For the TMT, age was significantly associated with RT in both the consecutive ($r = 0.29$, $P = 0.002$) and switch conditions ($r = 0.19$, $P = 0.048$) in older adults, but again not in younger adults ($r = 0.1$, $P = 0.270$; $r = 0.11$, $P = 0.260$). For the *n*-back task, neither association between task condition and age per group was significant ($P = 0.05$).

Prediction

CWI Congruent. The results are displayed in Table 2 and Fig. 2A.

CWI Incongruent. The results are displayed in Table 3 and Fig. 2B.

TMT Consecutive. The results are displayed in Table 4 and Fig. 2C.

TMT Switch. The results are displayed in Table 5 and Fig. 2D.

0-back. The results are displayed in Table 6 and Fig. 2E.

1-back. The results are displayed in Table 7 and Fig. 2F.

Mixed-measures ANOVA

To further assess age differences in prediction accuracy as well as network specificity and the impact of the task demand level, we submitted the prediction results represented by RMSE values for each of the 100 repetitions to a 2 (age group) × 4 (network) × 2 (demand level) mixed-measures ANOVA ($P < 0.00005$, Bonferroni-adjusted for our 10×100 cross-validation scheme).

In the first step, prediction results for all low-demand (i.e. 0-back, CWI congruent, and TMT consecutive) and all high-demand (i.e. 1-back, CWI incongruent, and TMT switch) conditions were combined into low-demand (LD) and high-demand (HD) compound scores, respectively. When Mauchly's test of sphericity was significant, Greenhouse–Geisser corrected results were interpreted. The ANOVA yielded significant main effects for the factors network [$F(2.268, 449.094) = 4{,}439.180$, $P = 6.36 \times 10^{-308}$, $\eta^2 = 0.957$], demand level [$F(1, 198) = 873.671$, $P = 1.551 \times 10^{-74}$, $\eta^2 = 0.815$], and age group [$F(1, 198) = 832.277$, $P = 7.697 \times 10^{-73}$, $\eta^2 = 0.808$] as well as the interactions age × network [$F(3, 594) = 192.666$, $P = 3.019 \times 10^{-87}$, $\eta^2 = 0.493$], age × demand level [$F(1, 198) = 24{,}857.207$, $P = 4.3025 \times 10^{-210}$, $\eta^2 = 0.992$], network × demand level [$F(2.190, 433.671) = 72.860$, $P = 4.813 \times 10^{-30}$, $\eta^2 = 0.269$], and age × network × demand level [$F(3, 594) = 156.714$, $P = 8.1004 \times 10^{-75}$, $\eta^2 = 0.442$]. To corroborate the main effects,

Table 1. Mean, standard deviation, and *t*-statistics of the performance of old and young subjects in the behavioral tasks.

	M_{old}	SD_{old}	M_{young}	SD_{young}	*t*-statistic	*P*-value
0-back	−0.02	0.62	−0.33	0.89	3.06	0.003
1-back	−0.00	0.68	−0.30	0.88	2.90	0.004
CWI_con	0.04	0.91	−0.30	0.90	2.82	0.005
CWI_inc	0.20	0.84	−0.47	0.62	6.82	< 0.001
TMT_con	0.12	0.67	−0.37	0.54	6.00	< 0.001
TMT_switch	0.07	0.77	−0.33	0.64	4.20	< 0.001

Note. CWI_con = congruent condition of color-word interference test, CWI_inc = incongruent condition of color-word interference test, TMT_con = consecutive condition of trail making test, TMT_switch = switching condition of trail making test.

Table 2. Prediction accuracies displayed as Pearson's *r*, mean absolute error, and root mean squared error for both age groups from prediction within the executive function-related, perceptuo-motor, whole-brain, and averaged random brain networks.

	r_{old}	r_{young}	MAE_{old}	MAE_{young}	$RMSE_{old}$	$RMSE_{young}$
EFN	0.1	−0.04	0.74	0.85	0.95	1.04
PercMot	0.01	0.05	0.75	0.76	0.96	0.98
Power	0.13	0.12	0.72	0.74	0.92	0.91
RandomAvg	0.06	0.00	0.72	0.79	0.94	0.98

Note. MAE = mean absolute error, RMSE = root mean squared error, EFN = executive function network, PercMot = perceptuo-motor network, RandomAvg = averaged 10 random networks.
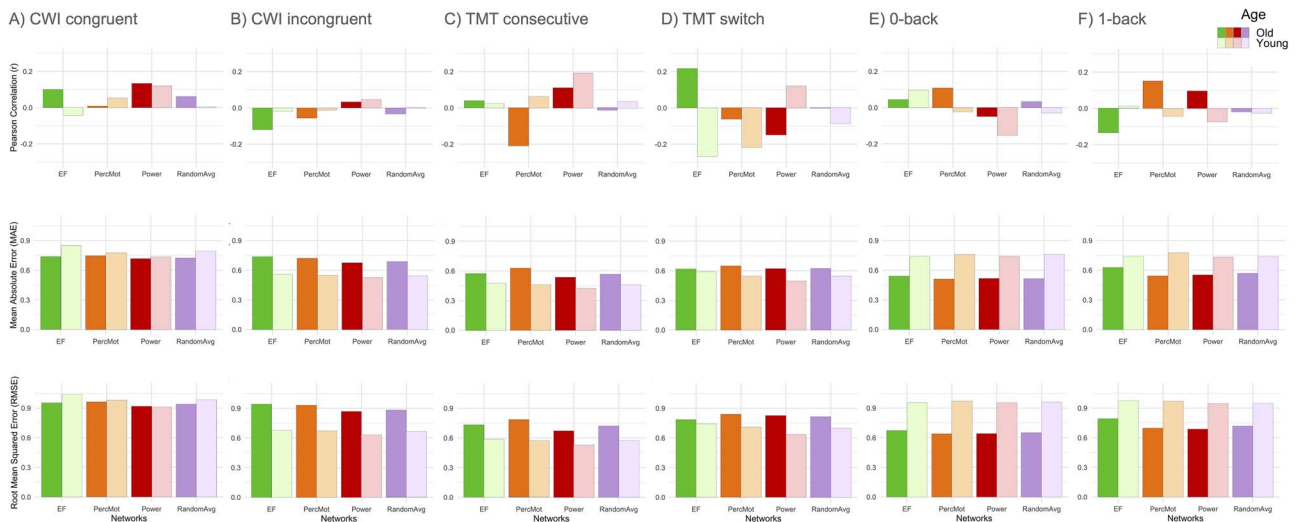
Fig. 2. Prediction accuracies expressed as mean Pearson correlations between observed and predicted scores, mean absolute error, as well as root mean squared error for the task conditions color-word interference (CWI) congruent A), CWI incongruent B), trail making test (TMT) consecutive C), TMT switch D), 0-back E), and 1-back F) based on executive function (EF), perceptuo-motor (PercMot), whole-brain (Power), and averaged random networks (RandomAvg), shown separately for the old (dark colors), and young (light colors) subsamples.

Table 3. Mean prediction accuracies displayed as Pearson's *r*, mean absolute error, and root mean squared error for each age group based on the functional connectivity within the executive function-related, perceptuo-motor, whole-brain, and averaged random brain networks.

	r_{old}	r_{young}	MAE_{old}	MAE_{young}	$RMSE_{old}$	$RMSE_{young}$
EFN	−0.12	−0.02	0.74	0.56	0.94	0.68
PercMot	−0.06	−0.01	0.72	0.55	0.93	0.67
Power	0.03	0.05	0.67	0.53	0.87	0.63
RandomAvg	−0.03	0.00	0.68	0.54	0.88	0.66

Note. MAE = mean absolute error, RMSE = root mean squared error, EFN = executive function network, PercMot = perceptuo-motor network, RandomAvg = averaged 10 random networks.

we conducted machine-learning-adjusted *t*-tests. The *t*-tests confirmed the main effects for age and all network differences. Only the main effect for demand level was not confirmed (see Table S5).

Overall, prediction accuracy (i.e. RMSE) was better for older than for younger subjects (see Fig. 3A and Table 8). Across demand level and age groups, prediction accuracy was best for the whole-brain connectome (i.e. Power nodes) compared with the averaged random, perceptuo-motor, and EF network (see Fig. 3B and Table 8). A figure showing the RMSE values of the 10 random networks separately can be found in the Supplementary Material (see Fig. S1).

Table 4. Mean prediction accuracies displayed as Pearson's *r*, mean absolute error, and root mean squared error for each age group based on the functional connectivity within the executive function-related, perceptuo-motor, whole-brain, and averaged random brain networks.

	r_{old}	r_{young}	MAE_{old}	MAE_{young}	$RMSE_{old}$	$RMSE_{young}$
EFN	0.04	0.02	0.57	0.47	0.73	0.59
PercMot	−0.21	0.06	0.63	0.46	0.79	0.57
Power	0.11	0.19	0.53	0.42	0.67	0.53
RandomAvg	−0.01	0.04	0.56	0.48	0.72	0.57

Note. MAE = mean absolute error, RMSE = root mean squared error, EFN = executive function network, PercMot = perceptuo-motor network, RandomAvg = averaged 10 random networks.

Table 5. Mean prediction accuracies displayed as Pearson's *r*, mean absolute error, and root mean squared error for each age group based on the functional connectivity within the executive function-related, perceptuo-motor, whole-brain, and averaged random brain networks.

	r_{old}	r_{young}	MAE_{old}	MAE_{young}	$RMSE_{old}$	$RMSE_{young}$
EFN	0.22	−0.27	0.62	0.59	0.78	0.74
PercMot	−0.06	−0.22	0.65	0.54	0.84	0.71
Power	−0.15	0.12	0.62	0.50	0.82	0.64
RandomAvg	0.00	−0.09	0.62	0.54	0.81	0.70

Note. MAE = mean absolute error, RMSE = root mean squared error, EFN = executive function network, PercMot = perceptuo-motor network, RandomAvg = averaged 10 random networks.

Table 6. Mean prediction accuracies displayed as Pearson's *r*, mean absolute error, and root mean squared error for each age group based on the functional connectivity within the executive function-related, perceptuo-motor, whole-brain, and averaged random brain networks.

	r_{old}	r_{young}	MAE_{old}	MAE_{young}	$RMSE_{old}$	$RMSE_{young}$
EFN	0.04	0.10	0.54	0.74	0.67	0.96
PercMot	0.11	−0.02	0.51	0.76	0.64	0.97
Power	−0.05	−0.15	0.51	0.74	0.64	0.95
RandomAvg	0.03	−0.03	0.51	0.76	0.65	0.96

Note. MAE = mean absolute error, RMSE = root mean squared error, EFN = executive function network, PercMot = perceptuo-motor network, RandomAvg = averaged 10 random networks.

Table 7. Mean prediction accuracies displayed as Pearson's *r*, mean absolute error, and root mean squared error for each age group based on the functional connectivity within the executive function-related, perceptuo-motor, whole-brain, and averaged random brain networks.

	r_{old}	r_{young}	MAE_{old}	MAE_{young}	$RMSE_{old}$	$RMSE_{young}$
EFN	−0.13	0.01	0.62	0.74	0.79	0.97
PercMot	0.15	−0.04	0.54	0.77	0.69	0.97
Power	0.10	−0.07	0.55	0.73	0.68	0.94
RandomAvg	−0.02	−0.03	0.57	0.74	0.72	0.95

Note. MAE = mean absolute error, RMSE = root mean squared error, EFN = executive function network, PercMot = perceptuo-motor network, RandomAvg = averaged 10 random networks.

Prediction accuracy was better for HD than LD (see Fig. 3C and Table 8) conditions across networks and age groups.

Prediction accuracy for both age groups was best for the Power nodes, as compared with the averaged random, perceptuo-motor, and EF network (see Fig. 4A and Table S6). Although prediction accuracy for older participants was better for LD than HD conditions, prediction accuracy for younger participants was better for HD (see Fig. 4B and Table S6) conditions. Finally, all networks performed better for HD than LD (see Fig. 4C and Table S6) task conditions.

T-tests between algorithms

For each age group, prediction accuracy, as indicated by RMSE, was significantly better for the nonlinear algorithm (RF) in all networks than for the linear algorithm PLSR (P < 0.001). The results of the adjusted *t*-test can be found in Table 9. Comparing PLS with CPM, prediction accuracy only differed significantly (P = 0.05) in younger adults, with PLSR slightly outperforming CPM (see

Table 10). Tables stating the prediction accuracies for RF and CPM can be found in the Supplementary Material (see Table S7 and S8).

Discussion

This study aimed to extend our understanding of the implementation of EF in the human brain, its age-related differences, and its potential as a marker for individual differences in EF performance. After deriving EF-related brain regions from previous neuroimaging meta-analyses, we examined to what degree individual EF abilities, as assessed via 3 classic tasks with high- and low-demand conditions, can be predicted from RSFC within this EFN in young and older adults. For comparison, we included 2 EF-unspecific networks (i.e. a perceptuo-motor network and a representation of the whole-brain connectome) as well as 10 pseudo-random networks with the same properties as the EFN. Eventually, we replicated our analysis, which used a linear prediction algorithm (i.e. PLSR), employing a nonlinear prediction

Table 8. Post-hoc pairwise comparisons of ANOVA effects.

Post-hoc pairwise comparisons

Factor	Mean (SE)$_{group1}$	Mean (SE)$_{group2}$	P-value
Age	0.792 (0.0002)$_{old}$	0.802 (0.0002)$_{young}$	7.70×10^{-73}
Network	0.820 (0.0005)$_{EF}$	0.809 (0.0004)$_{PercMot}$	1.61×10^{-40}
	0.820 (0.0005)$_{EF}$	0.767 (0.0003)$_{Power}$	6.17×10^{-172}
	0.820 (0.0005)$_{EF}$	0.786 (0.0004)$_{Random}$	2.81×10^{-124}
	0.809 (0.0004)$_{PercMot}$	767 (0.0003)$_{Power}$	1.42×10^{-160}
	0.809 (0.0004)$_{PercMot}$	793 (0.0001)$_{Random}$	1.39×10^{-89}
	767 (0.0003)$_{Power}$	793 (0.0001)$_{Random}$	1.92×10^{-161}
Demand Level	0.793 (0.0003)$_{HD}$	0.802 (0.0002)$_{LD}$	1.55×10^{-74}

Note. EF = executive-function-related network, PercMot = perceptuo-motor-related network, Power = Power et al.'s (2011) graph of putative functional areas, Random = average of 10 randomly computed brain networks, HD = high-demand, LD = low-demand, SE = standard error.
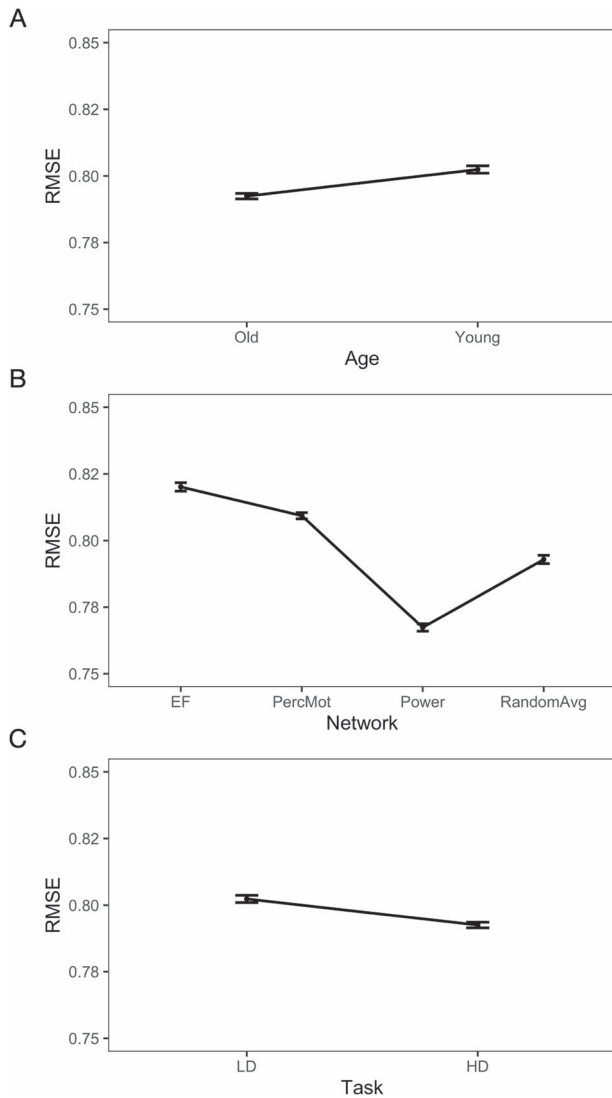

Fig. 3. Main effect for age, network, and task (mean ± standard error). Prediction accuracy (i.e. root mean squared error, RMSE) was better for older compared with younger subjects A); for high-demand (HD) compared with low-demand (LD) task conditions B); and best for the whole-brain connectome (i.e. Power nodes), compared with the executive function (EF), perceptuo-motor (PercMot), and averaged random networks (P < 0.00005; C).
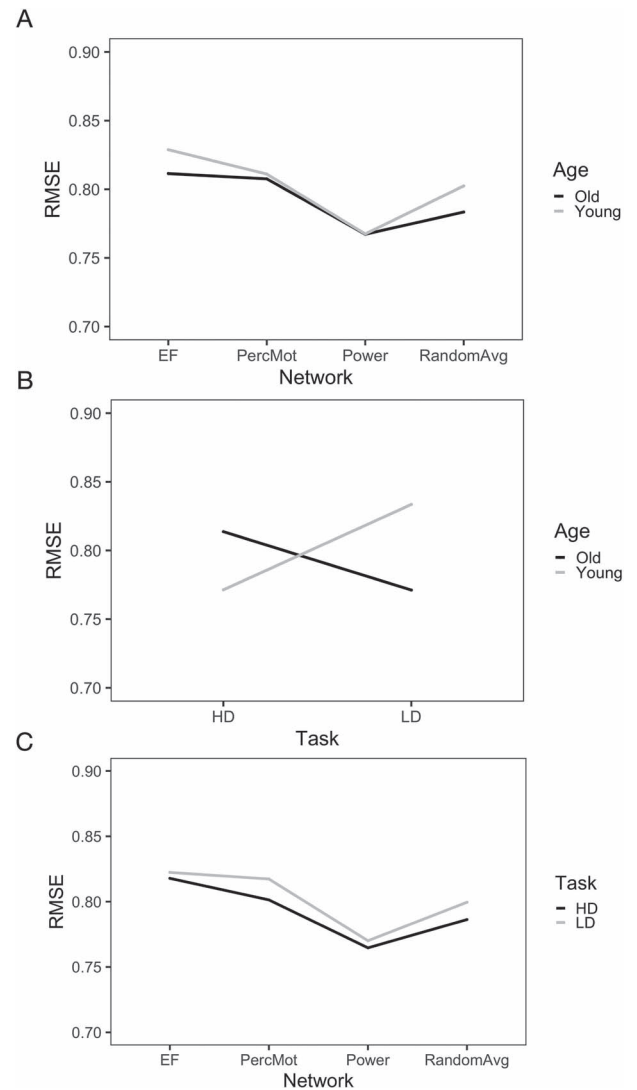
Fig. 4. Interaction effects for age ×network, age ×task, and network ×task. Prediction accuracy (i.e. root mean squared error, RMSE) was best for the whole-brain approach (i.e. Power nodes) for both age groups (P < 0.00005; A). Prediction accuracy for younger subjects was best for high-demand (HD) while prediction accuracy for older subjects was best for low-demand (LD) task conditions (P < 0.00005; B). Prediction accuracy was better for the high-demand task condition in all brain networks investigated (P < 0.00005; C).

algorithm (RF) as well as a data-driven feature-selection approach without any a priori network definition. As expected, younger adults were significantly faster than older adults in all tasks (i.e. n-back, CWI, and TMT) and both demand levels (i.e. HD vs. LD). Prediction accuracies, reflecting the strength of brain–behavior associations, were overall rather low. Surprisingly, at the given average low level of accuracy, EF-unspecific networks even outperformed the EF-specific network. Finally, prediction accuracy was better for HD than LD task conditions.

Prediction of EF abilities from network-level RSFC

Pearson's correlation coefficient (r), calculated to quantify the association between observed and predicted scores, is being widely used in neuroscience as an accuracy measure for predictive analyses or as a reference for a prediction's success. However, a correlation coefficient provides information about the strength of a linear relationship and not the accuracy of a model (Li 2017). RMSE is the standard deviation of the residuals. It provides information about how far away, on average, the data points are from the regression line and thus how accurate the prediction was. MAE measures the average magnitude of errors. RMSE and MAE do not depend on the variance of the outcome variable and thus allow comparisons across data sets. Therefore, RMSE, MAE, and correlation coefficient do not need to go hand in hand and offer different, yet complementary, information. A recent paper by Poldrack et al. (2020) discussing best practices evaluating prediction pointed out that high correlation coefficients can occur even if predicted data substantially differ from observed data (e.g. due to outliers or very heterogeneous groups). The authors recommend reporting and interpreting several measures of prediction accuracy, especially RMSE and MAE.

Here, we will discuss prediction accuracy as measured with RMSE (<0.8) and the respective correlation coefficient (note that cognitive performance was z-scored). This follows the rationale that RMSE provides information about the accuracy of a prediction and thus only the correlation coefficient of successful (i.e. meaningful) predictions will be further interpreted. As mentioned before, prediction accuracy for all networks and both age groups was overall rather low. Even for TMT switching, which featured the highest r, the predicted score based on within-EFN RSFC was not able to explain any variance in interindividual performance differences in older adults based on the coefficient of determination ($R^2 = -.02$; Scheinost et al. 2019). For within-perceptuo-motor network RSFC, similarly no variance was explained. The whole-brain approach, which achieved the overall best prediction accuracy (i.e. RMSE), was only able to explain 3.7% of the variance in the TMT consecutive condition in younger adults. Finally, RSFC within random networks, while resulting in second-best prediction accuracy, was not able to explain any variance in the target variables. Although previous research has made positive claims regarding the predictability of cognitive abilities from network-level RSFC, the explained variance rarely exceeded $r > 0.3$ (i.e. $r^2 = 0.09$; Ferguson et al. 2017; Greene et al. 2018; He et al. 2021), which is comparable with the strength of brain–behavior associations observed here [TMT_switch$_{old}$ from EFN ($r = 0.22$; $r^2 = 0.05$), 0back$_{old}$ ($r = 0.11$; $r^2 = 0.01$) and 1back$_{old}$ ($r = 0.15$; $r^2 = 0.02$) from the perceptuo-motor network as well as 1back$_{old}$ ($r = 0.10$; $r^2 = 0.01$), TMT_con$_{old}$ ($r = 0.11$; $r^2 = 0.01$), TMT_con$_{young}$ ($r = 0.19$; $r^2 = 0.04$) and TMT_switch$_{young}$ ($r = 0.12$, $r^2 = 0.01$) from the whole-brain connectome]. Note, that for the sake of comparability with published results, we here used the squared correlation coefficient as coefficient of determination.

In conclusion, when predicting EF ability from RSFC, we achieved overall rather low prediction accuracies, which suggests that the relationship between individual RSFC patterns and EF abilities is rather loose. Although the level of predictability is comparable with previous related research, it is rather unsatisfactory from a translational perspective aiming to discover biomarkers for individual-level prediction. Interestingly, our results also point out that RMSE, MAE, and correlation coefficient may offer different, yet complementary, information. This further brings into question whether prediction results evaluated only based on the correlation between predicted and observed scores should be interpreted at all. The question remains what kind of variance has been explained with the approach used and how meaningful it is, especially when the amount of variance explained is low to moderate.

Network specificity

The ANOVA yielded a main effect of network on prediction accuracy, and post-hoc pairwise comparisons revealed significant differences between all networks. It turned out that the best prediction accuracy was obtained with the whole-brain connectome, followed, in descending order, by the random networks, the perceptuo-motor network, and lastly—against expectations—the EFN. One possible explanation for the superiority of Power et al.'s (2011) version of the connectome in predicting individual EF abilities might be its greater feature space, which amounts to 34,716 connections, compared with 1,225 connections for the EFN. However, in our control analysis, 10 random networks of the same size as the EFN still resulted in a significantly better prediction accuracy than the EFN and the perceptuo-motor network, pointing out that it is not the sheer number of features that is responsible for the prediction outcome. As discussed before in the context of overall weak brain–behavior associations, the question of what part of the variance we are explaining becomes even more pressing if a whole-brain approach or even random networks outperform the network that is specifically involved in implementing EF in the human brain.

An explanation for the low prediction accuracy, reflecting rather loose RSFC–behavior associations, could be RSFC's unconstrained nature. Several recent studies have shown that prediction from brain activity during tasks (or movie watching) works better than from rest (Greene et al. 2018; Sripada et al. 2020; Finn and Bandettini 2021). Tasks modulate the functional brain state, which may yield important information about individual differences in brain functional organization and cognition (Greene et al. 2018). During tasks, FC changes are likely to subserve task-related processing and are thus more constrained. In fact, it has been shown that even only thoughts about a task already influence the FC pattern (Gregory et al. 2016). In contrast, the functional brain state during rest is much less externally constrained and therefore highly dependent on the participants themselves (Buckner et al. 2013; Tailby et al. 2015). Therefore, resting-state fMRI does not offer any certainty about what mental state is being captured, let alone experimental control over what is going on mentally (Finn and Bandettini 2021). Measurement conditions that involve movie watching, on the other hand, have been linked to the selection of specific pathways, such as higher-level brain regions integrating sensory information and an increase in inter-hemispheric exchange through a global reorganization of functional communities (Gilson et al. 2018).

Table 9. Results of adjusted two-sided *t*-tests comparing the prediction accuracies (RMSE) of the linear machine learning algorithm partial least squares and the nonlinear algorithm random forest.

PLSR vs. RF

	Mean (SD)$_{\text{old PLSR}}$	Mean (SD)$_{\text{old RF}}$	t-statistic$_{\text{old}}$ (Cohen's d)	Mean (SD)$_{\text{young PLSR}}$	Mean (SD)$_{\text{young RF}}$	t-statistic$_{\text{young}}$ (Cohen's d)
EF	0.81 (0.11)	0.76 (0.12)	4.19** (0.45)	0.83 (0.18)	0.76 (0.17)	6.35**(0.41)
PercMot	0.81 (0.13)	0.76 (0.12)	4.89** (0.42)	0.81 (0.18)	0.76 (0.17)	8.48**(0.29)
Power	0.77 (0.12)	0.75 (0.11)	2.50*(0.18)	0.77 (0.19)	0.75 (0.17)	2.22* (0.12)

Note. EF = executive-function-related network, PercMot = perceptuo-motor-related network, Power = Power et al.'s (2011) graph of putative functional areas, PLSR = partial least squares, RF = random forest, SD = standard deviation. **Significant at $P < 0.001$. *Significant at $P < 0.05$.

Table 10. Results of adjusted 2-sided *t*-tests comparing the prediction accuracies (RMSE) of 2 different feature selection algorithms: The linear machine learning algorithm partial least squares and the data-driven connectome-based predictive modeling.

PLSR vs. CPM

	Mean (SD)$_{\text{old PLSR}}$	Mean (SD)$_{\text{old CPM}}$	t-statistic$_{\text{old}}$ (Cohen's d)	Mean (SD)$_{\text{young PLSR}}$	Mean (SD)$_{\text{young CPM}}$	t-statistic$_{\text{young}}$ (Cohen's d)
Power	0.77 (0.12)	0.77 (0.11)	−1.10 (0.00)	0.77 (0.19)	0.78 (0.18)	−1.97* (0.06)

Note. Power = Power et al'.s graph of putative functional areas, PLSR = partial least squares, CPM = connectome-based predictive modeling, SD = standard deviation. *Significant at $P = 0.05$.

Along these lines, a recent paper titled "Is it time to put rest to rest?" brought into question if cognitive neuroscience using resting-state fMRI data has reached a plateau and suggested moving on to integrated designs that draw from the advantages of both rest (i.e. self-generated activity) and task (i.e. control and interpretability) (Finn 2021). Examples for integrated designs are task-signature echoes, where task paradigms are used to learn signatures of brain activity that correspond to particular task conditions, followed by searching for these signatures (an echo) in resting-state data; annotated rest, where introspection data are acquired about the subjective mental experience during or after the scan; state-informed approaches, where the brain state is monitored, for example through real-time neuroimaging, and tasks are given at certain intervals to causally test the influence of ongoing activity; or naturalistic designs such as movie watching. It has been shown that between-network communication increases during movie watching vs. within-network communication (i.e. network integration), whereas within-network communication increases during rest vs. between-network communication (i.e. network segregation; Betzel et al. 2020). Higher network integration has been associated with the processing of local, specialized information whereas higher network segregation has been linked to the transfer of inter-modular information (Shine et al. 2016; Fukushima et al. 2018). In the context of EF, higher network integration might explain why the whole-brain approach or even random networks outperformed the specific EFN in the current study, as they possible comprise important hubs of further relevant brain networks.

Therefore, another potential reason for low prediction accuracy from pre-defined brain networks might be that interindividual differences in complex mental functions such as EF are not so much related to single networks but rather global organizational properties of the brain (Pläschke et al. 2020). Brain regions crucial but not specific to EF—e.g. modulating between-network communication—might be missing from the meta-analytically derived networks but are covered by the whole-brain approach. It has been shown, for example, that the dynamic reconfiguration of frontoparietal and frontotemporal networks (i.e. network

flexibility) was able to predict individual performance in a working memory task (Braun et al. 2015), pointing out that between-network communication might be as important (if not more) as within-network communication. Here, we additionally looked at the features (i.e. RSFC edges) resulting from the data-driven feature-selection approach (i.e. features that were significantly associated with the EF target variables) and observed that the pattern of predictive edges was distributed throughout the entire brain and across intrinsic networks (see Figs. S2–S5, Tables S9 and S10). This finding supports the notion that between-network connectivity, for example, between the salience network and DMN, might contain more information about individual performance in EF-related tasks than does FC within networks.

We replicated our analyses with a data-driven feature-selection approach CPM (Finn et al. 2015; Shen et al. 2017) as well as with a nonlinear prediction algorithm (random forest) to investigate whether the association between within-network RSFC and our cognitive target variables might be better captured by selected, relevant features or if the relationship is nonlinear. For older adults, the data-driven feature-selection approach—where linear regression was applied for prediction—did not differ from feature selection based on eigen-decomposition using PLSR. For younger adults, PLSR (vs. CPM) resulted in a slightly but significantly higher prediction accuracy. Random forest, in turn, statistically significant but numerically small exceeded PLSR in both age groups. Therefore, overall, we were able to replicate the prediction results and patterns obtained with our initial algorithm (PLSR). Importantly, these results speak against the idea that between-network connectivity contained more information than within-network connectivity, as prediction accuracy of the data-driven feature-selection approach across the whole-brain did not outperform prediction limited to the pre-defined networks.

Another reason for interindividual differences in EF performance not being visible when predicting from the meta-analytically pre-defined EFN may be that meta-analyses result in regions showing the most consistent activation across subjects (i.e. possibly not capturing interindividual differences). This might similarly affect RSFC from within these networks. However, RSFC

from the whole-brain approach as well as from within random networks does not result in high prediction accuracy either, suggesting that RSFC patterns overall do not seem particularly associated with individual levels of performance. Returning to the question, what variance is explained by within-network RSFC. One possibility could be that within-network RSFC is suited for predicting individual performance in simple, but not complex, EF tasks. In complex EF tasks, the variance explained might be associated with perceptual and motor aspects rather than EF components of the task. These perceptual and motor aspects are more likely associated with global fitness of the brain, which is linked to many different parts of the brain rather than specific EF patterns. In addition, the EFN could consist of regions that are always involved during complex tasks but their individual level of involvement (i.e. individual level of interconnectivity) might not be important as long as they are recruited. In this scenario, there would be no association between the level of within-EFN RSFC and interindividual differences in EF ability, making a prediction impossible. Therefore, these regions might be important for EF but not for interindividual differences in EF. Lastly, another reason for individual differences not being visible when predicting from pre-defined brain networks derived from group-average maps might be the high interindividual variability in the brain's anatomy but also in the functional organization of the brain (Wang and Liu 2014). Therefore, a promising alternative for future studies would be prediction from individualized brain networks that acknowledge differences in size, location, and connectivity of brain regions. A recent study, for example, has shown that RSFC among individualized regions of interest was better at predicting fluid intelligence than was connectivity derived from group-level brain atlases (Li et al. 2019).

In summary, our results indicate that a comprehensively defined EFN is not better at predicting EF abilities than a perceptuo-motor network, random networks or the full connectome. In fact, the latter led to better predictions overall—irrespective of task demand level or age. This indicates that the whole-brain connectome and even random networks—which are structurally similar to the EFN and thus offer a direct comparison—contain more information about individual EF abilities than a comprehensive EF-related network. Prediction from task or integrated designs such as movie watching as well as from individualized brain networks might offer better prediction accuracy.

Age specificity

Overall prediction accuracy was better for older (vs. younger) adults, which is in line with previous work showing that brain–behavior associations become closer with advancing age (Pläschke et al. 2020). On the other hand, in all tasks except n-back, older adults showed greater variability (i.e. how scattered the behavioral scores were; see Table S11). However, this main effect was qualified by a crossed age × demand level interaction and will therefore not be discussed.

Interestingly, this interaction revealed that the highest prediction accuracy for younger adults was obtained when predicting HD task conditions, and the lowest prediction accuracy was obtained when predicting LD conditions. For older adults, the pattern was reversed. A possible explanation could be an age-related decline in within-network specificity or segregation between networks (Chan et al. 2017; Varangis et al. 2019). Although performance in LD tasks (i.e. less specific to EF) might still be maintained in advanced age by some form of compensation such as the additional recruitment of domain-general resources (Carp et al. 2010) subserved by the networks investigated here, performance

in HD tasks might not, resulting in low predictability. One reason might be ineffective compensatory attempts, as the EFN is already being recruited at LD so that the RSFC adjustments within the EFN cannot be linked to performance. Another possibility would be that the compensatory adjustment happens outside of the EFN. However, the latter explanation is not supported by the current results, as brain–behavior associations did not meaningfully improve with using the Power nodes.

In summary, LD tasks appear to be sensitive markers for EF abilities in older adults who show typical age-related cognitive decline and thus already need to compensate in LD. For younger adults, as expected, only HD tasks are sensitive enough for capturing meaningful interindividual differences in EF performance. This further points out the importance of adaptive testing, as only sensitive tests can result in interindividual differences that are meaningful and can be linked to trait-markers in the brain.

Conclusions and outlook

The current study investigated to what extent the individual EF performance of older and younger adults can be predicted from RSFC patterns in pre-defined functional brain networks. Overall, prediction accuracy was rather low to moderate, partially in line with previous studies. These results bear on 2 important issues. First, to meaningfully interpret such brain-based prediction results, it is important to not only report correlations between observed and predicted scores but also include additional measures of prediction accuracy, such as RMSE or MAE. Second, our results raise the question of whether an explanation of < 4% variance is meaningful and how results from RSFC may help to gain a better understanding of brain–behavior associations. Therefore, our results question the idea of finding biomarkers for individual EF performance in functional coupling patterns at rest. Furthermore, our results show that EF-unspecific networks outperformed a circumscribed EF-related network. This may indicate that interindividual differences in EF are not so much related to single networks but rather global properties of the brain (e.g. overall atrophy in older subjects or level of global fitness). In addition, central hubs modulating between-network communication might be missing from meta-analytically derived networks but are likely covered by the whole-brain connectome or by some random networks. Future studies may want to replicate the current approach with prediction from individualized brain networks that retain interindividual differences in size, location, and connectivity of brain regions. Furthermore, LD tasks appear to be sensitive markers for EF abilities in older adults who show typical age-related cognitive decrease and thus already need to compensate in LD. These findings suggest adaptive testing as only sensitive tests can capture meaningful brain–behavior relationships.

A limitation of the current study is the unconstrained nature of RS-fMRI and our limited understanding of it in the context of prediction analyses. It appears important to conceptually replicate our analyses with different brain modalities. A recent study, for example, has shown that prediction of crystallized intelligence from regional homogeneity (ReHo), which evaluates the time consistency of the BOLD signal in a particular brain region, resulted in a higher prediction accuracy than RSFC-based prediction (Larabi et al. 2021). Furthermore, replication with task-based FC, FC during movie watching, or other designs integrating the advantages of task and rest, might improve prediction accuracy and the association with the target variables. Another limitation was the task conditions available (high- vs. low-demand), which did not allow

for analyzing EF demand level as a continuous variable. Therefore, future research would benefit from EF tasks that allow for a continuous assessment of EF demand level. Similarly, it would be an advantage to investigate a continuous age distribution within a longitudinal design to assess the trajectory of changes throughout the lifespan.

Overall, our results reveal complex interactions among several factors and, importantly, an overall lack of specificity of neurobiologically plausible networks for predicting interindividual differences in EF abilities, questioning the idea of finding biomarkers for individual levels of EF performance in predefined networks applying RSFC. Our results call for future research replicating the current study in different task states, in different and larger samples as well as with different modalities.

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

References

Ashburner J, Friston KJ. Voxel-based morphometry—the methods. *NeuroImage*. 2000:11(6):805–821. https://doi.org/10.1006/nimg.2000.0582.

Betzel RF, Byrge L, Esfahlani FZ, Kennedy DP. Temporal fluctuations in the brain's modular architecture during movie-watching. *NeuroImage*. 2020:213:116687. https://doi.org/10.1016/j.neuroimage.2020.116687.

Braun U, Schäfer A, Walter H, Erk S, Romanczuk-Seiferth N, Haddad L, Schweiger JI, Grimm O, Heinz A, Tost H, et al. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc Natl Acad Sci U S A*. 2015:112(37):11678. https://doi.org/10.1073/pnas.1422487112.

Breiman L. Random forests. *Mach Learn*. 2001:45(1):5–32. https://doi.org/10.1023/A:1010933404324.

Brown CA, Schmitt FA, Smith CD, Gold BT. Distinct patterns of default mode and executive control network circuitry contribute to present and future executive function in older adults. *NeuroImage*. 2019:195:320–332. https://doi.org/10.1016/j.neuroimage.2019.03.073.

Buckner RL, Krienen FM, Yeo BTT. Opportunities and limitations of intrinsic functional connectivity MRI. *Nat Neurosci*. 2013:16(7):832–837. https://doi.org/10.1038/nn.3423.

Camilleri JA, Müller VI, Fox P, Laird AR, Hoffstaedter F, Kalenscher T, Eickhoff SB. Definition and characterization of an extended multiple-demand network. *NeuroImage*. 2018:165:138–147. https://doi.org/10.1016/j.neuroimage.2017.10.020.

Carp J, Gmeindl L, Reuter-Lorenz P. Age differences in the neural representation of working memory revealed by multi-voxel pattern analysis. *Front Hum Neurosci*. 2010:4:217. https://doi.org/10.3389/fnhum.2010.00217.

Chan MY, Alhazmi FH, Park DC, Savalia NK, Wig GS. Resting-state network topology differentiates task signals across the adult life span. *J Neurosci*. 2017:37(10):2734. https://doi.org/10.1523/JNEUROSCI.2406-16.2017.

Damoiseaux JS, Rombouts SARB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF. Consistent resting-state networks across healthy subjects. *Proc Natl Acad Sci U S A*. 2006:103(37):13848. https://doi.org/10.1073/pnas.0601417103.

Delis DC, Kaplan E, Kramer JH. Delis-Kaplan executive function system (D-KEFS). *APA PsycTests* 2001. https://doi.org/10.1037/t15082-000.

Delis DC, Kramer JH, Kaplan E, Holdnack J. Reliability and validity of the Delis-Kaplan executive function system: an update. *J Int Neuropsychol Soc*. 2004:10(2):301–303. https://doi.org/10.1017/S1355617704102191.

Diamond A. Executive functions. *Annu Rev Psychol*. 2013:64(1):135–168. https://doi.org/10.1146/annurev-psych-113011-143750.

Duncan J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci*. 2010:14(4):172–179. https://doi.org/10.1016/j.tics.2010.01.004.

Duncan J, Owen AM. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci*. 2000:23(10):475–483. https://doi.org/10.1016/S0166-2236(00)01633-7.

Eickhoff SB, Klaas SE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*. 2005:25(4):1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034.

Eickhoff SB, Paus T, Caspers S, Grosbras M-H, Evans AC, Zilles K, Amunts K. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage*. 2007:36(3):511–521. https://doi.org/10.1016/j.neuroimage.2007.03.060.

Ferguson MA, Anderson JS, Spreng RN. Fluid and flexible minds: intelligence reflects synchrony in the brain's intrinsic network architecture. *Network Neurosci*. 2017:1(2):192–207. https://doi.org/10.1162/NETN_a_00010.

Finn ES. Is it time to put rest to rest? *Trends Cogn Sci*. 2021:25(12):1021–1032. https://doi.org/10.1016/j.tics.2021.09.005.

Finn ES, Bandettini PA. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage*. 2021:235:117963. https://doi.org/10.1016/j.neuroimage.2021.117963.

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci*. 2015:18(11):1664–1671. https://doi.org/10.1038/nn.4135.

Fukushima M, Betzel RF, He Y, de Reus MA, van den Heuvel MP, Zuo X-N, Sporns O. Fluctuations between high- and low-modularity topology in time-resolved functional connectivity. *NeuroImage*. 2018:180:406–416. https://doi.org/10.1016/j.neuroimage.2017.08.044.

Gilson M, Deco G, Friston KJ, Hagmann P, Mantini D, Betti V, Romani GL, Corbetta M. Effective connectivity inferred from fMRI transition dynamics during movie viewing points to a balanced reconfiguration of cortical interactions. *NeuroImage*. 2018:180:534–546. https://doi.org/10.1016/j.neuroimage.2017.09.061.

Greene AS, Gao S, Scheinost D, Constable RT. Task-induced brain state manipulation improves prediction of individual traits. *Nat Commun*. 2018:9(1):2807. https://doi.org/10.1038/s41467-018-04920-3.

Gregory MD, Robertson EM, Manoach DS, Stickgold R. Thinking about a task is associated with increased connectivity in regions activated by task performance. *Brain Connectivity*. 2016:6(2):164–168. https://doi.org/10.1089/brain.2015.0386.

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*. 2014:95: 232–247. https://doi.org/10.1016/j.neuroimage.2014.03.034.

Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, Brensinger C, Gur RE. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *J Neurosci Methods*. 2010:187(2):254–262. https://doi.org/10.1016/j.jneumeth.2009.11.017.

Hausman HK, O'Shea A, Kraft JN, Boutzoukas EM, Evangelista ND, Van Etten EJ, Bharadwaj PK, Smith SG, Porges E, Hishaw GA, et al. The role of resting-state network functional connectivity in cognitive aging. *Front Aging Neurosci*. 2020:12:177. https://doi.org/10.3389/fnagi.2020.00177.

He L, Liu W, Zhuang K, Meng J, Qiu J. Executive function-related functional connectomes predict intellectual abilities. *Intelligence*. 2021:85:101527. https://doi.org/10.1016/j.intell.2021.101527.

Heckner MK, Cieslik EC, Küppers V, Fox PT, Eickhoff SB, Langner R. Delineating visual, auditory and motor regions in the human brain with functional neuroimaging: a BrainMap-based meta-analytic synthesis. *Sci Rep*. 2021:11(1):9942. https://doi.org/10.1038/s41598-021-88773-9.

Homack S, Lee D, Riccio CA. Test review: Delis-Kaplan executive function system. *Journal of Clinical and Experimental Neuropsychology*. 2005:27(5):599–609. https://doi.org/10.1080/13803390490918444.

Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial least squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage*. 2011:56(2):455–475. https://doi.org/10.1016/j.neuroimage.2010.07.034.

Kupis L, Goodman ZT, Kornfeld S, Hoang S, Romero C, Dirks B, Dehoney J, Chang C, Spreng RN, Nomi JS, et al. Brain dynamics underlying cognitive flexibility across the lifespan. *Cereb Cortex*. 2021:31(11):5263–5274. https://doi.org/10.1093/cercor/bhab156.

Langner R, Cieslik EC, Behrwind SD, Roski C, Caspers S, Amunts K, Eickhoff SB. Aging and response conflict solution: behavioural and functional connectivity changes. *Brain Struct Funct*. 2015:220(3):1739–1757. https://doi.org/10.1007/s00429-014-0758-0.

Langner R, Leiberg S, Hoffstaedter F, Eickhoff SB. Towards a human self-regulation system: common and distinct neural signatures of emotional and behavioural control. *Neurosci Biobehav Rev*. 2018:90:400–410. https://doi.org/10.1016/j.neubiorev.2018.04.022.

Larabi DI, Gell M, Amico E, Eickhoff SB, Patil KR. Highly accurate local functional fingerprints and their stability. *bioRxiv*. 2021, 2021:08(03):454862. https://doi.org/10.1101/2021.08.03.454862.

Li J. Assessing the accuracy of predictive models for numerical data: not r nor r2, why not? Then what? *PLoS One*. 2017:12(8):e0183250. https://doi.org/10.1371/journal.pone.0183250.

Li M, Wang D, Ren J, Langs G, Stoecklein S, Brennan BP, Lu J, Chen H, Liu H. Performing group-level functional image analyses based on homologous functional regions mapped in individuals. *PLoS Biol*. 2019:17(3):e2007032. https://doi.org/10.1371/journal.pbio.2007032.

Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, Donohue MR, Foran W, Miller RL, Hendrickson TJ, et al. Publisher correction: reproducible brain-wide association studies require thousands of individuals. *Nature*. 2022:605(7911):E11–E11. https://doi.org/10.1038/s41586-022-04692-3.

Miller G, Chapman J. Misunderstanding analysis of covariance. *J Abnorm Psychol*. 2001:110(1):40–48. https://doi.org/10.1037//0021-843x.110.1.40.

Miyake A, Friedman NP. The nature and organization of individual differences in executive functions: four general conclusions. *Curr Dir Psychol Sci*. 2012:21(1):8–14. https://doi.org/10.1177/0963721411429458.

Nadeau C, Bengio Y. Inference for the generalization error. *Mach Learn*. 2003:52(3):239–281. https://doi.org/10.1023/A:1024068626366.

Ng KK, Lo JC, Lim JKW, Chee MWL, Zhou J. Reduced functional segregation between the default mode network and the executive control network in healthy older adults: a longitudinal study. *NeuroImage*. 2016:133:321–330. https://doi.org/10.1016/j.neuroimage.2016.03.029.

Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. *NeuroImage*. 2019:203:116157. https://doi.org/10.1016/j.neuroimage.2019.116157.

Nooner K, Colcombe S, Tobe R, Mennes M, Benedict M, Moreno A, Panek L, Brown S, Zavitz S, Li Q, et al. The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Front Neurosci*. 2012:6:152. https://doi.org/10.3389/fnins.2012.00152.

Park DC, Reuter-Lorenz P. The adaptive brain: aging and neurocognitive scaffolding. *Annu Rev Psychol*. 2008:60(1):173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656.

Pat N, Wang Y, Bartonicek A, Candia J, Stringaris A. Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition. *Cereb Cortex*. 2022:bhac235. https://doi.org/10.1093/cercor/bhac235.

Persson J, Lustig C, Nelson JK, Reuter-Lorenz PA. Age differences in deactivation: a link to cognitive control? *J Cogn Neurosci*. 2007:19(6):1021–1032. https://doi.org/10.1162/jocn.2007.19.6.1021.

Pläschke RN, Patil KR, Cieslik EC, Nostro AD, Varikuti DP, Plachti A, Lösche P, Hoffstaedter F, Kalenscher T, Langner R, et al. Age differences in predicting working memory performance from network-based functional connectivity. *Cortex*. 2020:132:441–459. https://doi.org/10.1016/j.cortex.2020.08.012.

Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry*. 2020:77(5):534–540. https://doi.org/10.1001/jamapsychiatry.2019.3671.

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, et al. Functional network organization of the human brain. *Neuron*. 2011:72(4): 665–678. https://doi.org/10.1016/j.neuron.2011.09.006.

Prakash RS, Heo S, Voss MW, Patterson B, Kramer AF. Age-related differences in cortical recruitment and suppression: implications for cognitive performance. *Behav Brain Res*. 2012:230(1):192–200. https://doi.org/10.1016/j.bbr.2012.01.058.

Raichle ME. The Brain's default mode network. *Annu Rev Neurosci*. 2015:38(1):433–447. https://doi.org/10.1146/annurev-neuro-071013-014030.

Reineberg AE, Andrews-Hanna JR, Depue BE, Friedman NP, Banich MT. Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage*. 2015:104:69–78. https://doi.org/10.1016/j.neuroimage.2014.09.045.

Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, Fox PT, Eickhoff SB. Modelling neural correlates of working memory: a coordinate-based meta-analysis. *NeuroImage*. 2012:60(1):830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050.

Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*. 2014:90:449–468. https://doi.org/10.1016/j.neuroimage.2013.11.046.

Satterthwaite TD, Wolf DH, Erus G, Ruparel K, Elliott MA, Gennatas ED, Hopson R, Jackson C, Prabhakaran K, Bilker WB, et al. Functional maturation of the executive system during adolescence. *J Neurosci*. 2013:33(41):16249. https://doi.org/10.1523/JNEUROSCI.2345-13.2013.

Scheinost D, Noble S, Horien C, Greene AS, Lake EMR, Salehi M, Gao S, Shen X, O'Connor D, Barron DS, et al. Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*. 2019:193:35–45. https://doi.org/10.1016/j.neuroimage.2019.02.057.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci*. 2007:27(9):2349. https://doi.org/10.1523/JNEUROSCI.5587-06.2007.

Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, Constable RT. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat Protoc*. 2017:12(3):506–518. https://doi.org/10.1038/nprot.2016.178.

Shine JM, Bissett PG, Bell PT, Koyejo O, Balsters JH, Gorgolewski KJ, Moodie CA, Poldrack RA. The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron*. 2016:92(2):544–554. https://doi.org/10.1016/j.neuron.2016.09.018.

Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, et al. Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci U S A*. 2009:106(31):13040. https://doi.org/10.1073/pnas.0905267106.

Sripada C, Angstadt M, Rutherford S, Taxali A, Shedden K. Toward a "treadmill test" for cognition: improved prediction of general cognitive ability from the task activated brain. *Hum Brain Mapp*. 2020:41(12):3186–3197. https://doi.org/10.1002/hbm.25007.

Steffener J, Brickman AM, Rakitin BC, Gazes Y, Stern Y. The impact of age-related changes on working memory functional activity. *Brain Imaging Behav*. 2009:3(2):142–153. https://doi.org/10.1007/s11682-008-9056-x.

Tailby C, Masterton RAJ, Huang JY, Jackson GD, Abbott DF. Resting state functional connectivity changes induced by prior brain state are not network specific. *NeuroImage*. 2015:106:428–440. https://doi.org/10.1016/j.neuroimage.2014.11.037.

Taxali A, Angstadt M, Rutherford S, Sripada C. 2021. Boost in test–retest reliability in resting state fMRI with predictive modeling. *Cereb Cortex* 31(6):2822–2833. doi:https://doi.org/10.1093/cercor/bhaa390.

Teuber H. Unity and diversity of frontal lobe functions. *Acta Neurobiol Exp*. 1972:32(2):615–656.

Turner GR, Spreng RN. 2015. Prefrontal engagement and reduced default network suppression co-occur and are dynamically coupled in older adults: the default–executive coupling hypothesis of aging. *J Cogn Neurosci* 27(12):2462–2476. doi:https://doi.org/10.1162/jocn_a_00869.

Varangis E, Habeck CG, Razlighi QR, Stern Y. The effect of aging on resting state connectivity of predefined networks in the brain. *Front Aging Neurosci*. 2019:11. https://www.frontiersin.org/article/10.3389/fnagi.2019.00234.

Wang D, Liu H. Functional connectivity architecture of the human brain: not all the same. *Neuroscientist*. 2014:20(5):432–438. https://doi.org/10.1177/1073858414543290.

Worringer B, Langner R, Koch I, Eickhoff SB, Eickhoff CR, Binkofski FC. Common and distinct neural correlates of dual-tasking and task-switching: a meta-analytic review and a neuro-cognitive processing model of human multitasking. *Brain Struct Funct*. 2019:224(5):1845–1869. https://doi.org/10.1007/s00429-019-01870-4.