

Towards a “Treadmill Test” for Cognition: Improved Prediction of General Cognitive Ability from the Task Activated Brain

Chandra Sripada^{1*}, Mike Angstadt¹, Saige Rutherford¹, Aman Taxali¹, Kerby Sheddon

¹Department of Psychiatry, University of Michigan, Ann Arbor, MI, ²Department of Statistics, University of Michigan, Ann Arbor, MI

*Correspondence: sripada@umich.edu

Abstract

General cognitive ability (GCA) refers to a trait-like ability that contributes to performance across diverse cognitive tasks. Identifying brain-based markers of GCA has been a longstanding goal of cognitive and clinical neuroscience. Recently, predictive modeling methods have emerged that build whole-brain, distributed neural signatures for phenotypes of interest. In this study, we employ a predictive modeling approach to predict general cognitive ability based on fMRI task activation patterns during the *N*-back working memory task as well as six other tasks in the Human Connectome Project dataset ($n=967$), encompassing 15 task contrasts in total. We found tasks are a highly effective basis for prediction of GCA: The *2-back* versus *0-back* contrast achieved a 0.50 correlation with GCA scores in ten-fold cross-validation, and 13 out of 15 task contrasts afforded statistically significant prediction of GCA. Additionally, we found that task contrasts that produce greater fronto-parietal activation and default mode network deactivation—a brain activation pattern associated with executive processing and higher cognitive demand—are more effective in prediction of GCA. These results suggest a picture analogous to treadmill testing for cardiac function: Placing the brain in a more cognitively demanding task state significantly improves brain-based prediction of GCA.

1 Introduction

In addition to particular abilities associated with individual cognitive tasks, there is substantial evidence for an overarching general ability involved in performance across a diverse range of tasks.¹⁻⁵ Test batteries composed of diverse tasks can yield accurate estimates of this general ability, which we here refer to as general cognitive

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/hbm.25007](https://doi.org/10.1002/hbm.25007)

ability (GCA).^{6,7} GCA is a fundamental dimension of individual differences and is a key contributor to a number of important academic, occupational, health, and well-being-related outcomes.⁸⁻¹³ There is thus substantial interest in understanding the neural basis of GCA as well as the nature of inter-individual neural differences.

Functional imaging studies of brain activation patterns during cognitive tasks have yielded important insights into the neural basis of GCA¹⁴⁻¹⁷. In one key line of investigation, researchers identified a multiple demand network that activates across an array of cognitive tasks¹⁸⁻²¹. This network is hypothesized to support domain-general functions such as working memory^{22,23} and cognitive control²⁴⁻²⁶ that contribute to performance across tasks irrespective of their specific content. Subsequent work found activation in key regions of this network, including dorsal lateral prefrontal cortex and superior parietal regions, are correlated with measures of GCA or closely related constructs^{14,15,27,28}.

A notable feature of many of these previous task-based studies is that they are mainly concerned with localization and correlation: they mainly seek to identify specific brain regions whose activation correlates with GCA. Recently, however, another important goal has emerged in cognitive neuroscience: prediction²⁹⁻³¹. Unlike mass univariate approaches that are especially good for localization, predictive modeling approaches use multivariate methods that identify distributed patterns across the brain (“neurosignatures”). These distributed neurosignatures are often substantially more strongly related to phenotypes of interest than individual features because the neurosignatures aggregate information from across the entire brain³². However, because multivariate methods for constructing these distributed neurosignatures are highly parametrized, they are prone to overfitting. Predictive models are thus typically assessed by how well they predict unseen data, usually through the use of cross-validation^{33,34}.

Predictive modeling has been employed with a number of imaging modalities, including structural maps³⁵ and resting state connectomes³⁶⁻³⁹, to predict GCA or closely related constructs. A notable feature of these studies is that they mainly examined relatively stable, enduring features of the brain—features that are largely independent of the person’s current cognitive state, and in particular their actual exercise of the cognitive abilities that are relevant to GCA. An alternative approach for building predictive models of GCA, which appears to be relatively less utilized (cf. ^{40,41}), employs a rationale similar to that for cardiac treadmill testing. This approach attempts to first place the brain in an activated state that engages the cognitive abilities associated with GCA. By activating the brain in this way,

individual-differences in the neural basis of GCA may be rendered more “visible” for a predictive model to detect (see ⁴² for a suggestion along these lines).

In the current study, we adopted this second approach. Utilizing the Human Connectome Project’s (HCP) 1200 release, we began by constructing a highly reliable measure of GCA from 10 measures from the NIH Toolbox and Penn Neurocognitive Battery³⁷. We then used a predictive modeling framework to examine prediction of GCA from contrast maps derived from the *N*-back working memory task as well as six other fMRI tasks (15 task contrasts in total). We demonstrate two things. First, task-based brain activation patterns allow highly reliable prediction of GCA, with performance appreciably higher than that typically reported in other neuroimaging modalities. Second, tasks that produce greater fronto-parietal activation and default mode network deactivation, which is associated with higher cognitive demand, are more effective at GCA prediction.

2 Methods

2.1 Subjects and Data Acquisition

All subjects and data were from the HCP-1200 release^{43,44} and all research was performed in accordance with relevant guidelines and regulations. Subjects provided informed consent, and recruitment procedures and informed consent forms, including consent to share de-identified data, were approved by the Washington University institutional review board. Subjects completed two runs each of seven scanner tasks across two fMRI sessions, using a 32-channel head coil on a 3T Siemens Skyra scanner (TR = 720ms, TE = 33.1ms, 72 slices, 2mm isotropic voxels, multiband acceleration factor = 8) with right-to-left and left-to-right phase encoding directions. Comprehensive details are available elsewhere on HCP’s overall neuroimaging approach^{43,45} and HCP’s task fMRI dataset⁴⁶.

For the construction of a GCA factor, all subjects with available data were included. This analysis included 1,192 subjects. For the brain imaging analysis, subjects were eligible to be included if they had available task data in MSMAll format (information about both folding as well as function are used for cross-subject alignment⁴⁷) for both runs of all seven tasks, had full behavioral data, and no more than 25% of their volumes in each run exceeded a framewise displacement threshold of 0.5mm. These exclusions resulted in a sample of 967 subjects.

2.2 Data Preparation

Data was preprocessed through the HCP minimally preprocessed pipeline, which is presented in detail by Glasser *et al.*⁴⁸ Briefly, the pipeline includes gradient unwarping, motion correction, fieldmap distortion correction, brain-boundary based linear registration of functional to structural images, non-linear registration to MNI152 space, and grand-mean intensity normalization. Data then entered a surfaced-based preprocessing stream, followed by grayordinate-based processing, which involves data from the cortical ribbon being projected to surface space and combined with subcortical volumetric data.

2.3 FMRI Tasks

We used contrasts from seven HCP tasks, described in brief in Table 1 (detailed descriptions are available elsewhere^{44,46}).

At the single subject-level, fixed-effects analyses were conducted using FSL's FEAT to estimate the average effects across runs within-participants, using 2mm surface smoothed data. Some tasks permitted multiple contrasts beyond the standard experimental versus control condition (e.g., *N*-back allows additional contrasts based on all four stimulus types). To reduce the complexity of the analysis and avoid loss of power from smaller number of trials, we focused on the standard contrasts associated with these tasks. The Language Task and Emotion Task lacked fixation blocks. Thus we included the main condition contrasts (e.g., math-story and faces-shapes), but we did not include each of these conditions versus baseline. A full list of filenames of the contrast maps used can be found in the Supplement, Table S1.

2.4 Constructing a GCA Factor

We conducted an exploratory factor analysis utilizing the strategy and associated code made available by Dubois and colleagues (https://github.com/adolphslab/HCP_MRI-behavior), who recently investigated prediction of GCA from resting state fMRI in the HCP dataset³⁷. Unadjusted scores from ten cognitive tasks for 1181 HCP subjects were included in the analysis (subjects with missing data or MMSE < 26 were excluded), including seven tasks from the NIH Toolbox (Dimensional Change Cart Sort, Flanker Task, List Sort Test, Picture Sequence Test, Picture Vocabulary Test, Pattern Completion Test, Oral Reading Recognition Test) and three tasks from the Penn Neurocognitive Battery (Penn Progressive Matrices, Penn Word Memory Test, Variable Short Penn Line Orientation Test), with additional details supplied in³⁷.

We applied Dubois and colleagues' code to this data, which uses the omega function in the psych (v 1.8.4) package⁴⁹ in R (v3.4.4). In particular, the code performs maximum likelihood-estimated exploratory factor analysis (specifying a bifactor

model), oblimin factor rotation, followed by a Schmid-Leiman transformation⁵⁰ to find general factor loadings.

To assess reliability, in a separate analysis, we re-ran the factor analysis excluding 46 subjects that had Test/Retest sessions available. We then estimated factor scores for both sessions for these subjects and calculated test/retest reliability via intraclass correlation (we used ICC(2,1) in the Shrout and Fleiss scheme⁵¹).

We performed the preceding factor analysis on the entire dataset to characterize the factor structure (see Results §3.1). But importantly, we in addition repeated the factor analysis multiple times, each time within a fold of a 10-fold cross-validation analysis (see §2.6). This was to ensure the complete separation of train and test datasets during cross-validation.

2.5 Brain Basis Set Modeling

Our aim was to predict each subject's GCA scores from each of 15 task contrasts. To accomplish this, we used Brain Basis Set (BBS) modeling, previously described in detail⁵² and presented here in brief (Figure 1). Note that BBS was applied separately to each of the 15 task contrasts, and thus the steps that follow are performed separately for each contrast.

BBS assumes a train/test split of the dataset (see §2.6 below). In the train dataset, each subject's task contrast was vectorized and then concatenated yielding an n subjects \times m voxels matrix. This matrix was then submitted to principal components analysis using the `pca` function in MATLAB (2015b), yielding $n-1$ components ordered by descending eigenvalues, of which we retained the top 75 components.

We selected 75 as the number of components to retain based on prior analysis in which we estimated the number of intrinsic dimensions associated with each task contrast. This was accomplished by submitting each of the task contrast matrices to the dimensionality estimation procedure of Levina and Bickel⁵³. This is a maximum likelihood estimation method based on distance between close neighbors, which we previously successfully applied to HCP resting state data⁵⁴. Dimensionality estimation found a mean of 72 dimensions across the 15 task contrasts. Because prior studies by our group⁵² showed small differences in the number of components make little difference in classifier performance, and to increase comparability with recent studies that used 75 components^{38,52}, we chose to use 75 components for each task.

Next, in the training dataset, we calculate the expression scores for each of the components for each subject by projecting their data onto the 75 principal components. We then fit a linear regression model with these expression scores as predictors and the phenotype of interest (i.e., GCA) as the outcome, saving \mathbf{B} , the 75 x 1 vector of fitted coefficients, for later use. In a test dataset, we again calculate the expression scores for each of the 75 components for each subject. Our predicted phenotype for each test subject is the dot product of \mathbf{B} learned from the training dataset with the vector of component expression scores for that subject. We assessed performance of BBS-based prediction of GCA by calculating the correlation between predicted versus actual GCA in the test sample.

2.6 10-Fold Cross-Validation

To assess of the performance of BBS-based prediction models, we used 10-fold cross-validation. Because there is family structure in the HCP dataset, we ensured that family members always appeared within a single partition (and thus in no cases was the BBS classifier trained on a member of a family and tested on another member of that family).

To ensure complete separation of the train and test datasets, in each fold of the cross-validation, we did the following in the train dataset: First, a PCA was performed on the task contrast yielding a 75-component basis set. Second, the exploratory factor analysis described in §2.2 was performed yielding GCA scores for each train subject. In addition, the betas representing factor loadings for each behavioral task were applied to the test dataset, yielding GCA scores for the test subjects.

2.7 Accounting for Covariates in a Cross-Validation Framework

In each fold of cross-validation, BBS models were trained in the train partition with the following covariates (similar to³⁷): age, age squared, handedness, gender, brain size, multiband reconstruction algorithm version number (HCP variables: Age_In_Yrs, Handedness, Gender, FS_BrainSeg_Vol, fMRI_3T_ReconVrs), and mean framewise displacement (mean FD; task-specific values were used) and mean FD squared. Thus, our generative model for the data had the following form:

$$(1) y_{train} = X_{train}\beta + Z_{train}\gamma + \varepsilon$$

where y_{train} = train set response variable, X_{train} = train set brain features design matrix, β = train set brain features regression coefficients, Z_{train} = train set covariate design matrix, γ = train set covariate regression coefficients, and ε is Gaussian mean zero error.

When this model is estimated, we are particularly interested in the relationship between the following two terms:

$$(2) y_{train} - Z_{train}\hat{\gamma}$$

$$(3) X_{train}\hat{\beta}$$

where, $\hat{\gamma}$ = estimated train set covariate regression coefficients and $\hat{\beta}$ = estimated train set brain features regression coefficients. Term (2) represents the response variable adjusted for the estimated effects of the nuisance covariates, while term (3) represents the prediction of this covariate-adjusted response variable based on brain features. To be clear, $X_{train}\hat{\beta}$ is a prediction of the *covariate-adjusted* response because $\hat{\beta}$ is learned in a model with covariates.

To assess this same relationship in the test dataset, we compute quantities analogous to (2) and (3) in the test dataset. But to maintain the strict separation between train and test datasets needed in cross-validation, we compute these quantities using the coefficients learned in the train dataset. Thus we examine the relationship between:

$$(4) y_{test} - Z_{test}\hat{\gamma}$$

$$(5) X_{test}\hat{\beta}$$

where Z_{test} = test set design matrix, and $\hat{\gamma}$ = covariate regression coefficients learned from the train dataset, X_{test} = test set brain features design matrix, and $\hat{\beta}$ = brain features regression coefficients learned from the train dataset.

2.8 Evaluation of Cross-Validation Performance

Overall performance across the 10-fold cross-validation was assessed in three ways. Our primary measure is based on correlation between observed covariate-adjusted outcome variable and predicted outcome variable:

$$correlation(\tilde{y}_{test}, \hat{y}_{test})$$

where \tilde{y}_{test} is term (4) above, i.e., the test set response variable adjusted for covariates based on coefficients learned in the train dataset, and \hat{y}_{test} is term (5) above, i.e., the predicted covariate-adjusted response variable for the test set. Correlations were computed for each fold. To obtain the average correlation across folds, the per-fold correlations were Fisher r to z transformed, the transformed

Author Manuscript

correlations were averaged across all folds, and then this average was z to r transformed. Confidence intervals were estimated as 95% t intervals based on the mean and standard deviation over cross-validation folds.

In addition, we report cross-validated coefficient of determination R_{cv}^2 and mean square error (MSE), which are calculated as follows:

$$R_{cv}^2 = 1 - \frac{\sum_{i=1}^n (\tilde{y}_i - \hat{y}_i)^2}{\sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}$$
$$MSE = \frac{1}{n-1} \sum_{i=1}^n (\tilde{y}_i - \hat{y}_i)^2$$

where \tilde{y}_i = covariate-adjusted response variable for the test set for subject i , \hat{y}_i = predicted covariate-adjusted response variable for the test set for subject i , \bar{y} = mean value of the response variable for the train set, and n = the number of test set subjects. We calculate these values for each fold and then average across folds.

2.9 Permutation Tests

To assess the statistical significance of BBS models, we used non-parametric permutation methods. The distribution under chance of correlations between BBS-based predictions of neurocognitive scores and observed neurocognitive scores was generated by randomly permuting the subjects' neurocognitive scores 10,000 times. At each iteration, we performed the 10-fold cross validation procedure described above, which includes refitting BBS models at each fold of the cross-validation. We then recalculated the average correlation across folds between predicted versus actual neurocognitive scores. The average correlation across folds that was actually observed was located in this null distribution in terms of rank, and statistical significance was set as this rank value divided by 10,000.

Since the BBS models fit at each iteration of the permutation test included covariates, the procedure of Freedman and Lane was followed⁵⁵. In brief, a BBS model was first estimated with nuisance covariates alone, residuals were formed and were permuted. The covariate effect of interest was then included in the subsequent model, creating an approximate realization of data under the null hypothesis, and the statistical test of interest was calculated on this data (see FSL Randomise <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise/Theory> for a neuroimaging implementation).

2.10 Consensus Predictive Maps for Visualization

We used BBS with 75 whole-brain components to make predictions about GCA. To help convey overall patterns across the entire BBS predictive model, we constructed “consensus” predictive maps. We first multiplied each component map with its associated beta from the fitted BBS model. Next, we summed across all 75 components yielding a single map, and z scored the entries.

2.11 Analysis of Resting State Connectomes

To help contextualize results from predictive modeling applied to task contrast data, we applied this same predictive modeling stream to resting state connectomes. Data used was from the HCP-1200 release^{43,44}. Four runs of resting state fMRI data (14.4 minutes each; two runs per day over two days) were acquired using the same acquisition sequence described above in §2.1. Processed volumetric data from the HCP minimal preprocessing pipeline including ICA-FIX denoising were used. Full details of these steps can be found in Glasser⁴⁵ and Salimi-Korshidi⁵⁶.

Data then went through a number of resting state processing steps, including a motion artifact removal steps comparable to the type B (i.e., recommended) stream of Siegel et al.⁵⁷. These steps include linear detrending, CompCor to extract and regress out the top 5 principal components of white matter and CSF⁵⁸, bandpass filtering from 0.1-0.01Hz, and motion scrubbing of frames that exceed a framewise displacement of 0.5mm. We next calculated spatially-averaged time series for each of 264 4.24mm radius ROIs from the parcellation of Power et al.⁵⁹. We then calculated Pearson’s correlation coefficients between each ROI. These were then transformed using Fisher’s r to z transformation.

Subjects consisted of those subjects included in the main task contrast analysis who had 4 complete resting state fMRI runs (14m 24s each). In addition, subjects with more than 10% of resting state frames censored were excluded. This resulted in 903 subjects who entered a BBS predictive modeling analysis for prediction of GCA scores using the same BBS approach that is described above.

3 Results

3.1. Constructing a GCA Factor from Ten HCP Behavioral Tasks

We began by fitting a bifactor model to the behavioral data for the entire dataset. Similar to the findings of Dubois and colleagues³⁷ who examined a largely overlapping set of subjects, this model fit the data very well (CFI=0.989; RMSEA=0.036; SRMR=0.0200; BIC=0.782). The solution, which included a general factor and four group factors, is depicted in Figure 2. Similar to Dubois and

colleagues, we interpret the four group factors as: 1) Crystallized Ability; 2) Processing Speed; 3) Memory; and 4) Visuospatial Ability.

The general factor, which we refer to throughout as the GCA factor and which is the focus of this report, accounts for 58.6% of the variance (coefficient omega hierarchical ω^{60}), while group factors account for 18.0% of the variance. Based on the 46 subjects in the retest dataset for HCP, test-retest reliability for GCA was found to be 0.78, which is conventionally classified as very good (we used ICC(2,1) in the Shrout and Fleiss scheme⁵¹).

3.2 Contrasts associated with the *N*-Back task are highly effective at predicting GCA.

Because working memory has been strongly and consistently linked with GCA^{48,61,62}, we first investigated prediction of GCA based on the *N*-back working memory task. We used BBS modeling with 75 components and a 10-fold cross-validation procedure. The average correlation across folds between predicted GCA and actual GCA was 0.50, which was highly statistically significant (permutation-based $p < 0.0001$, observed correlation was higher than all 10,000 in the permutation distribution).

Figure 3 shows the top three components based on statistical significance displayed so that greater expression of these components predicts higher GCA. These components include large activations in supplementary motor area (SMA), precuneus, and dlPFC, as well as deactivations in anterior default mode network (DMN). To convey “average” patterns across all 75 components, we constructed consensus predictive maps (see Methods) and they are displayed in Figure 3. These show additional patterns predictive of GCA, including deactivation of posterior cingulate cortex and fronto-polar cortex.

We next trained additional BBS models on the *2-back vs. baseline* and *0-back vs. baseline* contrasts. The correlation across folds of the 10-fold cross-validation procedure between predicted GCA and actual GCA was 0.48 and 0.35, respectively. The consensus predictive maps, shown in Figure 4, revealed an interesting change in directionality across these contrasts. For example, pre-SMA strongly predicts higher GCA in the *2-back vs. baseline* contrast but the reverse is true in the *0-back vs. baseline* contrast. Additionally, less activation (i.e., deactivation) of the anterior DMN predicts higher GCA in the *2-back vs. baseline* contrast, but the reverse is true in the *0-back vs. baseline* contrast.

3.3 Looking across all 15 task contrasts, tasks involving more executive processing and higher cognitive demand are more effective in predicting GCA

We next examined the remaining 12 contrasts from the other six HCP tasks. As with the *N*-back task, we constructed BBS models predicting GCA scores from each contrast, and assessed performance of these models in 10-fold cross-validation analysis.

Results are shown in Figure 5 and Table 2. Using permutation-based statistical testing with 10,000 permutations, we found that 13 out of the 15 task contrasts produced statistically significant predictions of GCA (shown in blue and orange in Figure 5). The 2-back vs. 0-back contrast was the most effective single task contrast for GCA prediction, achieving a 0.50 correlation with GCA scores in 10-fold cross-validation. Other tasks involving executive processing were top performers, including the 2-back vs. 0-back contrast from the *N*-back task, the *relational vs. match* contrast from the relational processing task, and the *math vs. story* contrast from the language processing task. Resting state connectomes yielded prediction accuracy of $r=0.26$. In comparison, 13 out of 15 task contrasts performed better.

3.4 Mean activation levels of FPN and DMN predict which task contrasts are effective for GCA prediction

A number of studies have observed that tasks that are cognitively demanding produce activation in regions of frontoparietal network (FPN)^{18,25,26,63} and deactivation of regions of default mode network (DMN)⁶⁴⁻⁶⁷. Building on these observations, we hypothesized that more cognitively demanding task contrasts (operationalized in terms of activation levels of FPN and DMN) should be more effective in predicting GCA. We extracted mean activation across the seven networks in Yeo and colleagues' parcellation⁶⁸ and examined correlations with accuracy of GCA prediction across the 15 task contrasts (*prediction accuracy is measured with the cross-validated correlation between observed and predicted GCA scores*). We found that FPN activation was indeed strongly and statistically significantly related to accuracy of GCA prediction ($r=0.68$, $p=0.006$, 95% CI 0.25-0.88). DMN activation was also (inversely) related to accuracy of GCA prediction ($r=-0.20$), but the correlation did not reach statistical significance. We also created a regression model in which both FPN and DMN activation jointly predict accuracy of GCA prediction. The correlation across task contrasts between fitted predictions from the regression model and actual accuracy in predicting GCA was $r=0.82$ ($p = 0.001$; Figure 6). None of the other five Yeo networks were statistically significantly related to GCA prediction.

3.5. Across the 15 task contrasts, activation signatures of GCA are spatially distributed and task-specific

We next compared the consensus predictive maps associated with the 15 contrasts (five maps are shown in Figure 4, and the remaining maps are shown in Supplemental Figure S1). Signatures for predicting GCA associated with each task were highly distributed, with notable variation in these signatures across tasks. Prominently represented regions include: superior parietal cortex (reward vs. baseline, punishment vs. baseline), dlPFC (*math vs. story*), anterior insula (*relational v. match*), fronto-polar cortex (*math vs. story*), pre-SMA (*relational vs. match*), and visual cortex (*relational vs. match, reward vs. baseline, punishment vs. baseline*).

Of note, consensus predictive maps for all BBS models associated with each of the 15 task contrasts have been shared on Balsa, the Human Connectome Projects' website for sharing and hosting neuroimaging datasets, and can be accessed here: <https://balsa.wustl.edu/study/show/MZPv>.

4 Discussion

Task-based imaging provides a promising route for constructing brain-based predictive models of general cognitive ability (GCA) because tasks can potentially selectively activate brain regions responsible for effective cognitive performance. Thus, we systematically assessed neuroimaging-based prediction of GCA from 15 fMRI task conditions in the HCP dataset. Our first main finding is that whole-brain task activation patterns are a highly effective basis for prediction of GCA, with a model trained on activation during the *N*-back working memory task (*2-back vs. 0-back* contrast) achieving a 0.50 correlation with GCA scores in 10-fold cross-validation. Our second main finding is that more cognitively demanding tasks that more vigorously activate FPN and deactivate DMN are particularly effective for GCA prediction. These results highlight the utility of placing the brain in a cognitively demanding, activated task state for improved brain-based prediction of GCA.

Role of executive regions in prediction of GCA

The importance of FPN, as well as related executive regions (e.g., dorsal anterior cingulate), for GCA has been highlighted in previous work, especially in Jung and Haier's influential fronto-parietal integration theory⁶⁹. In a similar vein, Duncan, Owen, Fedorenko, and colleagues have proposed that "multiple demand" cortex—regions of the brain that activate across a broad range of cognitively demanding tasks¹⁸⁻²¹—are a primary substrate of GCA¹⁴. The present study extends these findings using a multivariate predictive modeling framework that identifies distributed neurosignatures across the brain that are predictive of GCA. We showed

that executive regions are important in these distributed neurosignatures in three complementary ways.

First, in looking across the set of 15 contrasts derived from seven HCP tasks, we found that tasks that tap executive processes were more predictive of GCA (e.g., *N-back 2-back vs. 0-back* contrast, relational reasoning *relational vs. match* contrast, and *math vs. story* contrast). Second, we found that FPN activation and DMN deactivation, highly associated with the cognitive demandingness of task conditions^{18,25,26,63–67}, predicts which task contrasts will be effective for GCA prediction. Third, within highly predictive contrasts, such the *2-back vs. 0-back* contrast and *math vs. story* contrast, activation patterns in executive regions were prominent among regions predictive of GCA.

Overall, the *N-back 2-back vs. 0-back* contrast performed best in GCA prediction. This is consistent with the finding that working memory is highly related to GCA^{48,61,62}. However, the differences in performance between the three main executive task contrasts—i.e., *2-back vs. 0-back*, *math vs. story*, and *relational vs. match*—were modest. Future studies with larger sample sizes should investigate whether all executive tasks are similarly effective with respect to GCA prediction, which would align well with the multiple demand network hypothesis. Or alternatively, there are subtle differences across executive tasks in affording GCA prediction.

Interestingly, for certain regions, the directionality of prediction of GCA exhibited some variability across task contrasts in a way suggestive of moderation by task difficulty (for example, see pre-SMA in *0-back* compared to *2-back* and in *match* compared to *relational*; we discuss moderation by cognitive load in these tasks further in⁷²). These observations are consistent with a neural efficiency model of GCA proposed by Neubauer and Fink⁷³. They propose that higher GCA is associated with greater processing efficiency in elementary cognitive tasks (leading to less activation in higher GCA individuals) but greater processing capacity in demanding cognitive tasks (leading to greater activation in higher GCA people), thus potentially explaining the flipped directions of activation observed across the easy and hard conditions of the *N-back* and other tasks.

While activation patterns in executive regions clearly play an important role in explaining the success of our task-based approach to GCA prediction, there is still clear evidence for discriminative information about GCA located outside executive regions. This is apparent in looking at the consensus predictive maps for each of the 15 task contrasts in Figure 4 as well Figure S1 in the Supplement. Non-executive

regions, such as lateral temporal cortex and temporal pole, are found in several of these consensus maps, indicating they too are important for the prediction of GCA.

Comparison of task-based prediction with other modalities

Previous studies have examined correlations between GCA and structural brain imaging features including cortical thickness^{74,75} and white matter structure⁷⁶, for reviews see ^{17,69,77}. It is notable that the correlations reported with these modalities tend to be modest. For example, the correlations with brain volume, one of the most studied variables, are typically reported to be between 0.1 to 0.3^{78,79}. In terms of functional MRI, recent studies have examined resting state connectivity patterns^{36-38,52}. In the present study, we found resting state connectomes, which entered the same BBS prediction pipeline as our task-based contrast maps, achieved a correlation of 0.26 with GCA (broadly similar to the results from our recent study using BBS modeling to predict neurocognition from resting state connectomes in 2,013 youth³⁸). These results, however, are appreciably smaller than the 0.50 correlation we found when applying BBS predictive modeling to the *2-back* vs. *0-back* task contrast in the present study.

There are two interrelated reasons why task-based fMRI might potentially offer more reliable prediction of GCA than other imaging modalities. The first appeals to the “treadmill testing” idea already mentioned: actively engaging in cognitive tasks has the potential to unmask critical GCA-relevant features of the brain that are otherwise invisible in other modalities such as structural or resting state brain imaging^{41,42}. A second potential advantage of task-based methods is specificity. Tasks are constructed by their designers to target specific psychological processes, often with control conditions that subtract away contributions from auxiliary processes of no interest. This will tend to make classification more accurate as the feature set is culled of a sizable number of uninformative features.

Future Directions

While we found strong predictivity of GCA from fMRI task contrasts, even the strongest performing task contrast explained only 28% of the variance (R_{cv}) in GCA scores. Thus, the majority of variance in GCA scores remains to be explained, which raises the question of how we might improve performance in future studies. In considering this question, it is notable that we used the set of imaging tasks that were included in the HCP dataset. These imaging tasks, in turn, were selected based on diverse considerations (see ⁴⁶), but maximizing prediction of GCA was not among them. Thus, it is plausible that one can do still better: It should be possible to intentionally design and optimize an imaging task battery to yield even more accurate task-based prediction of GCA.

Given our observation that tasks that more vigorously activate FPN and deactivate DMN afford better prediction of GCA, a natural approach is to focus on highly demanding tasks that produce this activation profile. One natural candidate is an N-back task with increased cognitive load (e.g., a 3-back^{80,81} or 4-back task) Other executive function tasks, such as tasks involving response inhibition, task switching, or higher-order reasoning, are also plausible. Moreover, it is possible that task contrasts from an executive task battery, as opposed to a contrast from a single task, could afford still better GCA prediction.

In sum, this study firmly establishes the effectiveness of task-based fMRI for prediction of GCA and demonstrates that tasks that are more cognitively demanding are associated with better prediction accuracy.

Data Availability Statement

Human Connectome Project data is publicly available to the research community at the project website: <http://www.humanconnectomeproject.org/>. Consensus component maps for all predictive models associated with each of the 15 task contrasts have been shared on BALSAs, the Human Connectome Projects' website for sharing and hosting neuroimaging datasets, and can be accessed here: <https://balsa.wustl.edu/study/show/MZPv>.

Competing Interests

The authors declare no conflicts of interest.

Funding

CS was supported by R01MH107741 and a grant from the Dana Foundation David Mahoney Neuroimaging Program.

References

1. Spearman, C. 'General Intelligence,' objectively determined and measured. *Am. J. Psychol.* **15**, 201–292 (1904).
2. Neisser, U. *et al.* Intelligence: Knowns and unknowns. *Am. Psychol.* **51**, 77 (1996).

3. Horn, J. L. & Noll, J. Human cognitive capabilities: Gf-Gc theory. (1997).
4. Mackintosh, N. & Mackintosh, N. J. *IQ and human intelligence*. (Oxford University Press, 2011).
5. Carroll, J. B. The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. in *The scientific study of general intelligence* 5–21 (Elsevier, 2003).
6. Carroll, J. B. *Human cognitive abilities: A survey of factor-analytic studies*. (Cambridge University Press, 1993).
7. McGrew, K. S. *CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research*. (Elsevier, 2009).
8. Gottfredson, L. S. Why g matters: The complexity of everyday life. *Intelligence* **24**, 79–132 (1997).
9. Strenze, T. Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* **35**, 401–426 (2007).
10. Ree, M. J., Earles, J. A. & Teachout, M. S. Predicting job performance: Not much more than g. *J. Appl. Psychol.* **79**, 518 (1994).
11. Whitley, E. *et al.* Intelligence in early adulthood and subsequent risk of assault: cohort study of 1,120,998 Swedish men. *Psychosom. Med.* **72**, 390 (2010).
12. Gale, C. R., Batty, G. D., Tynelius, P., Deary, I. J. & Rasmussen, F. Intelligence in early adulthood and subsequent hospitalisation and admission rates for the

whole range of mental disorders: longitudinal study of 1,049,663 men.

Epidemiol. Camb. Mass **21**, 70 (2010).

13. Batty, G. D., Mortensen, E. L. & Osler, M. Childhood IQ in relation to later psychiatric disorder: evidence from a Danish birth cohort study. *Br. J. Psychiatry* **187**, 180–181 (2005).
14. Duncan, J. *et al.* A neural basis for general intelligence. *Science* **289**, 457–460 (2000).
15. Gray, J. R., Chabris, C. F. & Braver, T. S. Neural mechanisms of general fluid intelligence. *Nat. Neurosci.* **6**, 316–322 (2003).
16. Schultz, D. H. & Cole, M. W. Higher intelligence is associated with less task-related brain network reconfiguration. *J. Neurosci.* **36**, 8551–8561 (2016).
17. Deary, I. J., Penke, L. & Johnson, W. The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.* **11**, 201 (2010).
18. Duncan, J. & Owen, A. M. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* **23**, 475–483 (2000).
19. Fedorenko, E., Duncan, J. & Kanwisher, N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci.* **110**, 16616–16621 (2013).
20. Duncan, J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* **14**, 172–179 (2010).

21. Shashidhara, S., Mitchell, D. J., Erez, Y. & Duncan, J. Progressive Recruitment of the Frontoparietal Multiple-demand System with Increased Task Complexity, Time Pressure, and Reward. *J. Cogn. Neurosci.* **31**, 1617–1630 (2019).
22. Baddeley, A. Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* **63**, 1–29 (2012).
23. D’Esposito, M., Postle, B. R. & Rypma, B. Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. in *Executive Control and the Frontal Lobe: Current Issues* (eds. Schneider, W. X., Owen, A. M. & Duncan, J.) 3–11 (Springer Berlin Heidelberg, 2000). doi:10.1007/978-3-642-59794-7_2.
24. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* **24**, 167–202 (2001).
25. Cole, M. W. & Schneider, W. The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage* **37**, 343–360 (2007).
26. Niendam, T. A. *et al.* Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn. Affect. Behav. Neurosci.* **12**, 241–268 (2012).
27. Lee, K. H. *et al.* Neural correlates of superior intelligence: stronger recruitment of posterior parietal cortex. *Neuroimage* **29**, 578–586 (2006).

28. DeYoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S. & Gray, J. R. Intellect as distinct from Openness: differences revealed by fMRI of working memory. *J. Pers. Soc. Psychol.* **97**, 883 (2009).
29. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
30. Rosenberg, M. D., Casey, B. J. & Holmes, A. J. Prediction complements explanation in understanding the developing brain. *Nat. Commun.* **9**, 1–13 (2018).
31. Varoquaux, G. & Poldrack, R. A. Predictive models avoid excessive reductionism in cognitive neuroimaging. *Curr. Opin. Neurobiol.* **55**, 1–6 (2019).
32. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20**, 365–377 (2017).
33. Scheinost, D. *et al.* Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage* (2019).
34. Poldrack, R. A., Huckins, G. & Varoquaux, G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry* (2019).
35. Cox, S. R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M. & Deary, I. J. Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence* **76**, 101376 (2019).

36. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).
37. Dubois, J., Galdi, P., Paul, L. K. & Adolphs, R. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, (2018).
38. Sripada, C. *et al.* Prediction of Neurocognition in Youth From Resting State fMRI. *Mol. Psychiatry* 1–19 (2019) doi:10.1101/495267.
39. Cui, Z. *et al.* Individual Variation in Functional Topography of Association Networks in Youth. *Neuron* (2020).
40. Stern, Y., Gazes, Y., Razlighi, Q., Steffener, J. & Habeck, C. A task-invariant cognitive reserve network. *Neuroimage* **178**, 36–45 (2018).
41. Greene, A. S., Gao, S., Scheinost, D. & Constable, R. T. Task-induced brain state manipulation improves prediction of individual traits. *Nat. Commun.* **9**, 1–13 (2018).
42. Finn, E. S. *et al.* Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage* (2017).
43. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80**, 62–79 (2013).
44. WU-Minn HCP. 1200 Subjects Data Release Reference Manual. (2017).

45. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).
46. Barch, D. M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
47. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
48. Engle, R. W., Kane, M. J. & Tuholski, S. W. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. (1999).
49. Revelle, W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. (2016).
50. Schmid, J. & Leiman, J. M. The development of hierarchical factor solutions. *Psychometrika* **22**, 53–61 (1957).
51. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**, 420–428 (1979).
52. Sripada, C. *et al.* Basic Units of Inter-Individual Variation in Resting State Connectomes. *Sci. Rep.* **9**, 1900 (2019).
53. Levina, E. & Bickel, P. J. Maximum Likelihood estimation of intrinsic dimension. in *Proceedings of the 17th International Conference on Neural Information Processing Systems* 777–784 (MIT Press, 2004).

54. Sripada, C. *et al.* Fundamental Differences: A Basis Set for Characterizing Inter-Individual Variation in Resting State Connectomes. *bioRxiv* 326082 (2018) doi:10.1101/326082.
55. Freedman, D. & Lane, D. A Nonstochastic Interpretation of Reported Significance Levels. *J. Bus. Econ. Stat.* **1**, 292–298 (1983).
56. Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage* **90**, 449–468 (2014).
57. Siegel, J. S. *et al.* Data Quality Influences Observed Links Between Functional Connectivity and Behavior. *Cereb. Cortex N. Y. N 1991* **27**, 4492–4502 (2017).
58. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* **37**, 90–101 (2007).
59. Power, J. D. *et al.* Functional Network Organization of the Human Brain. *Neuron* **72**, 665–678 (2011).
60. Zinbarg, R. E., Revelle, W., Yovel, I. & Li, W. Cronbach's α , Revelle's β , and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *psychometrika* **70**, 123–133 (2005).

61. Engle, R. W. & Kane, M. J. Executive attention, working memory capacity, and a two-factor theory of cognitive control. in *The Psychology of Learning and Motivation: Advances in Research and Theory* (Elsevier, 2004).
62. Duncan, J., Schramm, M., Thompson, R. & Dumontheil, I. Task rules, working memory, and fluid intelligence. *Psychon. Bull. Rev.* **19**, 864–870 (2012).
63. Cabeza, R. & Nyberg, L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci.* **12**, 1–47 (2000).
64. McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J. & Binder, J. R. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J Cogn Neurosci* **15**, 394–408 (2003).
65. Anticevic, A., Repovs, G., Shulman, G. L. & Barch, D. M. When less is more: TPJ and default network deactivation during encoding predicts working memory performance. *NeuroImage* **49**, 2638–2648 (2010).
66. Anticevic, A. *et al.* The role of default network deactivation in cognition and disease. *Trends Cogn. Sci.* **16**, 584–592 (2012).
67. Esposito, F. *et al.* Independent component model of the default-mode brain function: Assessing the impact of active thinking. *Brain Res. Bull.* **70**, 263–269 (2006).
68. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

69. Jung, R. E. & Haier, R. J. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* **30**, 135–154 (2007).
70. Conway, A. R., Kane, M. J. & Engle, R. W. Working memory capacity and its relation to general intelligence. *Trends Cogn. Sci.* **7**, 547–552 (2003).
71. Kyllonen, P. C. & Christal, R. E. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* **14**, 389–433 (1990).
72. Sripada, C., Angstadt, M., Rutherford, S. & Taxali, A. Brain Network Mechanisms of General Intelligence. *bioRxiv* 657205 (2019) doi:10.1101/657205.
73. Neubauer, A. C. & Fink, A. Intelligence and neural efficiency. *Neurosci. Biobehav. Rev.* **33**, 1004–1023 (2009).
74. Shaw, P. *et al.* Intellectual ability and cortical development in children and adolescents. *Nature* **440**, 676 (2006).
75. Colom, R. *et al.* Gray matter correlates of fluid, crystallized, and spatial intelligence: Testing the P-FIT model. *Intelligence* **37**, 124–135 (2009).
76. Turken, U. *et al.* Cognitive processing speed and the structure of white matter pathways: convergent evidence from normal variation and lesion studies. *Neuroimage* **42**, 1032–1044 (2008).
77. Luders, E., Narr, K. L., Thompson, P. M. & Toga, A. W. Neuroanatomical correlates of intelligence. *Intelligence* **37**, 156–163 (2009).

78. Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M. & Voracek, M. Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neurosci. Biobehav. Rev.* **57**, 411–432 (2015).
79. McDaniel, M. A. Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence* **33**, 337–346 (2005).
80. Braver, T. S. *et al.* A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* **5**, 49–62 (1997).
81. Pochon, J. B. *et al.* The neural system that bridges reward and cognition in humans: An fMRI study. *Proc. Natl. Acad. Sci.* **99**, 5669 (2002).

Figure Captions

Figure 1: Main Steps of Brain Basis Set (BBS) Modeling. *BBS is a multivariate predictive modeling method. It utilizes dimensionality reduction with principal components analysis (PCA) to construct a basis set for predicting phenotypes of interest.*

Figure 2: Bifactor Model of General Cognitive Ability. *We performed bifactor exploratory factor analysis on ten behavioral tasks in the Human Connectome Project (HCP) dataset. The resulting model consisted of a general factor (“GCA”) and four group factors and exhibited excellent fit with the data. C=Crystallized Cognitive Ability, S=Processing Speed, M=Memory, V=Visuospatial Ability.*

Figure 3: Visualization of the Three Components From the 2-Back vs. 0-Back Task Contrast Most Predictive of General Cognitive Ability (GCA). *We found the 2-back vs. 0-back contrast was highly effective for GCA prediction, achieving a 0.50 correlation with GCA scores in 10-fold cross-validation. From a 75-component brain basis set model trained to predict GCA scores, the three most statistically significant components are shown above.*

Figure 4: Consensus Predictive Maps for Five Task Contrasts Highly Predictive of General Cognitive Ability (GCA). *We found 13 out of 15 task contrast maps yielded highly statistically significant predictions of GCA in 10-fold cross-validation analysis. For the five most predictive task contrasts, we constructed consensus predictive maps that display brain activation patterns that were most predictive of GCA. Rel=Relational*

Figure 5: Prediction of General Cognitive Ability (GCA) Across 15 Task Contrasts. *We used the brain basis set (BBS) predictive modeling approach to predict GCA from each of 15 Human Connectome Project (HCP) task contrasts. The y-axis in the figure refers to accuracy of these BBS models in predicting GCA, as measured by the correlation between observed and predicted GCA scores in 10-fold cross-validation. For comparison, we additionally plot accuracy of GCA prediction using BBS methods applied to another modality: resting state connectomes. Error bars represent the 95% confidence interval; Blue = permutation-based $p < 0.0001$, observed correlation was higher than all 10,000 in the permutation distribution; Orange = permutation-based p -value < 0.05 ; Red = permutation-based p -value is not significant; TOM=Theory of Mind; Rel=Relational.*

Figure 6: FPN and DMN Activation Patterns and Effectiveness of Task Contrasts in Predicting General Cognitive Ability (GCA). *We hypothesized that placing the brain in an activated, cognitively demanding state improves prediction of GCA. We thus calculated FPN and DMN activation levels, which are thought to index cognitive demandingness, for each of the 15 task contrasts. We in addition calculated each of the 15 task contrast's accuracy in predicting GCA, as measured by the correlation between observed and predicted GCA scores in 10-fold cross-validation. In multiple regression analysis, we found that FPN/DMN activation levels for the 15 contrasts (x-axis) were indeed strongly related to the contrasts' accuracy in predicting GCA (y-axis). That is, contrasts that activated FPN/deactivated DMN more afforded higher accuracy in predicting GCA. Red dashed lines represent the 95% confidence interval.*

Towards a “Treadmill Test” for Cognition: Improved Prediction of General Cognitive Ability from the Task Activated Brain

Chandra Sripada^{1*}, Mike Angstadt¹, Saige Rutherford¹, Aman Taxali¹, Kerby Sheddon

¹Department of Psychiatry, University of Michigan, Ann Arbor, MI, ²Department of Statistics, University of Michigan, Ann Arbor, MI

*Correspondence: sripada@umich.edu

Abstract

General cognitive ability (GCA) refers to a trait-like ability that contributes to performance across diverse cognitive tasks. Identifying brain-based markers of GCA has been a longstanding goal of cognitive and clinical neuroscience. Recently, predictive modeling methods have emerged that build whole-brain, distributed neural signatures for phenotypes of interest. In this study, we employ a predictive modeling approach to predict general cognitive ability based on fMRI task activation patterns during the *N*-back working memory task as well as six other tasks in the Human Connectome Project dataset ($n=967$), encompassing 15 task contrasts in total. We found tasks are a highly effective basis for prediction of GCA: The *2-back* versus *0-back* contrast achieved a 0.50 correlation with GCA scores in ten-fold cross-validation, and 13 out of 15 task contrasts afforded statistically significant prediction of GCA. Additionally, we found that task contrasts that produce greater fronto-parietal activation and default mode network deactivation—a brain activation pattern associated with executive processing and higher cognitive demand—are more effective in prediction of GCA. These results suggest a picture analogous to treadmill testing for cardiac function: Placing the brain in a more cognitively demanding task state significantly improves brain-based prediction of GCA.

1 Introduction

In addition to particular abilities associated with individual cognitive tasks, there is substantial evidence for an overarching general ability involved in performance across a diverse range of tasks.¹⁻⁵ Test batteries composed of diverse tasks can yield accurate estimates of this general ability, which we here refer to as general cognitive

ability (GCA).^{6,7} GCA is a fundamental dimension of individual differences and is a key contributor to a number of important academic, occupational, health, and well-being-related outcomes.⁸⁻¹³ There is thus substantial interest in understanding the neural basis of GCA as well as the nature of inter-individual neural differences.

Functional imaging studies of brain activation patterns during cognitive tasks have yielded important insights into the neural basis of GCA¹⁴⁻¹⁷. In one key line of investigation, researchers identified a multiple demand network that activates across an array of cognitive tasks¹⁸⁻²¹. This network is hypothesized to support domain-general functions such as working memory^{22,23} and cognitive control²⁴⁻²⁶ that contribute to performance across tasks irrespective of their specific content. Subsequent work found activation in key regions of this network, including dorsal lateral prefrontal cortex and superior parietal regions, are correlated with measures of GCA or closely related constructs^{14,15,27,28}.

A notable feature of many of these previous task-based studies is that they are mainly concerned with localization and correlation: they mainly seek to identify specific brain regions whose activation correlates with GCA. Recently, however, another important goal has emerged in cognitive neuroscience: prediction²⁹⁻³¹. Unlike mass univariate approaches that are especially good for localization, predictive modeling approaches use multivariate methods that identify distributed patterns across the brain (“neurosignatures”). These distributed neurosignatures are often substantially more strongly related to phenotypes of interest than individual features because the neurosignatures aggregate information from across the entire brain³². However, because multivariate methods for constructing these distributed neurosignatures are highly parametrized, they are prone to overfitting. Predictive models are thus typically assessed by how well they predict unseen data, usually through the use of cross-validation^{33,34}.

Predictive modeling has been employed with a number of imaging modalities, including structural maps³⁵ and resting state connectomes³⁶⁻³⁹, to predict GCA or closely related constructs. A notable feature of these studies is that they mainly examined relatively stable, enduring features of the brain—features that are largely independent of the person’s current cognitive state, and in particular their actual exercise of the cognitive abilities that are relevant to GCA. An alternative approach for building predictive models of GCA, which appears to be relatively less utilized (cf. ^{40,41}), employs a rationale similar to that for cardiac treadmill testing. This approach attempts to first place the brain in an activated state that engages the cognitive abilities associated with GCA. By activating the brain in this way,

individual-differences in the neural basis of GCA may be rendered more “visible” for a predictive model to detect (see ⁴² for a suggestion along these lines).

In the current study, we adopted this second approach. Utilizing the Human Connectome Project’s (HCP) 1200 release, we began by constructing a highly reliable measure of GCA from 10 measures from the NIH Toolbox and Penn Neurocognitive Battery³⁷. We then used a predictive modeling framework to examine prediction of GCA from contrast maps derived from the *N*-back working memory task as well as six other fMRI tasks (15 task contrasts in total). We demonstrate two things. First, task-based brain activation patterns allow highly reliable prediction of GCA, with performance appreciably higher than that typically reported in other neuroimaging modalities. Second, tasks that produce greater fronto-parietal activation and default mode network deactivation, which is associated with higher cognitive demand, are more effective at GCA prediction.

2 Methods

2.1 Subjects and Data Acquisition

All subjects and data were from the HCP-1200 release^{43,44} and all research was performed in accordance with relevant guidelines and regulations. Subjects provided informed consent, and recruitment procedures and informed consent forms, including consent to share de-identified data, were approved by the Washington University institutional review board. Subjects completed two runs each of seven scanner tasks across two fMRI sessions, using a 32-channel head coil on a 3T Siemens Skyra scanner (TR = 720ms, TE = 33.1ms, 72 slices, 2mm isotropic voxels, multiband acceleration factor = 8) with right-to-left and left-to-right phase encoding directions. Comprehensive details are available elsewhere on HCP’s overall neuroimaging approach^{43,45} and HCP’s task fMRI dataset⁴⁶.

For the construction of a GCA factor, all subjects with available data were included. This analysis included 1,192 subjects. For the brain imaging analysis, subjects were eligible to be included if they had available task data in MSMAll format (information about both folding as well as function are used for cross-subject alignment⁴⁷) for both runs of all seven tasks, had full behavioral data, and no more than 25% of their volumes in each run exceeded a framewise displacement threshold of 0.5mm. These exclusions resulted in a sample of 967 subjects.

2.2 Data Preparation

Data was preprocessed through the HCP minimally preprocessed pipeline, which is presented in detail by Glasser *et al.*⁴⁸ Briefly, the pipeline includes gradient unwarping, motion correction, fieldmap distortion correction, brain-boundary based linear registration of functional to structural images, non-linear registration to MNI152 space, and grand-mean intensity normalization. Data then entered a surfaced-based preprocessing stream, followed by grayordinate-based processing, which involves data from the cortical ribbon being projected to surface space and combined with subcortical volumetric data.

2.3 FMRI Tasks

We used contrasts from seven HCP tasks, described in brief in Table 1 (detailed descriptions are available elsewhere^{44,46}).

At the single subject-level, fixed-effects analyses were conducted using FSL's FEAT to estimate the average effects across runs within-participants, using 2mm surface smoothed data. Some tasks permitted multiple contrasts beyond the standard experimental versus control condition (e.g., *N*-back allows additional contrasts based on all four stimulus types). To reduce the complexity of the analysis and avoid loss of power from smaller number of trials, we focused on the standard contrasts associated with these tasks. The Language Task and Emotion Task lacked fixation blocks. Thus we included the main condition contrasts (e.g., math-story and faces-shapes), but we did not include each of these conditions versus baseline. A full list of filenames of the contrast maps used can be found in the Supplement, Table S1.

2.4 Constructing a GCA Factor

We conducted an exploratory factor analysis utilizing the strategy and associated code made available by Dubois and colleagues (https://github.com/adolphslab/HCP_MRI-behavior), who recently investigated prediction of GCA from resting state fMRI in the HCP dataset³⁷. Unadjusted scores from ten cognitive tasks for 1181 HCP subjects were included in the analysis (subjects with missing data or MMSE < 26 were excluded), including seven tasks from the NIH Toolbox (Dimensional Change Cart Sort, Flanker Task, List Sort Test, Picture Sequence Test, Picture Vocabulary Test, Pattern Completion Test, Oral Reading Recognition Test) and three tasks from the Penn Neurocognitive Battery (Penn Progressive Matrices, Penn Word Memory Test, Variable Short Penn Line Orientation Test), with additional details supplied in³⁷.

We applied Dubois and colleagues' code to this data, which uses the omega function in the psych (v 1.8.4) package⁴⁹ in R (v3.4.4). In particular, the code performs maximum likelihood-estimated exploratory factor analysis (specifying a bifactor

model), oblimin factor rotation, followed by a Schmid-Leiman transformation⁵⁰ to find general factor loadings.

To assess reliability, in a separate analysis, we re-ran the factor analysis excluding 46 subjects that had Test/Retest sessions available. We then estimated factor scores for both sessions for these subjects and calculated test/retest reliability via intraclass correlation (we used ICC(2,1) in the Shrout and Fleiss scheme⁵¹).

We performed the preceding factor analysis on the entire dataset to characterize the factor structure (see Results §3.1). But importantly, we in addition repeated the factor analysis multiple times, each time within a fold of a 10-fold cross-validation analysis (see §2.6). This was to ensure the complete separation of train and test datasets during cross-validation.

2.5 Brain Basis Set Modeling

Our aim was to predict each subject's GCA scores from each of 15 task contrasts. To accomplish this, we used Brain Basis Set (BBS) modeling, previously described in detail⁵² and presented here in brief (Figure 1). Note that BBS was applied separately to each of the 15 task contrasts, and thus the steps that follow are performed separately for each contrast.

BBS assumes a train/test split of the dataset (see §2.6 below). In the train dataset, each subject's task contrast was vectorized and then concatenated yielding an n subjects \times m voxels matrix. This matrix was then submitted to principal components analysis using the `pca` function in MATLAB (2015b), yielding $n-1$ components ordered by descending eigenvalues, of which we retained the top 75 components.

We selected 75 as the number of components to retain based on prior analysis in which we estimated the number of intrinsic dimensions associated with each task contrast. This was accomplished by submitting each of the task contrast matrices to the dimensionality estimation procedure of Levina and Bickel⁵³. This is a maximum likelihood estimation method based on distance between close neighbors, which we previously successfully applied to HCP resting state data⁵⁴. Dimensionality estimation found a mean of 72 dimensions across the 15 task contrasts. Because prior studies by our group⁵² showed small differences in the number of components make little difference in classifier performance, and to increase comparability with recent studies that used 75 components^{38,52}, we chose to use 75 components for each task.

Next, in the training dataset, we calculate the expression scores for each of the components for each subject by projecting their data onto the 75 principal components. We then fit a linear regression model with these expression scores as predictors and the phenotype of interest (i.e., GCA) as the outcome, saving \mathbf{B} , the 75 x 1 vector of fitted coefficients, for later use. In a test dataset, we again calculate the expression scores for each of the 75 components for each subject. Our predicted phenotype for each test subject is the dot product of \mathbf{B} learned from the training dataset with the vector of component expression scores for that subject. We assessed performance of BBS-based prediction of GCA by calculating the correlation between predicted versus actual GCA in the test sample.

2.6 10-Fold Cross-Validation

To assess of the performance of BBS-based prediction models, we used 10-fold cross-validation. Because there is family structure in the HCP dataset, we ensured that family members always appeared within a single partition (and thus in no cases was the BBS classifier trained on a member of a family and tested on another member of that family).

To ensure complete separation of the train and test datasets, in each fold of the cross-validation, we did the following in the train dataset: First, a PCA was performed on the task contrast yielding a 75-component basis set. Second, the exploratory factor analysis described in §2.2 was performed yielding GCA scores for each train subject. In addition, the betas representing factor loadings for each behavioral task were applied to the test dataset, yielding GCA scores for the test subjects.

2.7 Accounting for Covariates in a Cross-Validation Framework

In each fold of cross-validation, BBS models were trained in the train partition with the following covariates (similar to³⁷): age, age squared, handedness, gender, brain size, multiband reconstruction algorithm version number (HCP variables: Age_In_Yrs, Handedness, Gender, FS_BrainSeg_Vol, fMRI_3T_ReconVrs), and mean framewise displacement (mean FD; task-specific values were used) and mean FD squared. Thus, our generative model for the data had the following form:

$$(1) y_{train} = X_{train}\beta + Z_{train}\gamma + \varepsilon$$

where y_{train} = train set response variable, X_{train} = train set brain features design matrix, β = train set brain features regression coefficients, Z_{train} = train set covariate design matrix, γ = train set covariate regression coefficients, and ε is Gaussian mean zero error.

When this model is estimated, we are particularly interested in the relationship between the following two terms:

$$(2) y_{train} - Z_{train}\hat{\gamma}$$

$$(3) X_{train}\hat{\beta}$$

where, $\hat{\gamma}$ = estimated train set covariate regression coefficients and $\hat{\beta}$ = estimated train set brain features regression coefficients. Term (2) represents the response variable adjusted for the estimated effects of the nuisance covariates, while term (3) represents the prediction of this covariate-adjusted response variable based on brain features. To be clear, $X_{train}\hat{\beta}$ is a prediction of the *covariate-adjusted* response because $\hat{\beta}$ is learned in a model with covariates.

To assess this same relationship in the test dataset, we compute quantities analogous to (2) and (3) in the test dataset. But to maintain the strict separation between train and test datasets needed in cross-validation, we compute these quantities using the coefficients learned in the train dataset. Thus we examine the relationship between:

$$(4) y_{test} - Z_{test}\hat{\gamma}$$

$$(5) X_{test}\hat{\beta}$$

where Z_{test} = test set design matrix, and $\hat{\gamma}$ = covariate regression coefficients learned from the train dataset, X_{test} = test set brain features design matrix, and $\hat{\beta}$ = brain features regression coefficients learned from the train dataset.

2.8 Evaluation of Cross-Validation Performance

Overall performance across the 10-fold cross-validation was assessed in three ways. Our primary measure is based on correlation between observed covariate-adjusted outcome variable and predicted outcome variable:

$$correlation(\tilde{y}_{test}, \hat{y}_{test})$$

where \tilde{y}_{test} is term (4) above, i.e., the test set response variable adjusted for covariates based on coefficients learned in the train dataset, and \hat{y}_{test} is term (5) above, i.e., the predicted covariate-adjusted response variable for the test set. Correlations were computed for each fold. To obtain the average correlation across folds, the per-fold correlations were Fisher r to z transformed, the transformed

Author Manuscript

correlations were averaged across all folds, and then this average was z to r transformed. Confidence intervals were estimated as 95% t intervals based on the mean and standard deviation over cross-validation folds.

In addition, we report cross-validated coefficient of determination R_{cv}^2 and mean square error (MSE), which are calculated as follows:

$$R_{cv}^2 = 1 - \frac{\sum_{i=1}^n (\tilde{y}_i - \hat{y}_i)^2}{\sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}$$
$$MSE = \frac{1}{n-1} \sum_{i=1}^n (\tilde{y}_i - \hat{y}_i)^2$$

where \tilde{y}_i = covariate-adjusted response variable for the test set for subject i , \hat{y}_i = predicted covariate-adjusted response variable for the test set for subject i , \bar{y} = mean value of the response variable for the train set, and n = the number of test set subjects. We calculate these values for each fold and then average across folds.

2.9 Permutation Tests

To assess the statistical significance of BBS models, we used non-parametric permutation methods. The distribution under chance of correlations between BBS-based predictions of neurocognitive scores and observed neurocognitive scores was generated by randomly permuting the subjects' neurocognitive scores 10,000 times. At each iteration, we performed the 10-fold cross validation procedure described above, which includes refitting BBS models at each fold of the cross-validation. We then recalculated the average correlation across folds between predicted versus actual neurocognitive scores. The average correlation across folds that was actually observed was located in this null distribution in terms of rank, and statistical significance was set as this rank value divided by 10,000.

Since the BBS models fit at each iteration of the permutation test included covariates, the procedure of Freedman and Lane was followed⁵⁵. In brief, a BBS model was first estimated with nuisance covariates alone, residuals were formed and were permuted. The covariate effect of interest was then included in the subsequent model, creating an approximate realization of data under the null hypothesis, and the statistical test of interest was calculated on this data (see FSL Randomise <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise/Theory> for a neuroimaging implementation).

2.10 Consensus Predictive Maps for Visualization

We used BBS with 75 whole-brain components to make predictions about GCA. To help convey overall patterns across the entire BBS predictive model, we constructed “consensus” predictive maps. We first multiplied each component map with its associated beta from the fitted BBS model. Next, we summed across all 75 components yielding a single map, and z scored the entries.

2.11 Analysis of Resting State Connectomes

To help contextualize results from predictive modeling applied to task contrast data, we applied this same predictive modeling stream to resting state connectomes. Data used was from the HCP-1200 release^{43,44}. Four runs of resting state fMRI data (14.4 minutes each; two runs per day over two days) were acquired using the same acquisition sequence described above in §2.1. Processed volumetric data from the HCP minimal preprocessing pipeline including ICA-FIX denoising were used. Full details of these steps can be found in Glasser⁴⁵ and Salimi-Korshidi⁵⁶.

Data then went through a number of resting state processing steps, including a motion artifact removal steps comparable to the type B (i.e., recommended) stream of Siegel et al.⁵⁷. These steps include linear detrending, CompCor to extract and regress out the top 5 principal components of white matter and CSF⁵⁸, bandpass filtering from 0.1-0.01Hz, and motion scrubbing of frames that exceed a framewise displacement of 0.5mm. We next calculated spatially-averaged time series for each of 264 4.24mm radius ROIs from the parcellation of Power et al.⁵⁹. We then calculated Pearson’s correlation coefficients between each ROI. These were then transformed using Fisher’s r to z transformation.

Subjects consisted of those subjects included in the main task contrast analysis who had 4 complete resting state fMRI runs (14m 24s each). In addition, subjects with more than 10% of resting state frames censored were excluded. This resulted in 903 subjects who entered a BBS predictive modeling analysis for prediction of GCA scores using the same BBS approach that is described above.

3 Results

3.1. Constructing a GCA Factor from Ten HCP Behavioral Tasks

We began by fitting a bifactor model to the behavioral data for the entire dataset. Similar to the findings of Dubois and colleagues³⁷ who examined a largely overlapping set of subjects, this model fit the data very well (CFI=0.989; RMSEA=0.036; SRMR=0.0200; BIC=0.782). The solution, which included a general factor and four group factors, is depicted in Figure 2. Similar to Dubois and

colleagues, we interpret the four group factors as: 1) Crystallized Ability; 2) Processing Speed; 3) Memory; and 4) Visuospatial Ability.

The general factor, which we refer to throughout as the GCA factor and which is the focus of this report, accounts for 58.6% of the variance (coefficient omega hierarchical ω^{60}), while group factors account for 18.0% of the variance. Based on the 46 subjects in the retest dataset for HCP, test-retest reliability for GCA was found to be 0.78, which is conventionally classified as very good (we used ICC(2,1) in the Shrout and Fleiss scheme⁵¹).

3.2 Contrasts associated with the *N*-Back task are highly effective at predicting GCA.

Because working memory has been strongly and consistently linked with GCA^{48,61,62}, we first investigated prediction of GCA based on the *N*-back working memory task. We used BBS modeling with 75 components and a 10-fold cross-validation procedure. The average correlation across folds between predicted GCA and actual GCA was 0.50, which was highly statistically significant (permutation-based $p < 0.0001$, observed correlation was higher than all 10,000 in the permutation distribution).

Figure 3 shows the top three components based on statistical significance displayed so that greater expression of these components predicts higher GCA. These components include large activations in supplementary motor area (SMA), precuneus, and dlPFC, as well as deactivations in anterior default mode network (DMN). To convey “average” patterns across all 75 components, we constructed consensus predictive maps (see Methods) and they are displayed in Figure 3. These show additional patterns predictive of GCA, including deactivation of posterior cingulate cortex and fronto-polar cortex.

We next trained additional BBS models on the *2-back vs. baseline* and *0-back vs. baseline* contrasts. The correlation across folds of the 10-fold cross-validation procedure between predicted GCA and actual GCA was 0.48 and 0.35, respectively. The consensus predictive maps, shown in Figure 4, revealed an interesting change in directionality across these contrasts. For example, pre-SMA strongly predicts higher GCA in the *2-back vs. baseline* contrast but the reverse is true in the *0-back vs. baseline* contrast. Additionally, less activation (i.e., deactivation) of the anterior DMN predicts higher GCA in the *2-back vs. baseline* contrast, but the reverse is true in the *0-back vs. baseline* contrast.

3.3 Looking across all 15 task contrasts, tasks involving more executive processing and higher cognitive demand are more effective in predicting GCA

We next examined the remaining 12 contrasts from the other six HCP tasks. As with the *N*-back task, we constructed BBS models predicting GCA scores from each contrast, and assessed performance of these models in 10-fold cross-validation analysis.

Results are shown in Figure 5 and Table 2. Using permutation-based statistical testing with 10,000 permutations, we found that 13 out of the 15 task contrasts produced statistically significant predictions of GCA (shown in blue and orange in Figure 5). The 2-back vs. 0-back contrast was the most effective single task contrast for GCA prediction, achieving a 0.50 correlation with GCA scores in 10-fold cross-validation. Other tasks involving executive processing were top performers, including the 2-back vs. 0-back contrast from the *N*-back task, the *relational vs. match* contrast from the relational processing task, and the *math vs. story* contrast from the language processing task. Resting state connectomes yielded prediction accuracy of $r=0.26$. In comparison, 13 out of 15 task contrasts performed better.

3.4 Mean activation levels of FPN and DMN predict which task contrasts are effective for GCA prediction

A number of studies have observed that tasks that are cognitively demanding produce activation in regions of frontoparietal network (FPN)^{18,25,26,63} and deactivation of regions of default mode network (DMN)⁶⁴⁻⁶⁷. Building on these observations, we hypothesized that more cognitively demanding task contrasts (operationalized in terms of activation levels of FPN and DMN) should be more effective in predicting GCA. We extracted mean activation across the seven networks in Yeo and colleagues' parcellation⁶⁸ and examined correlations with accuracy of GCA prediction across the 15 task contrasts (*prediction accuracy is measured with the cross-validated correlation between observed and predicted GCA scores*). We found that FPN activation was indeed strongly and statistically significantly related to accuracy of GCA prediction ($r=0.68$, $p=0.006$, 95% CI 0.25-0.88). DMN activation was also (inversely) related to accuracy of GCA prediction ($r=-0.20$), but the correlation did not reach statistical significance. We also created a regression model in which both FPN and DMN activation jointly predict accuracy of GCA prediction. The correlation across task contrasts between fitted predictions from the regression model and actual accuracy in predicting GCA was $r=0.82$ ($p = 0.001$; Figure 6). None of the other five Yeo networks were statistically significantly related to GCA prediction.

3.5. Across the 15 task contrasts, activation signatures of GCA are spatially distributed and task-specific

We next compared the consensus predictive maps associated with the 15 contrasts (five maps are shown in Figure 4, and the remaining maps are shown in Supplemental Figure S1). Signatures for predicting GCA associated with each task were highly distributed, with notable variation in these signatures across tasks. Prominently represented regions include: superior parietal cortex (reward vs. baseline, punishment vs. baseline), dlPFC (*math vs. story*), anterior insula (*relational v. match*), fronto-polar cortex (*math vs. story*), pre-SMA (*relational vs. match*), and visual cortex (*relational vs. match, reward vs. baseline, punishment vs. baseline*).

Of note, consensus predictive maps for all BBS models associated with each of the 15 task contrasts have been shared on Balsa, the Human Connectome Projects' website for sharing and hosting neuroimaging datasets, and can be accessed here: <https://balsa.wustl.edu/study/show/MZPv>.

4 Discussion

Task-based imaging provides a promising route for constructing brain-based predictive models of general cognitive ability (GCA) because tasks can potentially selectively activate brain regions responsible for effective cognitive performance. Thus, we systematically assessed neuroimaging-based prediction of GCA from 15 fMRI task conditions in the HCP dataset. Our first main finding is that whole-brain task activation patterns are a highly effective basis for prediction of GCA, with a model trained on activation during the *N*-back working memory task (*2-back vs. 0-back* contrast) achieving a 0.50 correlation with GCA scores in 10-fold cross-validation. Our second main finding is that more cognitively demanding tasks that more vigorously activate FPN and deactivate DMN are particularly effective for GCA prediction. These results highlight the utility of placing the brain in a cognitively demanding, activated task state for improved brain-based prediction of GCA.

Role of executive regions in prediction of GCA

The importance of FPN, as well as related executive regions (e.g., dorsal anterior cingulate), for GCA has been highlighted in previous work, especially in Jung and Haier's influential fronto-parietal integration theory⁶⁹. In a similar vein, Duncan, Owen, Fedorenko, and colleagues have proposed that "multiple demand" cortex—regions of the brain that activate across a broad range of cognitively demanding tasks¹⁸⁻²¹—are a primary substrate of GCA¹⁴. The present study extends these findings using a multivariate predictive modeling framework that identifies distributed neurosignatures across the brain that are predictive of GCA. We showed

that executive regions are important in these distributed neurosignatures in three complementary ways.

First, in looking across the set of 15 contrasts derived from seven HCP tasks, we found that tasks that tap executive processes were more predictive of GCA (e.g., *N-back 2-back vs. 0-back* contrast, relational reasoning *relational vs. match* contrast, and *math vs. story* contrast). Second, we found that FPN activation and DMN deactivation, highly associated with the cognitive demandingness of task conditions^{18,25,26,63-67}, predicts which task contrasts will be effective for GCA prediction. Third, within highly predictive contrasts, such the *2-back vs. 0-back* contrast and *math vs. story* contrast, activation patterns in executive regions were prominent among regions predictive of GCA.

Overall, the *N-back 2-back vs. 0-back* contrast performed best in GCA prediction. This is consistent with the finding that working memory is highly related to GCA^{48,61,62}. However, the differences in performance between the three main executive task contrasts—i.e., *2-back vs. 0-back*, *math vs. story*, and *relational vs. match*—were modest. Future studies with larger sample sizes should investigate whether all executive tasks are similarly effective with respect to GCA prediction, which would align well with the multiple demand network hypothesis. Or alternatively, there are subtle differences across executive tasks in affording GCA prediction.

Interestingly, for certain regions, the directionality of prediction of GCA exhibited some variability across task contrasts in a way suggestive of moderation by task difficulty (for example, see pre-SMA in *0-back* compared to *2-back* and in *match* compared to *relational*; we discuss moderation by cognitive load in these tasks further in⁷²). These observations are consistent with a neural efficiency model of GCA proposed by Neubauer and Fink⁷³. They propose that higher GCA is associated with greater processing efficiency in elementary cognitive tasks (leading to less activation in higher GCA individuals) but greater processing capacity in demanding cognitive tasks (leading to greater activation in higher GCA people), thus potentially explaining the flipped directions of activation observed across the easy and hard conditions of the *N-back* and other tasks.

While activation patterns in executive regions clearly play an important role in explaining the success of our task-based approach to GCA prediction, there is still clear evidence for discriminative information about GCA located outside executive regions. This is apparent in looking at the consensus predictive maps for each of the 15 task contrasts in Figure 4 as well Figure S1 in the Supplement. Non-executive

regions, such as lateral temporal cortex and temporal pole, are found in several of these consensus maps, indicating they too are important for the prediction of GCA.

Comparison of task-based prediction with other modalities

Previous studies have examined correlations between GCA and structural brain imaging features including cortical thickness^{74,75} and white matter structure⁷⁶, for reviews see ^{17,69,77}. It is notable that the correlations reported with these modalities tend to be modest. For example, the correlations with brain volume, one of the most studied variables, are typically reported to be between 0.1 to 0.3^{78,79}. In terms of functional MRI, recent studies have examined resting state connectivity patterns^{36-38,52}. In the present study, we found resting state connectomes, which entered the same BBS prediction pipeline as our task-based contrast maps, achieved a correlation of 0.26 with GCA (broadly similar to the results from our recent study using BBS modeling to predict neurocognition from resting state connectomes in 2,013 youth³⁸). These results, however, are appreciably smaller than the 0.50 correlation we found when applying BBS predictive modeling to the *2-back* vs. *0-back* task contrast in the present study.

There are two interrelated reasons why task-based fMRI might potentially offer more reliable prediction of GCA than other imaging modalities. The first appeals to the “treadmill testing” idea already mentioned: actively engaging in cognitive tasks has the potential to unmask critical GCA-relevant features of the brain that are otherwise invisible in other modalities such as structural or resting state brain imaging^{41,42}. A second potential advantage of task-based methods is specificity. Tasks are constructed by their designers to target specific psychological processes, often with control conditions that subtract away contributions from auxiliary processes of no interest. This will tend to make classification more accurate as the feature set is culled of a sizable number of uninformative features.

Future Directions

While we found strong predictivity of GCA from fMRI task contrasts, even the strongest performing task contrast explained only 28% of the variance (R_{cv}) in GCA scores. Thus, the majority of variance in GCA scores remains to be explained, which raises the question of how we might improve performance in future studies. In considering this question, it is notable that we used the set of imaging tasks that were included in the HCP dataset. These imaging tasks, in turn, were selected based on diverse considerations (see ⁴⁶), but maximizing prediction of GCA was not among them. Thus, it is plausible that one can do still better: It should be possible to intentionally design and optimize an imaging task battery to yield even more accurate task-based prediction of GCA.

Given our observation that tasks that more vigorously activate FPN and deactivate DMN afford better prediction of GCA, a natural approach is to focus on highly demanding tasks that produce this activation profile. One natural candidate is an N-back task with increased cognitive load (e.g., a 3-back^{80,81} or 4-back task) Other executive function tasks, such as tasks involving response inhibition, task switching, or higher-order reasoning, are also plausible. Moreover, it is possible that task contrasts from an executive task battery, as opposed to a contrast from a single task, could afford still better GCA prediction.

In sum, this study firmly establishes the effectiveness of task-based fMRI for prediction of GCA and demonstrates that tasks that are more cognitively demanding are associated with better prediction accuracy.

Data Availability Statement

Human Connectome Project data is publicly available to the research community at the project website: <http://www.humanconnectomeproject.org/>. Consensus component maps for all predictive models associated with each of the 15 task contrasts have been shared on BALSAs, the Human Connectome Projects' website for sharing and hosting neuroimaging datasets, and can be accessed here: <https://balsa.wustl.edu/study/show/MZPv>.

Competing Interests

The authors declare no conflicts of interest.

Funding

CS was supported by R01MH107741 and a grant from the Dana Foundation David Mahoney Neuroimaging Program.

References

1. Spearman, C. 'General Intelligence,' objectively determined and measured. *Am. J. Psychol.* **15**, 201–292 (1904).
2. Neisser, U. *et al.* Intelligence: Knowns and unknowns. *Am. Psychol.* **51**, 77 (1996).

3. Horn, J. L. & Noll, J. Human cognitive capabilities: Gf-Gc theory. (1997).
4. Mackintosh, N. & Mackintosh, N. J. *IQ and human intelligence*. (Oxford University Press, 2011).
5. Carroll, J. B. The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. in *The scientific study of general intelligence* 5–21 (Elsevier, 2003).
6. Carroll, J. B. *Human cognitive abilities: A survey of factor-analytic studies*. (Cambridge University Press, 1993).
7. McGrew, K. S. *CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research*. (Elsevier, 2009).
8. Gottfredson, L. S. Why g matters: The complexity of everyday life. *Intelligence* **24**, 79–132 (1997).
9. Strenze, T. Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* **35**, 401–426 (2007).
10. Ree, M. J., Earles, J. A. & Teachout, M. S. Predicting job performance: Not much more than g. *J. Appl. Psychol.* **79**, 518 (1994).
11. Whitley, E. *et al.* Intelligence in early adulthood and subsequent risk of assault: cohort study of 1,120,998 Swedish men. *Psychosom. Med.* **72**, 390 (2010).
12. Gale, C. R., Batty, G. D., Tynelius, P., Deary, I. J. & Rasmussen, F. Intelligence in early adulthood and subsequent hospitalisation and admission rates for the

whole range of mental disorders: longitudinal study of 1,049,663 men.

Epidemiol. Camb. Mass **21**, 70 (2010).

13. Batty, G. D., Mortensen, E. L. & Osler, M. Childhood IQ in relation to later psychiatric disorder: evidence from a Danish birth cohort study. *Br. J. Psychiatry* **187**, 180–181 (2005).
14. Duncan, J. *et al.* A neural basis for general intelligence. *Science* **289**, 457–460 (2000).
15. Gray, J. R., Chabris, C. F. & Braver, T. S. Neural mechanisms of general fluid intelligence. *Nat. Neurosci.* **6**, 316–322 (2003).
16. Schultz, D. H. & Cole, M. W. Higher intelligence is associated with less task-related brain network reconfiguration. *J. Neurosci.* **36**, 8551–8561 (2016).
17. Deary, I. J., Penke, L. & Johnson, W. The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.* **11**, 201 (2010).
18. Duncan, J. & Owen, A. M. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* **23**, 475–483 (2000).
19. Fedorenko, E., Duncan, J. & Kanwisher, N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci.* **110**, 16616–16621 (2013).
20. Duncan, J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* **14**, 172–179 (2010).

21. Shashidhara, S., Mitchell, D. J., Erez, Y. & Duncan, J. Progressive Recruitment of the Frontoparietal Multiple-demand System with Increased Task Complexity, Time Pressure, and Reward. *J. Cogn. Neurosci.* **31**, 1617–1630 (2019).
22. Baddeley, A. Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* **63**, 1–29 (2012).
23. D’Esposito, M., Postle, B. R. & Rypma, B. Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. in *Executive Control and the Frontal Lobe: Current Issues* (eds. Schneider, W. X., Owen, A. M. & Duncan, J.) 3–11 (Springer Berlin Heidelberg, 2000). doi:10.1007/978-3-642-59794-7_2.
24. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* **24**, 167–202 (2001).
25. Cole, M. W. & Schneider, W. The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage* **37**, 343–360 (2007).
26. Niendam, T. A. *et al.* Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn. Affect. Behav. Neurosci.* **12**, 241–268 (2012).
27. Lee, K. H. *et al.* Neural correlates of superior intelligence: stronger recruitment of posterior parietal cortex. *Neuroimage* **29**, 578–586 (2006).

28. DeYoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S. & Gray, J. R. Intellect as distinct from Openness: differences revealed by fMRI of working memory. *J. Pers. Soc. Psychol.* **97**, 883 (2009).
29. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
30. Rosenberg, M. D., Casey, B. J. & Holmes, A. J. Prediction complements explanation in understanding the developing brain. *Nat. Commun.* **9**, 1–13 (2018).
31. Varoquaux, G. & Poldrack, R. A. Predictive models avoid excessive reductionism in cognitive neuroimaging. *Curr. Opin. Neurobiol.* **55**, 1–6 (2019).
32. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20**, 365–377 (2017).
33. Scheinost, D. *et al.* Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage* (2019).
34. Poldrack, R. A., Huckins, G. & Varoquaux, G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry* (2019).
35. Cox, S. R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M. & Deary, I. J. Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence* **76**, 101376 (2019).

36. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).
37. Dubois, J., Galdi, P., Paul, L. K. & Adolphs, R. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, (2018).
38. Sripada, C. *et al.* Prediction of Neurocognition in Youth From Resting State fMRI. *Mol. Psychiatry* 1–19 (2019) doi:10.1101/495267.
39. Cui, Z. *et al.* Individual Variation in Functional Topography of Association Networks in Youth. *Neuron* (2020).
40. Stern, Y., Gazes, Y., Razlighi, Q., Steffener, J. & Habeck, C. A task-invariant cognitive reserve network. *Neuroimage* **178**, 36–45 (2018).
41. Greene, A. S., Gao, S., Scheinost, D. & Constable, R. T. Task-induced brain state manipulation improves prediction of individual traits. *Nat. Commun.* **9**, 1–13 (2018).
42. Finn, E. S. *et al.* Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage* (2017).
43. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80**, 62–79 (2013).
44. WU-Minn HCP. 1200 Subjects Data Release Reference Manual. (2017).

45. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).
46. Barch, D. M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
47. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
48. Engle, R. W., Kane, M. J. & Tuholski, S. W. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. (1999).
49. Revelle, W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. (2016).
50. Schmid, J. & Leiman, J. M. The development of hierarchical factor solutions. *Psychometrika* **22**, 53–61 (1957).
51. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**, 420–428 (1979).
52. Sripada, C. *et al.* Basic Units of Inter-Individual Variation in Resting State Connectomes. *Sci. Rep.* **9**, 1900 (2019).
53. Levina, E. & Bickel, P. J. Maximum Likelihood estimation of intrinsic dimension. in *Proceedings of the 17th International Conference on Neural Information Processing Systems* 777–784 (MIT Press, 2004).

54. Sripada, C. *et al.* Fundamental Differences: A Basis Set for Characterizing Inter-Individual Variation in Resting State Connectomes. *bioRxiv* 326082 (2018) doi:10.1101/326082.
55. Freedman, D. & Lane, D. A Nonstochastic Interpretation of Reported Significance Levels. *J. Bus. Econ. Stat.* **1**, 292–298 (1983).
56. Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage* **90**, 449–468 (2014).
57. Siegel, J. S. *et al.* Data Quality Influences Observed Links Between Functional Connectivity and Behavior. *Cereb. Cortex N. Y. N 1991* **27**, 4492–4502 (2017).
58. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* **37**, 90–101 (2007).
59. Power, J. D. *et al.* Functional Network Organization of the Human Brain. *Neuron* **72**, 665–678 (2011).
60. Zinbarg, R. E., Revelle, W., Yovel, I. & Li, W. Cronbach's α , Revelle's β , and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *psychometrika* **70**, 123–133 (2005).

61. Engle, R. W. & Kane, M. J. Executive attention, working memory capacity, and a two-factor theory of cognitive control. in *The Psychology of Learning and Motivation: Advances in Research and Theory* (Elsevier, 2004).
62. Duncan, J., Schramm, M., Thompson, R. & Dumontheil, I. Task rules, working memory, and fluid intelligence. *Psychon. Bull. Rev.* **19**, 864–870 (2012).
63. Cabeza, R. & Nyberg, L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci.* **12**, 1–47 (2000).
64. McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J. & Binder, J. R. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J Cogn Neurosci* **15**, 394–408 (2003).
65. Anticevic, A., Repovs, G., Shulman, G. L. & Barch, D. M. When less is more: TPJ and default network deactivation during encoding predicts working memory performance. *NeuroImage* **49**, 2638–2648 (2010).
66. Anticevic, A. *et al.* The role of default network deactivation in cognition and disease. *Trends Cogn. Sci.* **16**, 584–592 (2012).
67. Esposito, F. *et al.* Independent component model of the default-mode brain function: Assessing the impact of active thinking. *Brain Res. Bull.* **70**, 263–269 (2006).
68. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

69. Jung, R. E. & Haier, R. J. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* **30**, 135–154 (2007).
70. Conway, A. R., Kane, M. J. & Engle, R. W. Working memory capacity and its relation to general intelligence. *Trends Cogn. Sci.* **7**, 547–552 (2003).
71. Kyllonen, P. C. & Christal, R. E. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* **14**, 389–433 (1990).
72. Sripada, C., Angstadt, M., Rutherford, S. & Taxali, A. Brain Network Mechanisms of General Intelligence. *bioRxiv* 657205 (2019) doi:10.1101/657205.
73. Neubauer, A. C. & Fink, A. Intelligence and neural efficiency. *Neurosci. Biobehav. Rev.* **33**, 1004–1023 (2009).
74. Shaw, P. *et al.* Intellectual ability and cortical development in children and adolescents. *Nature* **440**, 676 (2006).
75. Colom, R. *et al.* Gray matter correlates of fluid, crystallized, and spatial intelligence: Testing the P-FIT model. *Intelligence* **37**, 124–135 (2009).
76. Turken, U. *et al.* Cognitive processing speed and the structure of white matter pathways: convergent evidence from normal variation and lesion studies. *Neuroimage* **42**, 1032–1044 (2008).
77. Luders, E., Narr, K. L., Thompson, P. M. & Toga, A. W. Neuroanatomical correlates of intelligence. *Intelligence* **37**, 156–163 (2009).

78. Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M. & Voracek, M. Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neurosci. Biobehav. Rev.* **57**, 411–432 (2015).
79. McDaniel, M. A. Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence* **33**, 337–346 (2005).
80. Braver, T. S. *et al.* A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* **5**, 49–62 (1997).
81. Pochon, J. B. *et al.* The neural system that bridges reward and cognition in humans: An fMRI study. *Proc. Natl. Acad. Sci.* **99**, 5669 (2002).

Figure Legends

Figure 1: Main Steps of Brain Basis Set (BBS) Modeling. *BBS is a multivariate predictive modeling method. It utilizes dimensionality reduction with principal components analysis (PCA) to construct a basis set for predicting phenotypes of interest.*

Figure 2: Bifactor Model of General Cognitive Ability. *We performed bifactor exploratory factor analysis on ten behavioral tasks in the Human Connectome Project (HCP) dataset. The resulting model consisted of a general factor (“GCA”) and four group factors and exhibited excellent fit with the data. C=Crystallized Cognitive Ability, S=Processing Speed, M=Memory, V=Visuospatial Ability.*

Figure 3: Visualization of the Three Components From the 2-Back vs. 0-Back Task Contrast Most Predictive of General Cognitive Ability (GCA). *We found the 2-back vs. 0-back contrast was highly effective for GCA prediction, achieving a 0.50 correlation with GCA scores in 10-fold cross-validation. From a 75-component brain basis set model trained to predict GCA scores, the three most statistically significant components are shown above.*

Figure 4: Consensus Predictive Maps for Five Task Contrasts Highly Predictive of General Cognitive Ability (GCA). *We found 13 out of 15 task contrast maps yielded highly statistically significant predictions of GCA in 10-fold cross-validation analysis. For the five most predictive task contrasts, we constructed consensus predictive maps that display brain activation patterns that were most predictive of GCA. Rel=Relational*

Figure 5: Prediction of General Cognitive Ability (GCA) Across 15 Task Contrasts. *We used the brain basis set (BBS) predictive modeling approach to predict GCA from each of 15 Human Connectome Project (HCP) task contrasts. The y-axis in the figure refers to accuracy of these BBS models in predicting GCA, as measured by the correlation between observed and predicted GCA scores in 10-fold cross-validation. For comparison, we additionally plot accuracy of GCA prediction using BBS methods applied to another modality: resting state connectomes. Error bars represent the 95% confidence interval; Blue = permutation-based $p < 0.0001$, observed correlation was higher than all 10,000 in the permutation distribution; Orange = permutation-based p -value < 0.05 ; Red = permutation-based p -value is not significant; TOM=Theory of Mind; Rel=Relational.*

Figure 6: FPN and DMN Activation Patterns and Effectiveness of Task Contrasts in Predicting General Cognitive Ability (GCA). *We hypothesized that placing the brain in an activated, cognitively demanding state improves prediction of GCA. We thus calculated FPN and DMN activation levels, which are thought to index cognitive demandingness, for each of the 15 task contrasts. We in addition calculated each of the 15 task contrast's accuracy in predicting GCA, as measured by the correlation between observed and predicted GCA scores in 10-fold cross-validation. In multiple regression analysis, we found that FPN/DMN activation levels for the 15 contrasts (x-axis) were indeed strongly related to the contrasts' accuracy in predicting GCA (y-axis). That is, contrasts that activated FPN/deactivated DMN more afforded higher accuracy in predicting GCA. Red dashed lines represent the 95% confidence interval.*

We investigated prediction of general cognitive ability (GCA) based on fMRI task activation patterns with 15 task contrasts in the Human Connectome Project dataset. The 2 back versus 0 back contrast achieved a 0.50 correlation with GCA scores in ten-fold cross-validation analysis. Additionally, we found that task contrasts that produce greater fronto-parietal activation and default mode network deactivation—a brain activation pattern associated with executive processing and higher cognitive demand—are more effective in GCA prediction.

<i>N-back task</i>	Participants respond when the picture shown on the screen is the same as the one two trials back (=2- <i>back</i> condition) or the same as one shown at the start of the block (=0- <i>back</i> condition).
<i>Incentive Processing</i>	Participants guess whether the number on a mystery card will be more or less than 5 and win or lose money (<i>reward</i> condition = mostly wins; <i>loss</i> condition = mostly losses)
<i>Motor</i>	Participants move fingers, toes, and tongue
<i>Language Task</i>	Participants answer questions about Aesop's fables (=story condition) or math problems (=math condition).
<i>Social Cognition Task</i>	Participants watch video clips of objects interacting in an agentive way (=theory of mind condition) or random way (=random condition).
<i>Relational Task</i>	Participants identify the dimension along which a cue pair of objects differs and determine if a target pair differs along same dimension (=relational condition). Or they determine if a cue object matches a member of a target pair along a given dimension (=match condition).
<i>Emotion Task</i>	Participants decide whether one of two presented faces match one at the top of the screen (=face condition) or else they perform the same task with shapes (=shape condition)

Table 1: Seven Human Connectome Project (HCP) fMRI Tasks.

Task	R_{cv}^2	Mean Squared Error (MSE)
2back-0back	0.280	0.576
2back	0.265	0.604
Math-Story	0.227	0.623
Rel-Match	0.195	0.639
Random	0.172	0.658
0back	0.172	0.659
TOM	0.172	0.654
Reward	0.165	0.662
Rel	0.156	0.664
Match	0.155	0.669
Punish	0.132	0.686
TOM-Random	0.130	0.692
Faces-Shapes	0.127	0.691
Motor	0.049	0.751
Punish-Reward	0.033	0.763
Resting Connectome	0.078	0.755

Table 2: Prediction of General Intelligence Across 15 Task Contrasts. *We used the brain basis set (BBS) predictive modeling approach to predict general intelligence from each of 15 HCP task contrasts. For comparison, we additionally include results from predicting general intelligence with resting state connectomes. The table shows model performance assessed with cross-validated r squared (R_{cv}^2) and mean square error (MSE). TOM=Theory of Mind; Rel=Relational.*