A distributed brain network predicts general intelligence from resting-state human neuroimaging data

Julien Dubois[1,4], Paola Galdi[5], Lynn K. Paul[1,3] , Ralph Adolphs[1,2,3]

[1] Division of Humanities and Social Sciences, California Institute of Technology, Pasadena CA 91125, USA
[2] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena CA 91125, USA
[3] Chen Neuroscience Institute, California Institute of Technology, Pasadena CA 91125, USA
[4] Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[5] Department of Management and Innovation Systems, University of Salerno, Fisciano, Salerno, Italy

Correspondence:
Julien Dubois
Department of Neurosurgery
Cedars-Sinai Medical Center, Los Angeles, CA 90048
jcrdubois@gmail.com

Abstract

Individual people differ in their ability to reason, solve problems, think abstractly, plan and learn. A reliable measure of this general ability, also known as intelligence, can be derived from scores across a diverse set of cognitive tasks. There is great interest in understanding the neural underpinnings of individual differences in intelligence, since it is the single best predictor of long-term life success, and since individual differences in a similar broad ability are found across animal species. The most replicated neural correlate of human intelligence to date is total brain volume. However, this coarse morphometric correlate gives no insights into mechanisms; it says little about function. Here we ask whether measurements of the activity of the resting brain (resting-state fMRI) might also carry information about intelligence. We used the final release of the Young Adult Human Connectome Project dataset (N=884 subjects after exclusions), providing a full hour of resting-state fMRI per subject; controlled for gender, age, and brain volume; and derived a reliable estimate of general intelligence from scores on multiple cognitive tasks. Using a cross-validated predictive framework, we predicted 20% of the variance in general intelligence in the sampled population from their resting-state fMRI data. Interestingly, no single anatomical structure or network was responsible or necessary for this prediction, which instead relied on redundant information distributed across the brain.

Introduction

Most psychologists agree that there is, in addition to specific cognitive abilities, a very general mental capability to reason, think abstractly, solve problems, plan, and learn across domains [1]. This ability, intelligence, does not refer to a person's sheer amount of knowledge but rather to their ability to recognize, acquire, organize, update, select and apply this knowledge [2], to reason and make comparisons [3]. There are large and reliable individual differences in intelligence across species: some people are smarter than others, and some rats are smarter than others [4].

What's more, these differences matter. Intelligence is one of the most robust predictors of conventional measures of educational achievement [5], job performance [6], socio-economic success [7], social mobility [8], health [9] and longevity [10,11]; and as life becomes increasingly complex, intelligence may play an ever increased role in life outcome [2]. Despite this overwhelming convergence of evidence for the construct of "intelligence" there is considerable debate about what it really is, how best to measure it, and in particular what predictors and mechanisms for it we could find in the brain.

Measuring intelligence: the structure of cognitive abilities

Although apparent in its consequences in everyday life, this is not how intelligence has been operationalized psychometrically. Instead, it is tied to performance on very specific tests, or factor analyses derived from tests. Intelligence tests are among the most reliable, and valid, of all psychological tests and assessments [1]. Psychologists are so confident in the psychometric properties of intelligence tests that, almost 100 years ago, Edwin G. Boring infamously wrote: "Intelligence is what the tests test." [12].
A comprehensive modern intelligence assessment (such as the Wechsler Adult Intelligence Scale, Fourth Edition or WAIS-IV [13]) comprises tasks that assess several aspects of intelligence: some assess verbal comprehension (e.g. word definition, general knowledge, verbal reasoning), some assess visuo-spatial reasoning (e.g., puzzle construction, matrix reasoning, visual perception), some assess working memory (e.g., digit span, mental arithmetic, mental manipulation), and some some assess mental processing speed (e.g., reaction time for detection). The scores on all of these tasks are tallied and compared to a normative, age-matched sample to calculate a standardized Full Scale Intelligence Quotient (FSIQ) score. One of the most important findings in intelligence research is that performances on all these seemingly disparate tasks -- and many other cognitive tasks -- are positively correlated - individuals who perform above average on, say, visual perception, also tend to perform above average on, say, word definition. Spearman [14] discovered this phenomenon and described it as the *positive manifold*, and since then it has been described in a number of species [15]. To explain this empirical observation, he posited the existence of a general factor of intelligence, the 'g-factor', or simply '*g*', which no single task can perfectly measure, but which can be derived

from performance on several cognitive tasks through factor analysis. The g-factor explains around half of people's intellectual differences [16] and shows good reliability across sets of cognitive tasks [17].

The empirical observation of the positive manifold is well-established [18] (it is sometimes referred to as the most replicated result in all of psychology [19]). The descriptive value of Spearman's *g* is beyond doubt, however its interpretation -- that psychometric *g* reflects a general aspect of brain functioning -- was challenged early on [20] and remains a topic of debate to this day among intelligence researchers [21–23]. The leading alternative theory posits that each cognitive test involves several mental processes, and that the sampled mental processes overlap across tests; in this situation, performance on all tests appears to be positively correlated. The common factor, in this framework, is a consequence of the positive manifold, rather than its cause. The most recent formulation of this view was coined "Process Overlap Theory" [19], and the response it received demonstrates the division in the field [24]. The main issue, in this ongoing debate, is that there is no statistical means of distinguishing between alternate theories on the basis of the psychometric data alone [22]. Thus the interpretation of g as general intelligence may be sufficient, but it is not a necessary explanation of the positive manifold.

Beyond the debate on the nature of the *g*-factor, the structure of intelligence is often described in terms of broad abilities, the exact nature of which is also still debated. Prominent theories provide differing descriptions of those broad abilities. The Cattell–Horn–Carroll (CHC) theory [25] recognizes 9 broad domains of ability: Gf (fluid reasoning), Gc (comprehension/knowledge), Gv (visual-spatial processing), Ga (auditory processing), Gsm (short-term memory), Glr (long-term storage and retrieval), Gs (cognitive processing speed), Gq (quantitative knowledge), and Grw (reading and writing). At the root of the CHC theory is the distinction between a "fluid" intelligence, which reflects the capacity to reason and solve problems without using knowledge; and a "crystallized" intelligence, which reflects the ability to use knowledge and experience [26]. The main contending theory is the verbal–perceptual–image rotation (VPR) theory [27,28], which simply posits three broad abilities and does not distinguish whether prior knowledge is used or not. There is one major finding that supports a distinction between fluid and crystallized intelligence [29]: though the two are highly correlated, they change differentially with age [30,31]. Fluid intelligence starts declining around age 25, whereas crystallized intelligence increases until about age 50 and then shows little decline.

The search for biological substrates of intelligence

Differential psychology -- the psychological discipline that studies individual differences between people -- has three main aims: to describe the trait of interest accurately, to establish its impact in real life, and to understand its aetiology, including its biological basis [32]. Much headway has been made with respect to the first two aims for intelligence, as described above. The third aim, despite much effort, has remained elusive.

Individual differences in intelligence are relatively stable: one of the best predictors of intelligence in old age is - perhaps unsurprisingly - intelligence in childhood [33]. Intelligence has a strong genetic component; heritability estimates from twin studies range between .20 (for

children) and .80 (for adults) [5] (see also [34] for data from adoption studies). Genome-Wide Association Studies (GWAS) yield similar heritability estimates [35], and suggest that intelligence is highly polygenic (no single gene accounts for a large fraction of the variance) [36,37].

While high heritability points to genes as an undeniable distal cause of intelligence differences, the proximal cause for individual differences in cognitive ability is to be found in brain function. Arguably, studying the proximal cause may yield more direct insight about the actual mechanisms of intelligence than genetic studies have so far. Since intelligence is a relatively stable trait, one might reason that its aetiology should be found in stable aspects of brain function, and hence also predicted by aspects of brain structure. Most of the data on the neural basis of intelligence comes from neuroimaging and from lesion studies. There are several in-depth reviews of the neurobiological substrates of intelligence, to which the interested reader is referred for a more complete treatment [32,38–40]. Kievit et al. [41] recently described the current state of the neuroscience of intelligence as “an embarrassment of riches” -- a plethora of neuroimaging-derived properties of the brain, structural and functional, have been linked to intelligence over the years. However, many of these claims may not replicate, due to methodological concerns such as data overfitting in small subject samples. Here we briefly highlight some key results to help set the stage for our study.

It is worth noting, from the outset, that the naive search for a simple neurobiological correlate of intelligence faces a major theoretical hurdle, which is best understood through analogy [41]. Imagine a researcher trying to find the biological basis of the construct of “physical fitness”. If they search for a single physical property, they would likely fall short of their goal. Indeed, physical fitness is a composite of several physical properties (cardiorespiratory endurance, muscular strength, muscular endurance, body composition, flexibility), and cannot be equated with any single one, or even any specific combination of these factors. It is very likely that intelligence is of a similar composite nature, as suggested by genetic data [36]. Furthermore, individuals may score identically on an IQ test by using different cognitive strategies, or different brain structures [32]. This is an important picture to hold in mind as one searches for neural correlates of intelligence.

Structural studies

The best replicated brain correlate of intelligence is a fairly simple one: size matters [42]. People who score higher on IQ tests tend to also have bigger brains, as assessed nowadays using total brain volume derived from structural MRI scans; the correlation coefficient is about r=0.24 according to the meta-analytical estimates [43] (or maybe as much as r=0.40 [44]). Initial correlation estimates using small sample sizes were as high as 0.51 [45], and were revised down multiple times [46]. The volume of gray-matter seems slightly more strongly related to intelligence than the volume of white matter [47].

Finer-grained measures, such as cortical thickness or volume of specific regions across the whole brain, have also been shown to correlate positively with intelligence [47], though usually with small sample sizes which may undermine statistical power and replicability.

Recent studies with larger sample sizes have reported associations between intelligence and white matter tract structure (from diffusion imaging [48,49]) as well as white matter integrity (volume of white matter hyperintensities, number of microbleeds, number of iron deposits) [50]; a combination of these with brain volume and cortical thickness has been shown to account for as much as 20% of the variance in g [51]. The role of white matter is further corroborated by several MR spectroscopy studies which find a relationship between N-acetyl aspartate, a metabolite of the oligodendrocytes that constitute the myelin sheath [52], and IQ scores. Neuroimaging studies remain correlative in nature, and causal inference is limited. Lesion studies in humans are an opportunity to establish causation [53], although they come with their own set of limitations [54]. A recent study using a large sample of patients (N=241) used voxel-based lesion mapping and found specific lesion-deficit relationships for the g-factor and several broad ability domains, in frontal and parietal cortices [55].

Functional studies

Extant functional neuroimaging studies (using EEG, PET, rCBF and fMRI) have been summarized as supporting the notion that intelligence is a network property of the brain, related to neural efficiency [38,56]. While this may generally be correct, foundational studies had very low sample sizes [56], and so did a recent study making a similar claim using graph analysis in N=19 subjects [57]. More recent, better powered studies have found correlations between intelligence and the global connectivity of a small region in lateral prefrontal cortex (N=78) [58]; the nodal efficiency of hub regions in the salience network (N=54) [59]; the modularity of frontal and parietal networks (N=309) [60]; and several other seemingly disparate reports.

A neurobiological model of intelligence

A 2007 review of existing structural and functional studies lead to the Parieto-Frontal Integration Theory (P-FIT) of intelligence [38], regarded as “the best available answer to the question of where in the brain intelligence resides” [32]. There aren’t many detractors of the P-FIT model, yet it is a very generic model, and falls short of a mechanistic understanding. Because the brain regions discussed in the model include most of the cortex, it seems difficult to falsify the theory experimentally -- leaving most studies little choice but to “generally support” the P-FIT model.

The current study

Current literature on the fMRI-derived substrates of intelligence suffers from the same caveats as most fMRI-based individual differences research to date [61]: small sample sizes, and lack of out-of-sample generalization. A predictive framework was first used in a recent study [62], which found that fluid intelligence could be predicted from functional connectivity matrices in an early release (N=118) of the Human Connectome Project (HCP) dataset [63], with a correlation r_{LOSO}=0.50 between observed and predicted scores (the LOSO subscript denotes the use of a leave-one-subject-out cross-validation framework). Since this early report, the authors revised the effect size down to r_{LOSO}=0.22 using later data releases (N=606) [64]. Recent work in our

group, in which we further control for confounding effects of age, gender, brain size and motion, as well as use a leave-one-family-out cross-validation framework (LOFO) instead of the original leave-one-subject-out framework (thus accounting for the family structure of the HCP dataset), further revised the effect size down to about r_{LOFO}=0.09 using methods matched as closely as possible to the original study [62]; yet, using improvements including better inter-subject alignment and multivariate modeling, we found r_{LOFO}=0.263 (N=884) [65]. Note that this effect size is comparable to recent estimates of the relationship between brain size and intelligence [43]. Though the explained variance is small, it would in fact fall around the 65th percentile of correlations observed in individual differences research [66] (with the caveat that r_{LOFO} correlation is derived from a cross-validation procedure, which breaks the assumption of independence between individual data points).

According to recent guidelines [44], the assessment of general intelligence with HCP's 24-item Progressive Matrices would be considered of "fair" quality (1 test, 1 cognitive dimension, testing time less than 19 minutes), and would be expected to correlate with general intelligence in the range 0.50 to 0.71. This rather low measurement quality of intelligence itself is likely to attenuate the magnitude of the relationship between neural data and general intelligence. Fortunately, there are several other measures of cognition in the HCP, which we here decided to leverage to derive a better estimate of general intelligence -- one that would meet criteria for an "excellent" quality measurement (more than 9 tests, more than 3 dimensions, testing time more than 40 minutes), and thus be expected to correlate with the general factor of intelligence above 0.95.

Our main aims in this study were to: i) predict an excellent estimate of general intelligence from resting-state functional connectivity in a large sample of subjects from the HCP; ii) depending on the success of i), gain some anatomical insight on which functional connections matter for these predictions. The current study paves the way for a reliable neuroimaging-based science of intelligence differences (large sample size; predictive framework; valid, reliable psychometric construct).

Methods

Many of the methods, in particular the preprocessing of fMRI data and the predictive analyses, were developed and described in our recent publication [65]. Consequently, some sections here were reproduced verbatim from the previous paper.

Dataset

We used data from a public repository, the 1200 subjects release of the Human Connectome Project (HCP) [63]. The HCP provides MRI data and extensive behavioral assessment from almost 1200 subjects. Acquisition parameters and "minimal" preprocessing of the resting-state fMRI data is described in the original publication [67]. Briefly, each subject underwent two sessions of resting-state fMRI on separate days, each session with two separate 15 minute acquisitions generating 1200 volumes (customized Siemens Skyra 3 Tesla MRI

scanner, TR = 720 ms, TE = 33 ms, flip angle= 52°, voxel size = 2 mm isotropic, 72 slices, matrix = 104 x 90, FOV = 208 mm x 180 mm, multiband acceleration factor = 8). The two runs acquired on the same day differed in the phase encoding direction, left-right and right-left (which leads to differential signal intensity especially in ventral temporal and frontal structures). The HCP data was downloaded in its minimally preprocessed form, i.e. after motion correction, B_0 distortion correction, coregistration to T_1-weighted images and normalization to MNI space.

Cognitive ability tasks

Previous studies of the neural correlates of intelligence in the HCP [62,64] have relied on the number of correct responses on form A of the 24(+3 bonus)-item Penn Matrix Reasoning Test (PMAT), a test of non-verbal reasoning ability which can be administered in under 10 minutes (mean=4.6, std= 3 min; [68]), and is included in the University of Pennsylvania Computerized Neurocognitive Battery (Penn CNB, [69–71]). The PMAT is designed to parallel many of the psychometric properties of the Raven's Standard Progressive Matrices test (RSPM, originally published in 1938), which comprises 60 items [72]. A computerized version of RSPM takes approximately 17 minutes to administer [73], which is too long for large-scale studies requiring an efficient assessment of all major neurocognitive domains. Bilker and colleagues developed an abbreviated 9-item version of the RSPM test which achieves similar item- and test-level characteristics to 60-item based scores, with an administration time of about 4 minutes [74]. This paved the way for the development of the computerized, adaptive PMAT test, whose items were designed from scratch to have similar psychometric characteristics as RSPM items, while limiting learning effects and expanding the representation of the abstract reasoning construct (Ruben Gur, personal communication).
Assessment of cognitive ability in the HCP [75] also includes several tasks from the Blueprint-funded NIH Toolbox for Assessment of Neurological and Behavioral function (http://www.nihtoolbox.org), as well as other tasks from the Penn computerized neurocognitive battery [70]. These other tasks can be leveraged to derive a better measure of the general intelligence factor [44]. We included all cognitive tasks listed in the HCP Data Dictionary, except for: i) the delay discounting task, which is not a measure of ability (i.e. there is not a correct response); and ii) the Short Penn Continuous Performance Test which is about sustained attention rather than cognitive ability, and whose distribution departed too much from normality (data not shown) . Our initial selection thus consisted of 10 tasks, which are listed in **Table 1** along with a brief description (the descriptions are copied almost word for word from the HCP Data Dictionary). When several outcome measures were available for a given task, we selected the one that best captured ability; when both age-adjusted and unadjusted scores were available, we included the unadjusted scores. Though some of the NIH toolbox scores combine accuracy and reaction time, we only considered accuracies for the Penn CNB tasks (to avoid confounding ability and processing speed; but, see [71]).

Test Name	Variable(s)	Test Activity / Outcome
NIH Toolbox Dimensional Change Card Sort Test	**CardSort_Unadj** [Unadjusted Scale Score]	Two target pictures are presented that vary along two dimensions (e.g., shape and color). Participants are asked to match a series of bivalent test pictures (e.g., yellow balls and blue trucks) to the target pictures, first according to one dimension (e.g., color) and then, after a number of trials, according to the other dimension (e.g., shape). "Switch" trials are also employed, in which the participant must change the dimension being matched. For example, after 4 straight trials matching on shape, the participant may be asked to match on color on the next trial and then go back to shape, thus requiring the cognitive flexibility to quickly choose the correct stimulus. Scoring is based on a combination of accuracy and reaction time, and the test takes approximately 4 minutes to administer.
NIH Toolbox Flanker Inhibitory Control and Attention Test	**Flanker_Unadj** [Unadjusted Scale Score]	The test requires the participant to focus on a given stimulus while inhibiting attention to arrows flanking it. Sometimes the middle stimulus is pointing in the same direction as the "flankers" (congruent) and sometimes in the opposite direction (incongruent). Scoring is based on a combination of accuracy and reaction time, and the test takes approximately 3 minutes to administer.
NIH Toolbox List Sorting Working Memory Test	**ListSort_Unadj** [Unadjusted Scale Score]	This task assesses working memory and requires the participant to sequence different visually- and orally-presented stimuli. Pictures of different foods and animals are displayed with both a sound clip and written text that name the item. The task has two different conditions: 1-List and 2-List. In the 1-List condition, participants are required to order a series of objects (either food or animals) in size order from smallest to largest. In the 2-List condition, participants are presented both food and animals and are asked to report the food in size order, followed by the animals in size order. Participants have two practice items, in which the images briefly "flash" sequentially on the screen in each condition.
NIH Toolbox Picture Sequence Memory Test	**PicSeq_Unadj** [Unadjusted Scale Score]	The Picture Sequence Memory Test is a measure developed for the assessment of episodic memory. It involves recalling increasingly lengthy series of illustrated objects and activities that are presented in a particular order on the computer screen. The participants are asked to recall the sequence of pictures that is demonstrated over two learning trials; sequence length varies from 6-18 pictures, depending on age. Participants are given credit for each adjacent pair of pictures (i.e., if pictures in locations 7 and 8 and placed in that order and adjacent to each other anywhere – such as slots 1 and 2 – one point is awarded) they correctly place, up to the maximum value for the sequence, which is one less than the sequence length (if there are 18 pictures in the sequence, the maximum score is 17, because that is the number of adjacent pairs of pictures that exist). The test takes approximately 7 minutes to administer.
NIH Toolbox Picture Vocabulary Test	**PicVocab_Unadj** [Unadjusted Scale Score]	This measure of receptive vocabulary is administered in a computerized adaptive format. The respondent is presented with an audio recording of a word and four photographic images on the computer screen and is asked to select the picture that most closely matches the meaning of the word. This test takes approximately 4 minutes to administer.
NIH Toolbox Pattern Comparison Processing Speed Test	**ProcSpeed_Unadj** [Unadjusted Scale Score]	This test measures speed of processing by asking participants to discern whether two side-by-side pictures are the same or not. Participants' raw score is the number of items correct in a 90-second period. The items are designed to be simple to most purely measure processing speed. The test overall takes approximately 3 minutes to administer.
NIH Toolbox Oral Reading Recognition Test	**ReadEng_Unadj** [Unadjusted Scale Score]	The participant is asked to read and pronounce letters and words as accurately as possible. The test administrator scores them as right or wrong. The test is given via a computerized adaptive format and requires approximately 3 minutes.
Penn CNB Penn Progressive Matrices	**PMAT24_A_CR** [# of Correct Resp]	Fluid intelligence is measured using Raven's Progressive Matrices . We use Form A of an abbreviated version of the Raven's developed by Gur and colleagues. Participants are presented with patterns made up of 2x2, 3x3 or 1x5 arrangements of squares, with one of the squares missing. The participant must pick one of five response choices that best fits the missing square on the pattern. The task has 24 items and 3 bonus items, arranged in order of increasing difficulty. However, the task discontinues if the participant makes 5 incorrect responses in a row.
Penn CNB Penn Word Memory Test	**IWRD_TOT** [Total # of Correct Resp]	Verbal episodic memory is measured using Form A of the Penn Word Memory Test. Participants are shown 20 words and asked to remember them for a subsequent memory test. They are then shown 40 words (the 20 previously presented words and 20 new words matched on memory related characteristics). They decide whether they have seen the word previously by choosing among "definitely yes," "probably yes," "probably no," and "definitely no."
Penn CNB Variable Short Penn Line Orientation	**VSPLOT_TC** [total # correct]	Spatial orientation processing is measured using the Variable Short Penn Line Orientation Test. Participants are shown two lines with different orientations. They have to rotate one of the lines (a moveable blue one) so that is parallel to the other line (a fixed red line). The rotation of the blue line is accomplished by clicking buttons on the keyboard that rotate the lines either clockwise or counterclockwise. Across trials, the lines vary in their relative location on the screen, though the distance between the centers of the two lines is always the same. The length of the red line is always the same, but the length of the blue line can be either short or long. There are a total of 24 trials

Table 1. List of cognitive measures included in our analyses.

Subject selection

The total number of subjects in the 1200-subject release of the HCP dataset is N=1206. We applied the following criteria to include/exclude subjects from our analyses (listing in parentheses the HCP database field codes). i) Complete neuropsychological datasets. Subjects must have completed all relevant neuropsychological testing (PMAT_Compl=True, NEO-FFI_Compl=True, Non-TB_Compl=True, VisProc_Compl=True, SCPT_Compl=True, IWRD_Compl=True, VSPLOT_Compl=True) and the Mini Mental Status Exam (MMSE_Compl=True). Any subjects with missing values in any of the tests or test items were discarded. This left us with N = 1183 subjects. ii) Cognitive compromise. We excluded subjects with a score of 26 or below on the MMSE, which could indicate marked cognitive impairment in this highly educated sample of adults under age 40 [76]. This left us with N = 1181 subjects (638 females, 28.8 +/- 3.7 y.o., range 22-37 y.o). This is the sample of subjects available for factor analyses.

Furthermore, iii) subjects must have completed all resting-state fMRI scans (3T_RS-fMRI_PctCompl=100), which leaves us with N = 988 subjects. Finally, iv) we further

excluded subjects with a root-mean-squared frame-to-frame head motion estimate (Movement_Relative_RMS.txt) exceeding 0.15mm in any of the 4 resting-state runs (threshold similar to [62]). This left us with the final sample of N = 884 subjects (475 females, 28.6 +/- 3.7 y.o., range 22-36 y.o.) for predictive analyses based on resting-state data.

Deriving the general factor of intelligence, *g*

There are several methods in the literature to derive a general factor of intelligence from scores on a set of cognitive tasks. The simplest consists in using a standardized sum score composite -- simply summing the standardized scores from all selected cognitive tasks. This is the conventional approach when all scores come from a well-validated battery. However, since we are here including scores from two different cognitive batteries (NIH toolbox and Penn CNB), we sought to characterize the structure of cognitive abilities in our sample using factor analysis, and then derive the scores for the general factor. We conducted an exploratory factor analysis (EFA), specifying the bi-factor model of intelligence -- a common factor g which loads on all test scores, and several group factors that each load on subsets of the test scores; all latent factors are orthogonal to one another -- using the *psych* (v1.7.8) package [77] in *R* (v3.4.2). We specifically used the *omega* function, which conducts a factor analysis (with maximum likelihood estimation) of the data set, rotates the factors obliquely (using “oblimin” rotation), factors the resulting correlation matrix, then does a Schmid-Leiman transformation [78] to find general factor loadings. Model fit is assessed using several commonly used statistics in factor analysis [79]: the Comparative Fit Index (CFI; should be as close to 1 as possible; values >0.95 are considered a good fit); the Root Mean Squared Error of Approximation (RMSEA; should be as close to 0 as possible; values <0.06 are considered a good fit); the Standardized Root Mean Squared Residual (SRMR; should be as close to 0 as possible; values <0.08 are considered a good fit); and the Bayesian Information Criterion (BIC; better models have lower values, can be negative). Factor scores can be derived using different methods [80], for example the regression method. This approach mimics the one taken by [55]. It is however usually preferable to derive scores from a confirmatory factor analysis (CFA). The main difference between EFA and CFA is that in EFA, observed task scores are allowed to cross-load freely on several group factors, while in CFA such cross-loadings can be forbidden. For the purpose of deriving the general factor of intelligence, there is little difference between using CFA and EFA in practice; we conducted a CFA using the *lavaan* (v0.5-23.1097) package [81] in *R* to verify this (see **Supplementary Material**).

Assessment and removal of potential confounds

We computed the correlation of the general factor of intelligence *g* with Gender (HCP variable: *Gender*), Handedness and Age (restricted HCP variables: *Handedness*, *Age_in_Yrs*). We also looked for differences in *g* in our subject sample with variables that are likely to affect FC matrices, such as brain size (we used *FS_BrainSeg_Vol*), motion (we computed the sum of framewise displacement in each run), and the multiband reconstruction algorithm which

changed in the third quarter of HCP data collection (*fMRI_3T_ReconVrs*). We then used multiple linear regression to regress these variables from g scores and remove their confounding effects.

Note that we do not control for differences in cortical thickness and other morphometric features, which have been reported to be correlated with intelligence (e.g. [82]). These likely interact with FC measures and should eventually be accounted for in a full model, yet this was deemed outside the scope of the present study.

Data preprocessing

Resting-state data must be preprocessed beyond “minimal preprocessing”, due to the presence of multiple noise components, such as subject motion and physiological fluctuations. Several approaches have been proposed to remove these noise components and clean the data, however the community has not yet reached a consensus on the “best” denoising pipeline for resting-state fMRI data [83–86]. Most of the steps taken to denoise resting-state data have limitations, and it is unlikely that there is a set of denoising steps that can completely remove noise without also discarding some of the signal of interest. Categories of denoising operations that have been proposed comprise tissue regression, motion regression, noise component regression, temporal filtering and volume censoring. Each of these categories may be implemented in several ways. There exist several excellent reviews of the pros and cons of various denoising steps [84,87–89].

We recently explored the effects of several preprocessing pipeline on the prediction of personality factors and PMAT scores in the HCP dataset [65]. Here we adopt a preprocessing pipeline which was found to perform well in that study. The pipeline reproduces as closely as possible the strategy described in [62] and consists of seven consecutive steps: 1) the signal at each voxel is z-score normalized; 2) using tissue masks, temporal drifts from cerebrospinal fluid (CSF) and white matter (WM) are removed with third-degree Legendre polynomial regressors; 3) the mean signals of CSF and WM are computed and regressed from gray matter voxels; 4) translational and rotational realignment parameters and their temporal derivatives are used as explanatory variables in motion regression; 5) signals are low-pass filtered with a Gaussian kernel with a standard deviation of 1 TR, i.e. 720ms in the HCP dataset; 6) the temporal drift from gray matter signal is removed using a third-degree Legendre polynomial regressor; 7) global signal regression is performed. These operations were performed using an in-house, Python (v2.7.14)-based pipeline (mostly based on open source libraries and frameworks for scientific computing, including SciPy (v0.19.0), Numpy (v1.11.3), NiLearn (v0.2.6), NiBabel (v2.1.0), Scikit-learn (v0.18.1) [90–94]).

Inter-subject alignment, parcellation, and functional connectivity matrix generation

We use surface-based multi-modally aligned cortical data (MSM-All [95]), together with a parcellation that was derived from this data using an objective semi-automated neuroanatomical

approach [96]. The parcellation has 360 nodes, 180 for each hemisphere. These nodes can be attributed to the major resting state networks [97] (**Figure 4a**).

Time series extraction simply consisted in averaging data from vertices within each parcel, and matrix generation in pairwise correlating parcel time series (Pearson correlation coefficient). We concatenated time series across runs to derive average FC matrices (REST1: from concatenated REST1_LR and REST1_RL time series; REST2: from concatenated REST2_LR and REST2_RL time series; REST12: from concatenated REST1_LR, REST1_RL, REST2_LR and REST2_RL time series).

There are (360 * 359)/2 = 64620 undirected edges in a network of 360 nodes. This is the dimensionality of the feature space for prediction.

Prediction models

We use a univariate feature filtering approach to reduce the number of features, discarding edges for which the p-value of the correlation with the behavioral score is greater than a set threshold, for example $p<0.01$ (as in [62]).We then use Elastic Net regression to learn the relationship with behavior. Elastic Net is a regularized regression method that linearly combines L1- (lasso) and L2- (ridge) penalties to shrink some of the regressor coefficients toward zero, thus retaining just a subset of features. The lasso model performs continuous shrinkage and automatic variable selection simultaneously, but in the presence of a group of highly correlated features, it tends to arbitrarily select one feature from the group. With high-dimensional data and few examples, the ridge model has been shown to outperform lasso; yet it cannot produce a sparse model since all the predictors are retained. Combining the two approaches, elastic net is able to do variable selection and coefficient shrinkage while retaining groups of correlated variables. Here however, based on preliminary experiments and on the fact that it is unlikely that just a few edges contribute to prediction, we fixed the L1 ratio (which weights the L1- and L2- regularizations) to 0.01, which amounts to almost pure ridge regression. We used 3-fold nested cross-validation (with balanced “classes”, based on a partitioning of the training data into quartiles) to choose the alpha parameter (among 50 possible values) that weighs the penalty term.

Cross-validation scheme

In the HCP dataset, several subjects are genetically related (in our final subject sample there were 410 unique families). To avoid biasing the results due to this family structure (e.g., perhaps having a sibling in the training set would facilitate prediction for a test subject, if both intelligence and functional connectivity are heritable), we implemented a leave-one-family-out cross-validation scheme for all predictive analyses.

Statistical assessment of predictions

Several measures can be used to assess the quality of prediction. A typical approach is to plot observed vs. predicted values (rather than predicted vs. observed [98]). The Pearson

correlation coefficient between observed scores and predicted scores is often reported as a measure of prediction (e.g. [62]), given its clear graphical interpretation. However, in the context of cross-validation, it is incorrect to square this correlation coefficient to obtain the coefficient of determination R^2, which is often taken to reflect the proportion of variance explained by the model [99]; instead, the coefficient of determination R^2 should be calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{1}$$

where *n* is the number of observations (subjects), y is the observed response variable, $\bar{y}$ is its mean, and $\hat{y}$ is the corresponding predicted value. Equation (1) therefore measures the size of the residuals from the model compared with the size of the residuals for a null model where all of the predictions are the same, i.e., the mean value $\bar{y}$. In a cross-validated prediction context, R^2 can actually take negative values (in cases when the denominator is larger than the numerator, i.e. when the sum of squared errors is larger than that of the null model)! Yet another, related statistic to evaluate prediction outcome is the Root Mean Square Deviation (RMSD), defined in [98] as:

$$RMSD = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{2}$$

RMSD as defined in (2) represents the standard deviation of the residuals. To facilitate interpretation, it can be normalized by dividing it by the standard deviation of the observed values:

$$nRMSD = \frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y}_i)^2}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}} = \sqrt{1 - R^2} \tag{3}$$

nRMSD thus has a very direct link to R^2 (3); it is interpretable as the average deviation of each predicted value to the corresponding observed value, and is expressed as a fraction of the standard deviation of the observed values.

In a cross-validation scheme, the folds are not independent of each other. This means that statistical assessment of the cross-validated performance using parametric statistical tests is problematic [100,101]. Proper statistical assessment should thus be done using permutation testing on the actual data. To establish the empirical distribution of chance, we ran our predictive analysis using 1000 random permutations of the scores (shuffling scores randomly between subjects, keeping everything else exactly the same, including the family structure).

Results

A general factor accounts for 58% of the covariance structure of cognitive tasks in the HCP sample

All selected cognitive task scores (**Table 1**) correlate positively with one another, as expected from the well-known positive manifold (**Figure 1a)**. A parallel analysis suggests an underlying 4-factor structure **(Figure 1b)**. An exploratory bifactor analysis with a general factor *g* and 4 group factors fits the data very well (CFI=0.990; RMSEA=0.0311; SRMR=0.0201; BIC=-0.519), much better than a single factor model (CFI=0.719; RMSEA=0.1398; SRMR=0.0887; BIC=591.172). The solution is depicted in **Figure 1c**. The four factors can naturally be interpreted as: 1) Crystallized Ability [cry] (PicVocab_Unadj + ReadEng_Unadj); 2) Processing Speed [spd] (CardSort_Unadj + Flanker_Unadj + ProcSpeed_Unadj); 3) Visuospatial Ability [vis] (PMAT24_A_CR + VSPLOT_TC); and 4) Memory [mem] (IWRD_TOT + PicSeq_Unadj + ListSort_Unadj).

Across all cognitive task scores, the general factor accounts for 58.5% of the variance (coefficient omega_hierarchical ω_h [102–104]), while group factors account for 18.2% of the variance (with 15.5% of the variance unaccounted for). Another important metric is coefficient omega_subscale ω_s [105] which quantifies the reliable variance across the tasks accounted for by each subscale, beyond that accounted for by the general factor; we find ω_s^{cry}=38.7%, ω_s^{spd}=57.4%, ω_s^{vis}=9.9% and ω_s^{mem}=27.8%. While some of these subscale factors account for a substantive proportion of the variance across their respective tasks, their measurement quality is at most fair [44] due to the limited number of constituent tasks; thus we choose to focus on the general factor *g* only.

Factor scores are indeterminate, and several alternate methods exist to derive them from a structural model [80]. To avoid this issue, most researchers prefer to remain in latent space for further analyses [41]. However, for subsequent analyses we require factor scores for g, which we derive using regression-based weights (“Thurstone” method).

We compare the general factor score derived from this exploratory factor analysis (EFA) with a simple composite score consisting of the sum of standardized observed test scores. As expected, we find that the simple composite score correlates highly with the EFA-derived g (r=0.91).

Though deriving factor scores from an EFA is often done by empirical researchers [106], it is theoretically preferable to use a confirmatory factor analysis (CFA) framework. A major difference is that a CFA model typically restricts cross-loadings (an observed variable loading on several latent factors), while the EFA allows them; this can reduce the size of the *g* factor in the EFA. We conduct a CFA (see **Supplementary Material**), and find that the *g* factor extracted from the CFA model correlates almost perfectly with the *g* factor extracted from the EFA model (r=.99). Therefore, we simply proceed with the EFA-derived *g*-factor in the following (in keeping with previous literature [55]).

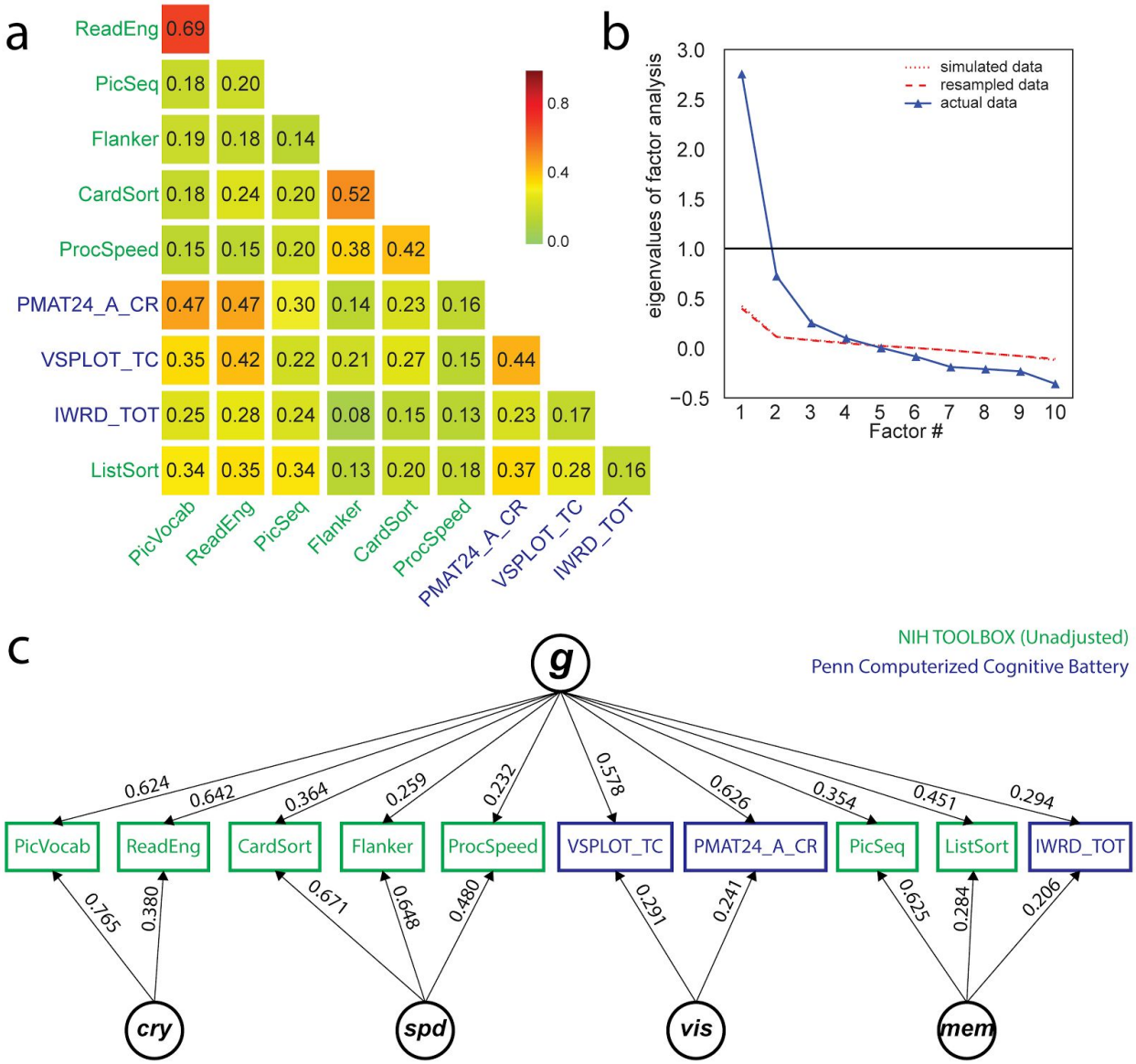


Figure 1. Exploratory factor analysis of select cognitive tasks (Table 1) in the Human Connectome Project (HCP) dataset, using N=1181 subjects. a) All cognitive task scores correlate positively with one another, reflecting the well-established positive manifold (see also **Supplementary Figure 2** for scatter plots). **b)** A parallel analysis suggests the presence of 4 latent factors from the covariance structure of cognitive task scores. Note that the simulated and resampled data lines are nearly indistinguishable. **c)** A bifactor analysis fits the data well (see fit statistics in text), and yields a theoretically plausible solution with a general factor (***g***) and four group factors which can be interpreted as crystallized ability (***cry***), processing speed (***spd***), visuospatial ability (***vis***) and memory (***mem***). Loadings less than 0.2 are not displayed.

Brain size, gender, and motion are correlated with g

There are known effects of gender [107,108], age [109,110], in-scanner motion [111–113] and brain size [114] on the functional connectivity patterns measured in the resting-state with fMRI. It is thus necessary to control for these variables [115]: indeed, if intelligence is correlated with gender, one would be able to predict some of the variance in intelligence solely from functional connections that are related to gender. The easiest way to control for these confounds is to remove any relationship between the confounding variables and the score of interest in our sample of subjects, which can be done using multiple regression. Note that this approach may be too conservative, and that more work remains to be done on dealing with such confounds (see **Discussion**).

We characterize the relationship between intelligence and each of the confounding variables listed above in our subject sample (**Figure 2**). Intelligence is correlated with gender (men scored higher in our sample), age (younger scored higher in our sample -- note the limited age range 22-36 y.o.), and brain size (larger brains score higher) [43,44]. There is no relationship between handedness and intelligence in our sample ($r=2x10^{-6}$). Motion, quantified as the sum of frame-to-frame displacement over the course of a run (and averaged separately for REST1 and REST2) is correlated with intelligence: subjects scoring lower on intelligence moved more during the resting-state. Note that motion in REST1 is highly correlated ($r=0.72$) with motion in REST2, indicating that motion itself may be a stable trait, and correlated with other traits. While the interpretation of these complex relationships would require further work outside the scope of this study (using partial correlations, and mediation models, to disentangle effects), it is critical to remove shared variance between intelligence and the primary confounding variables before proceeding further. This ensures that our model is trained specifically to predict intelligence, rather than confounds that covary with it in our subject sample [115]. However there are several other variables that we do not explicitly account for here, for example the Openness personality trait which we previously found to be correlated with intelligence [65].

Another possible confound, specific to the HCP dataset, is a difference in the image reconstruction algorithm between subjects collected prior to and after April 2013. The reconstruction version leaves a notable signature on the data that can make a large difference in the final analyses produced [116]. We find a small, but significant correlation with the intelligence in our sample (indicating that subjects imaged with the old reconstruction version are, on average, less intelligent than the ones imaged with the newer reconstruction version). This confound is, of course, a simple sampling bias artifact with no meaning. Yet, this significant artifactual correlation must be removed, by including the reconstruction factor as a confound variable.

Importantly, the multiple linear regression used for removing the variance shared with confounds is fitted on the training data (in each cross-validation fold during the prediction analysis), and then the fitted weights are applied to remove the effects of confounds in both the training and test data. This is critical to avoid any leakage of information, however negligible, from the test data into the training data.

Authors of the HCP-MegaTrawl have used transformed variables (Age^2) and interaction terms (Gender x Age, Gender x Age^2) as further confounds [117]. After accounting for the confounds described above, we do not find sizeable correlations with these additional terms (all correlations <0.011), and thus we do not use these additional terms in our confound regression.

Figure 2. Correlation of the general intelligence factor scores with PMAT24_A_CR scores, and with potential confounds, in the sample of subjects used for prediction analyses (N=884). All of these variables except for PMAT24_A_CR scores were regressed out of the training set data to obtain an unconfounded measure of g.

Resting-state FC predicts 20% of the variance in *g* across subjects

We compute a resting-state functional connectivity matrix for each subject from 1 hour of resting-state data (REST12), yielding a very reliable estimate of the stable functional network of each individual -- with 60 minutes of scan time, the test-retest reliability of the FC matrix is above r=.96 [118].

We use a leave-one-family-out cross-validation scheme to train a regularized linear model and predict general intelligence from functional connectivity matrices (features are the 64620 undirected edges), in our sample of 884 subjects. We find a significant correlation between observed and predicted *g* scores (r=0.457, p_{1000}<0.001, based on 1000 permutations), a coefficient of determination that differs significantly from chance (R^2=0.206; p_{1000}<0.001), and a normalized root mean square deviation that is significantly lower than its null distribution (nRMSD=0.892;p_{1000}<0.001) (**Figure 3**).

For comparison, we find that the prediction of PMAT24_A_CR captures less variance (r=0.263,p_{1000}<0.001; R^2=0.047, p_{1000}<0.001; nRMSD=0.977, p_{1000}<0.001). It is likely that the moderating effect of measurement quality [44] limits prediction performance.

Similarly, using only 30min of resting-state data to derive the functional connectivity matrices has a moderating effect on prediction performance. With 30 minutes of scan time, test-retest reliability of FC matrices falls to about r=.92 [118]. Predicting *g* using REST1, we find r=0.419, R^2=0.170, nRMSD=0.912; using REST2, we find r=0.312, R^2=0.067, nRMSD=0.966.

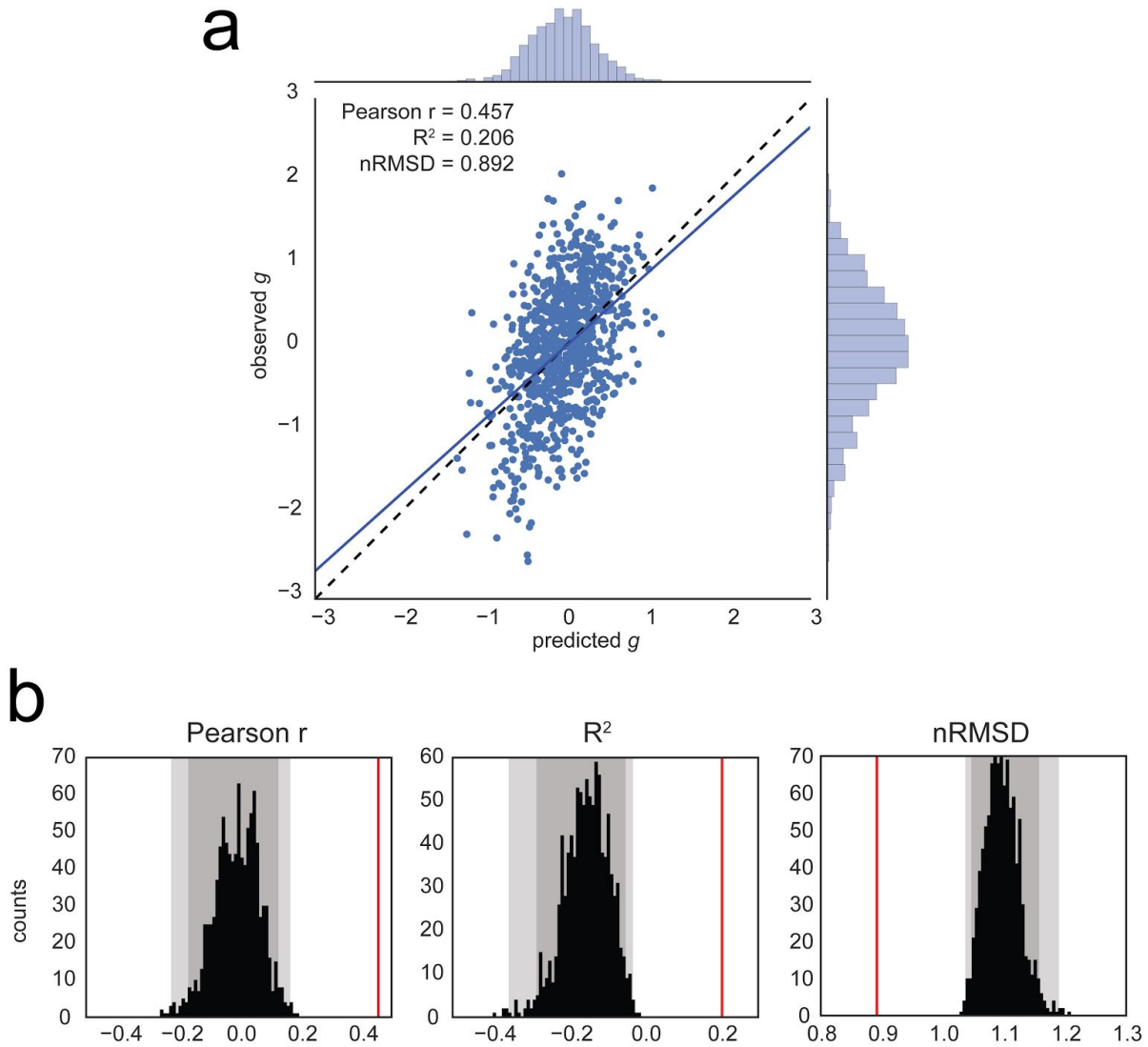


Figure 3. Prediction of the general factor of intelligence *g* from resting-state functional connectivity, averaging all resting-state runs for each subject (REST12, totaling 1h of fMRI data). **a)** Observed vs. predicted values of the general factor of intelligence. The regression line has a slope close to 1, as expected theoretically [98]. The correlation coefficient is r=0.457 (REST1 only, r=0.419; REST2 only, r=0.312). **b)** Evaluation of prediction performance according to several statistics and their distributions under the null hypothesis, as simulated through permutation testing (with 1000 surrogate datasets). All fit statistics fall far out of the confidence intervals under the null hypothesis. *Left:* the correlation between observed and predicted values; *Center:* the coefficient of determination R^2, interpretable as the proportion of explained variance; *Right:* the normalized Root Mean Square Deviation nRMSD, which indicates the average difference between observed and predicted scores. Faint gray shade: $p<0.05$; darker gray shade: $p<0.01$, permutation tests.

Predictive edges are distributed in FPN, CON, DMN, and VIS networks

We have demonstrated that a substantial and statistically highly significant amount of variance (about 20%) in general intelligence *g* across our sample of subjects can be predicted from resting-state functional connectivity. Is there a specific set of edges (connectivity between specific anatomical parcels) in the brain that carry most of the information? The Parieto-Frontal Integration Theory (P-FIT) of intelligence [38] postulates roles for cortical regions in the prefrontal (BA 6, 9–10, 45–47), parietal (BA 7, 39–40), occipital (BA 18–19), and temporal association cortex (BA 21, 37).

To address this question, we use a descriptive network selection/elimination approach. We focus on the 7 major resting-state networks [119]; Ito and colleagues recently assigned each region of the parcellation used here to these functional networks using the Generalized Louvain method for community detection with resting-state fMRI data [97] (**Figure 4a**). This network assignment indeed groups regions that have similar connectivity patterns, at the level of single subjects (**Figure 4b**). First, we ask how well we can predict *g* keeping edges within only one network ($\binom{7}{1} = 7$ combinations, **Figure 4c**), or within/between two networks ($\binom{7}{2} = 21$ combinations, **Figure 4d**). For all these analyses we carried out exactly the same methods as described above (including the univariate feature selection step), but training and testing on a reduced feature space (corresponding to only those features (edges) under investigation for questions about the specific networks). Prediction performance within a single network, or with two networks, is much impaired compared to performance with the full set of edges (one network, maximum performance r=0.327; two networks, maximum performance r=0.373); however, some networks carry more information than others: the most informative networks are CON, DMN, FPN and VIS, while DAN, AUD and SMN carry very little information. These results are in good agreement with the P-FIT; in particular, in addition to the eponymous frontal and parietal regions which have been reported in other studies already [55,62], we evidence information in the VIS network as postulated by P-FIT. We next explore how the virtual lesioning of networks affects prediction: we lesion a single network ($\binom{7}{6} = 7$, **Figure 4f**) or two networks ($\binom{7}{5} = 21$, **Figure 4e**). We find that lesioning one or two networks has a very small effect on the prediction of *g* (lesion one network, minimum performance r=0.409; lesion two networks, minimum performance r=0.373), indicating there is distributed and redundant information about *g* in functional connectivity patterns across several brain networks.

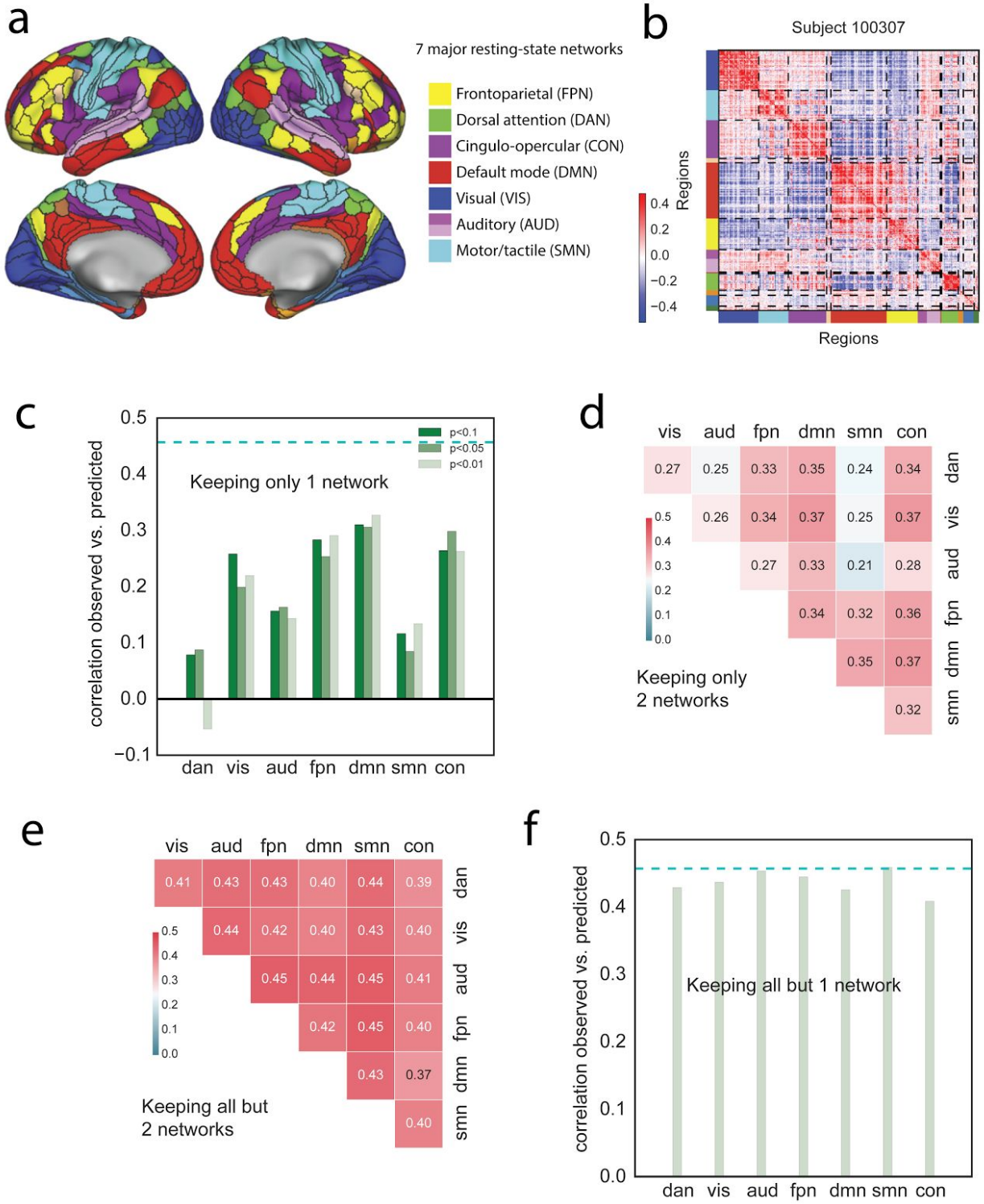


Figure 4. A distributed neural basis for g. a) Assignment of parcels to major resting-state networks, reproduced from Ito et al. [97]. **b)** Example REST12 functional connectivity matrix ordered by network, for an individual subject (id=100307). **c)** Prediction performance for REST12 matrices (as Pearson correlation between observed and predicted scores) using only

within network edges, for the 7 main resting-state networks listed in **a**. DMN, FPN, CON and VIS carry the most information about *g*. The three shades of green correspond to three univariate thresholds for initial feature selection (we used $p<0.01$ as in the main analysis; as well as $p<0.05$ and $p<0.1$ to make sure that the results were not limited by the inclusion of too few edges). The dashed cyan line shows for comparison the prediction performance for the whole-brain connectivity matrix (same data as in Figure 3). **d)** Prediction performance with two networks only (univariate feature selection with $p<0.01$). **e)** Prediction performance for REST12 matrices after "lesioning" two networks (univariate feature selection with $p<0.01$). **f)** Prediction performance after lesioning one network (univariate feature selection with $p<0.01$).

Discussion

Deary et al. recently wrote [120]: "the effort to understand the psychobiology of intelligence has a resemblance with digging the tunnel between England and France: We hope, with workers on both sides having a good sense of direction, that we can meet and marry brain biology and cognitive differences." The present study is a step in this direction, offering to our knowledge the to-date most robust investigation specifically focused on predicting intelligence from resting-state fMRI data. Here we used factor analysis of the scores on 10 cognitive tasks to derive a bi-factor model of intelligence, including a common g-factor and broad ability factors, which is the standard in the field of intelligence research [28,44]. We used reliable estimates of functional connectivity in a large sample of subjects, from one hour of high-quality resting-state fMRI data per subject. We used the best available inter-subject alignment algorithm (MSM-All), a stringent control for confounding variables, and out-of-sample prediction. With these state-of-the-art methods on both ends of the tunnel, we demonstrated a strong relationship between general intelligence *g* and resting-state functional connectivity (at least as strong as the well established relationship of intelligence with brain size [43,44]). We further established that predictive network edges were fairly distributed through the brain, though they mostly fell within 4 of the 7 major resting-state networks: the fronto-parietal network, the default mode network, the control network, and the visual network. These findings are in general agreement with the parieto-frontal integration theory (P-FIT) of intelligence. They considerably extend a previous report [62] which had hinted at a relationship between resting-state functional connectivity and intelligence using a much smaller subject sample (N=117), no account of potential confounding variables, a cross-validation scheme that did not respect family structure, a less functionally accurate inter-subject alignment, and a lower quality measure of intelligence.

We indeed found evidence that measurement quality, both on the behavioral and on the neural side, moderated the effect size of the relationship between brain and behavior [44]. We achieved better prediction performance using *g* rather than the number of correct responses on the PMAT24_A test; and we achieved better performance using REST12 matrices (1h of data per subject) rather than REST1 or REST2 matrices (30min of data per subject). Though this is of course expected statistically -- the noise ceiling gets lower as noise increases for the two variables that are correlated -- this is an important observation for future explorations in other datasets. In many aspects the neural and behavioral data in the HCP are of higher quality than

most other large-scale neuroimaging projects; conducting similar analyses in other datasets may yield smaller effect sizes solely because of lower data quality and, just as importantly, quantity.

The general factor of intelligence *g* that we derived from 10 cognitive task scores is as reliable as it can be in this dataset, and would be unlikely to improve substantially even if additional measures were available. An interesting question for future studies will be to look at the predictability of the subscales: crystallized ability, processing speed, visuospatial reasoning, and memory. Addition of tasks in each of these subdomains would increase the reliability of these specific ability factors, and allow for a more precise exploration of their neural bases. This would require a much longer assessment and many more ability tests. Rather than build an entirely new dataset from scratch, the possibility of testing all HCP subjects again on a lengthier, diverse cognitive ability battery should be considered [121].

Another factor that can moderate the relationship between variables is range restriction of variables [122]. Here, there is some concern that the range of intelligence scores in the HCP subject sample may be restricted to the higher end of the distribution. While published normative data is currently unavailable for the Penn matrix reasoning test and other Penn CNB tests in the age range of our subjects, the NIH toolbox tests provide age-normed scores. Inspection of these scores indicates that the HCP subject sample is indeed biased towards higher scores (in particular for crystallized abilities; see **Supplementary Figure 3**). This sampling bias is a well-known, systemic issue in experimental psychology [123], and one that is difficult to avoid despite efforts to recruit from the entire population. A natural question to ask is whether the neural bases of mental retardation, and of genius at the other end of the spectrum, lie in the same continuum as what we describe in this study. Future studies in samples with a larger range of intelligence should explore this important question.

Though we qualified our approach as state-of-the-art on the behavioral side, intelligence researchers would likely object that deriving factor scores is a thing of the past, and that analyses should be conducted in latent space (in part because of factor indeterminacy). This objection can mostly be ignored in our situation, where we only looked at the general factor *g*, given that it can be so precisely estimated in our dataset. However to study specific factors, it would be advisable to overhaul our predictive analysis so it can be performed in latent space, within a structural equation model. Others have explored such a framework in the study of brain-behavior relationships [41].

We were very careful to regress out several potential confounds [115], such as brain size, gender and age, before performing predictions. While this gives us some comfort that the results reported here are indeed specific to general intelligence, the confound regression approach could certainly be improved further. There are two main concerns: one the one hand, we may be throwing out relevant variance and injecting noise into the *g* scores by bluntly regressing out confounding variables -- a more careful cleanup should be attempted, for example using a well-specified structural equation model [124]. However, the approach we took is superior to ignoring the issue of potential confounds altogether, which is likely to inflate predictability and compromise interpretation. On the other hand, the list of confounding variables that we considered was not exhaustive: for example, we did not regress out variance in the Openness factor of personality, which we have previously found to be correlated with

intelligence [65], and there are certain to be other confounds that were not measured at all. Furthermore, it is very likely that regressing out the mean framewise displacement does not properly account for non-linear effects of motion. Cleaning resting-state fMRI data of the effects of motion remains a very intense topic of research for studies of brain-behavior relationships [83,85,125]. While we are confident that our current results are not solely explained by motion in the scanner, a full quantification of this issue remains warranted.

It is worth mentioning a related, entirely data-driven study that was recently conducted by Smith and colleagues on an earlier release of the same HCP dataset (N=461) [126]. Using canonical correlation analysis, the authors demonstrated that a network of brain regions that closely resembled the default mode network was highly related to a linear combination of behavioral scores that they label a “positive-negative mode of population covariation”. In essence, this combination is a neurally derived general factor, encompassing cognitive and other behavioral tasks. Our study is in good agreement with these results, as we found high predictability of general intelligence *g* from connections within the default mode network.

Where do we go from here? As we know that task functional MRI [38,56] and structural MRI data (brain size [43], as well as morphometric features [82]) also hold information that is predictive of cognitive ability, a natural question is whether combining functional and structural data will allow us to account for more variance in the general intelligence factor *g*. The more variance we can account for, the more trustworthy and thus interpretable our models become, and we can hope to further refine our understanding of the neural bases of general intelligence.

Of course, mere prediction does not yet illuminate mechanisms, and we would ultimately wish to have a much more detailed causal model that explains how genetic factors, brain structure, brain function, and individual differences in variables such as *g* and personality relate to real-life outcomes. Given that *g* is already known to predict outcomes such as lifespan and salary, a structural model incorporating all of these variables should provide us with the most comprehensive understanding of the mechanisms, and the most effective information for targeted interventions.

Finally, we would like to situate this paper in the broader context of this special issue. Intelligence can be quantified across species and is highly heritable. Are similar brain networks the most predictive of variability in intelligence across mammals? Are there measures of heritability or brain structure, as compared to brain function, that might be better predictors in some species than others? It would be intriguing to find that humans share with other species a core set of genetically specified constraints on intelligence, but that humans are unique in the extent to which education and learning can modify intelligence through the incorporation of additional variability in brain function.

Author contributions

J. Dubois and P. Galdi developed the overall general analysis framework and conducted some of the initial analyses for the paper. J. Dubois conducted all final analyses and produced all figures. L. Paul helped with literature search, analysis of behavioral data, and interpretation of the results. J. Dubois and R. Adolphs wrote the initial manuscript and all authors contributed

to the final manuscript. All authors contributed to planning and discussion on this project. The authors declare no conflict of interest.

Acknowledgments

This work was supported by NIMH grant 2P50MH094258 (PI: RA), the Chen Neuroscience Institute, the Carver Mead Seed Fund, and a NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation (PI: JD). We thank Stuart Ritchie, Gilles Gignac, William Revelle, and Ruben Gur for invaluable advice on the behavioral side of the analyses -- though the final analytical choices rest solely with the authors.

Data Sharing

The Young Adult HCP dataset is publicly available at https://www.humanconnectome.org/study/hcp-young-adult. Analysis scripts are available in the following public repository: https://github.com/adolphslab/HCP_MRI-behavior.

References

1. Gottfredson LS. 1997 Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography.

2. Gottfredson LS. 1997 Why g matters: The complexity of everyday life. *Intelligence* **24**, 79–132.

3. Jensen AR. 1981 *Straight talk about mental tests*. The Free Press.

4. Burkart JM, Schubiger MN, van Schaik CP. 2016 The evolution of general intelligence. *Behav. Brain Sci.* , 1–65.

5. Plomin R, Deary IJ. 2015 Genetics and intelligence differences: five special findings. *Mol. Psychiatry* **20**, 98–108.

6. Kuncel NR, Hezlett SA. 2010 Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions. *Curr. Dir. Psychol. Sci.* **19**, 339–345.

7. Strenze T. 2007 Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* **35**, 401–426.

8. Deary IJ, Taylor MD, Hart CL, Wilson V, Smith GD, Blane D, Starr JM. 2005 Intergenerational social mobility and mid-life status attainment: Influences of childhood intelligence, childhood social factors, and education. *Intelligence* **33**, 455–472.

9. Gottfredson LS. 2004 Intelligence: is it the epidemiologists' elusive' fundamental cause' of

social class inequalities in health? *J. Pers. Soc. Psychol.* **86**, 174.

10. Deary I. 2008 Why do intelligent people live longer? *Nature* **456**, 175–176.

11. Calvin CM, Deary IJ, Fenton C, Roberts BA, Der G, Leckenby N, Batty GD. 2011 Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. *Int. J. Epidemiol.* **40**, 626–644.

12. Boring EG. 1923 Intelligence as the tests test it. *New Republic* **36**, 35–37.

13. Wechsler D. 2008 *WAIS-IV: Wechsler Adult Intelligence Scale*. Pearson.

14. Spearman C. 1904 ‘ General Intelligence,’ Objectively Determined and Measured. *Am. J. Psychol.* **15**, 201–292.

15. Kolata S, Light K, Matzel LD. 2008 Domain-Specific and Domain-General Learning Factors are Expressed in Genetically Heterogeneous CD-1 mice. *Intelligence* **36**, 619–629.

16. Jensen AR. 1998 *The g factor: The science of mental ability*. Praeger Westport, CT.

17. Johnson W, Nijenhuis J te, Bouchard TJ. 2008 Still just 1 g: Consistent results from five test batteries. *Intelligence* **36**, 81–95.

18. Carroll JB. 1993 *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press.

19. Kovacs K, Conway ARA. 2016 Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychol. Inq.* **27**, 151–177.

20. Thomson GH. 1916 A hierarchy without a general factor. *Br. J. Psychol.* **8**, 271–281.

21. van der Maas HLJ, Dolan CV, Grasman RPPP, Wicherts JM, Huizenga HM, Raijmakers MEJ. 2006 A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol. Rev.* **113**, 842–861.

22. Bartholomew DJ, Deary IJ, Lawn M. 2009 A new lease of life for Thomson’s bonds model of intelligence. *Psychol. Rev.* **116**, 567–579.

23. McFarland DJ. 2012 A single g factor is not necessary to simulate positive correlations between cognitive tests. *J. Clin. Exp. Neuropsychol.* **34**, 378–384.

24. Kovacs K, Conway ARA. 2016 Has g Gone to POT? *Psychol. Inq.* **27**, 241–253.

25. McGrew KS. 2009 CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* **37**, 1–10.

26. Horn JL, Blankson N. 2005 Foundations for a better understanding of cognitive abilities. In *Contemporary intellectual assessment: Theories, tests, and issues* (eds DP Flanagan, PL Harrison), pp. 41–68.

27. Johnson W, Bouchard TJ. 2005 The structure of human intelligence: It is verbal, perceptual,

and image rotation (VPR), not fluid and crystallized. *Intelligence* **33**, 393–416.

28. Major JT, Johnson W, Deary IJ. 2012 Comparing models of intelligence in Project TALENT: The VPR model fits better than the CHC and extended Gf–Gc models. *Intelligence* **40**, 543–559.

29. Horn JL. 1989 Models of intelligence. *Intelligence: Measurement, theory, and public policy* , 29–73.

30. Horn JL, Cattell RB. 1967 Age differences in fluid and crystallized intelligence. *Acta Psychol.* **26**, 107–129.

31. Hartshorne JK, Germine LT. 2015 When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychol. Sci.* **26**, 433–443.

32. Deary IJ, Penke L, Johnson W. 2010 The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.* **11**, 201–211.

33. Deary IJ, Pattie A, Starr JM. 2013 The stability of intelligence from age 11 to age 90 years: the Lothian birth cohort of 1921. *Psychol. Sci.* **24**, 2361–2368.

34. Plomin R, Fulker DW, Corley R, DeFries JC. 1997 Nature, Nurture, and Cognitive Development from 1 to 16 Years: A Parent-Offspring Adoption Study. *Psychol. Sci.* **8**, 442–447.

35. Davies G *et al.* 2011 Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol. Psychiatry* **16**, 996–1005.

36. Davies G *et al.* 2015 Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). *Mol. Psychiatry* **20**, 183.

37. Arslan RC, Penke L. 2015 Zeroing in on the Genetics of Intelligence. *Journal of Intelligence* **3**, 41–45.

38. Jung RE, Haier RJ. 2007 The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* **30**, 135–54; discussion 154–87.

39. Colom R, Karama S, Jung RE, Haier RJ. 2010 Human intelligence and brain networks. *Dialogues Clin. Neurosci.* **12**, 489–501.

40. Haier RJ. 2016 *The Neuroscience of Intelligence*. Cambridge University Press.

41. Kievit RA, van Rooijen H, Wicherts JM, Waldorp LJ, Kan K-J, Scholte HS, Borsboom D. 2012 Intelligence and the brain: A model-based approach. *Cogn. Neurosci.* **3**, 89–97.

42. Galton F. 1888 Head Growth in Students at the University of Cambridge. *Nature* **38**, 14.

43. Pietschnig J, Penke L, Wicherts JM, Zeiler M, Voracek M. 2015 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.

44. Gignac GE, Bates TC. 2017 Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence* **64**, 18–29.

45. Willerman L, Schultz R, Neal Rutledge J, Bigler ED. 1991 In vivo brain size and intelligence. *Intelligence* **15**, 223–228.

46. McDaniel MA. 2005 Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence* **33**, 337–346.

47. Narr KL, Woods RP, Thompson PM, Szeszko P, Robinson D, Dimtcheva T, Gurbani M, Toga AW, Bilder RM. 2007 Relationships between IQ and regional cortical gray matter thickness in healthy adults. *Cereb. Cortex* **17**, 2163–2171.

48. Penke L, Maniega SM, Bastin ME, Valdés Hernández MC, Murray C, Royle NA, Starr JM, Wardlaw JM, Deary IJ. 2012 Brain white matter tract integrity as a neural foundation for general intelligence. *Mol. Psychiatry* **17**, 1026–1030.

49. Kievit RA, Davis SW, Griffiths J, Correia MM, Cam-Can, Henson RN. 2016 A watershed model of individual differences in fluid intelligence. *Neuropsychologia* **91**, 186–198.

50. Penke L, Valdés Hernandéz MC, Maniega SM, Gow AJ, Murray C, Starr JM, Bastin ME, Deary IJ, Wardlaw JM. 2012 Brain iron deposits are associated with general cognitive ability and cognitive aging. *Neurobiol. Aging* **33**, 510–517.e2.

51. Ritchie SJ *et al.* 2015 Beyond a bigger brain: Multivariable structural brain imaging and intelligence. *Intelligence* **51**, 47–56.

52. Jung RE, Gasparovic C, Chavez RS, Caprihan A, Barrow R, Yeo RA. 2009 Imaging intelligence with proton magnetic resonance spectroscopy. *Intelligence* **37**, 192–198.

53. Adolphs R. 2016 Human Lesion Studies in the 21st Century. *Neuron* **90**, 1151–1153.

54. Rorden C, Karnath H-O. 2004 Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nat. Rev. Neurosci.* **5**, 813–819.

55. Gläscher J, Rudrauf D, Colom R, Paul LK, Tranel D, Damasio H, Adolphs R. 2010 Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences* **107**, 4705–4709.

56. Neubauer AC, Fink A. 2009 Intelligence and neural efficiency. *Neurosci. Biobehav. Rev.* **33**, 1004–1023.

57. van den Heuvel MP, Stam CJ, Kahn RS, Hulshoff Pol HE. 2009 Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* **29**, 7619–7624.

58. Cole MW, Yarkoni T, Repovs G, Anticevic A, Braver TS. 2012 Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *J. Neurosci.* **32**, 8988–8999.

59. Hilger K, Ekman M, Fiebach CJ, Basten U. 2017 Efficient hubs in the intelligent brain: Nodal

efficiency of hub regions in the salience network is associated with general intelligence. *Intelligence* **60**, 10–25.

60. Hilger K, Ekman M, Fiebach CJ, Basten U. 2017 Intelligence is associated with the modular structure of intrinsic brain networks. *Sci. Rep.* **7**, 16088.

61. Dubois J, Adolphs R. 2016 Building a Science of Individual Differences from fMRI. *Trends Cogn. Sci.* **20**, 425–443.

62. Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. 2015 Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671.

63. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K, WU-Minn HCP Consortium. 2013 The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79.

64. Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D. 2017 Influences on the Test-Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. *Cereb. Cortex* , 1–15.

65. Dubois J, Galdi P, Han Y, Paul LK, Adolphs R. 2017 Predicting personality traits from resting-state fMRI. *bioRxiv*. , 215129. (doi:10.1101/215129)

66. Gignac GE, Szodorai ET. 2016 Effect size guidelines for individual differences researchers. *Pers. Individ. Dif.* **102**, 74–78.

67. Glasser MF *et al.* 2013 The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124.

68. Gur RC *et al.* 2012 Age group and sex differences in performance on a computerized neurocognitive battery in children age 8-21. *Neuropsychology* **26**, 251–265.

69. Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE. 2001 Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology* **25**, 766–776.

70. Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, Brensinger C, Gur RE. 2010 A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *J. Neurosci. Methods* **187**, 254–262.

71. Moore TM, Reise SP, Gur RE, Hakonarson H, Gur RC. 2015 Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology* **29**, 235–246.

72. Raven JC. 1938 *Raven's progressive matrices (1938) : sets A, B, C, D, E*. Melbourne: Australian Council for Educational Research.

73. Williams JE, McCord DM. 2006 Equivalence of standard and computerized versions of the Raven Progressive Matrices Test. *Comput. Human Behav.* **22**, 791–800.

74. Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC. 2012 Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment* **19**, 354–369.

75. Barch DM *et al.* 2013 Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189.

76. Crum RM, Anthony JC, Bassett SS, Folstein MF. 1993 Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA* **269**, 2386–2391.

77. Revelle W. 2016 psych: Procedures for Psychological, Psychometric, and Personality Research [Software].

78. Schmid J, Leiman JM. 1957 The development of hierarchical factor solutions. *Psychometrika* **22**, 53–61.

79. Hu L, Bentler PM. 1999 Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Modeling* **6**, 1–55.

80. Grice JW. 2001 Computing and evaluating factor scores. *Psychol. Methods* **6**, 430–450.

81. Rosseel Y. 2012 lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*. **48**, 1–36.

82. Seidlitz J *et al.* 2018 Morphometric Similarity Networks Detect Microscale Cortical Organization and Predict Inter-Individual Cognitive Variation. *Neuron* **97**, 231–247.e7.

83. Siegel JS, Mitra A, Laumann TO, Seitzman BA, Raichle M, Corbetta M, Snyder AZ. 2016 Data Quality Influences Observed Links Between Functional Connectivity and Behavior. *Cereb. Cortex* (doi:10.1093/cercor/bhw253)

84. Caballero-Gaudes C, Reynolds RC. 2016 Methods for cleaning the BOLD fMRI signal. *Neuroimage* (doi:10.1016/j.neuroimage.2016.12.018)

85. Ciric R *et al.* 2017 Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage* (doi:10.1016/j.neuroimage.2017.03.020)

86. Murphy K, Fox MD. 2017 Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *Neuroimage* **154**, 169–173.

87. Liu TT. 2016 Noise contributions to the fMRI signal: An overview. *Neuroimage* **143**, 141–151.

88. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. 2014 Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341.

89. Murphy K, Birn RM, Bandettini PA. 2013 Resting-state fMRI confounds and cleanup. *Neuroimage* **80**, 349–359.

90. Walt S van der, Colbert SC, Varoquaux G. 2011 The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **13**, 22–30.

91. Pedregosa F *et al.* 2011 Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830.

92. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS. 2011 Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* **5**, 13.

93. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, Varoquaux G. 2014 Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8**, 14.

94. Gorgolewski KJ *et al.* 2017 nipy/nipype: Release 0.13.1. (doi:10.5281/zenodo.581704)

95. Robinson EC, Jbabdi S, Glasser MF, Andersson J, Burgess GC, Harms MP, Smith SM, Van Essen DC, Jenkinson M. 2014 MSM: a new flexible framework for Multimodal Surface Matching. *Neuroimage* **100**, 414–426.

96. Glasser MF *et al.* 2016 A multi-modal parcellation of human cerebral cortex. *Nature* (doi:10.1038/nature18933)

97. Ito T, Kulkarni KR, Schultz DH, Mill RD, Chen RH, Solomyak LI, Cole MW. 2017 Cognitive task information is transferred between brain regions via resting-state network topology. *Nat. Commun.* **8**, 1027.

98. Piñeiro G, Perelman S, Guerschman JP, Paruelo JM. 2008 How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Modell.* **216**, 316–322.

99. Alexander DLJ, Tropsha A, Winkler DA. 2015 Beware of R(2): Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J. Chem. Inf. Model.* **55**, 1316–1322.

100. Noirhomme Q, Lesenfants D, Gomez F, Soddu A, Schrouff J, Garraux G, Luxen A, Phillips C, Laureys S. 2014 Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions. *Neuroimage Clin* **4**, 687–694.

101. Combrisson E, Jerbi K. 2015 Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* **250**, 126–136.

102. McDonald RP. 1970 THE THEORETICAL FOUNDATIONS OF PRINCIPAL FACTOR ANALYSIS, CANONICAL FACTOR ANALYSIS, AND ALPHA FACTOR ANALYSIS. *Br. J. Math. Stat. Psychol.* **23**, 1–21.

103. Zinbarg RE, Revelle W, Yovel I, Li W. 2005 Cronbach's α, Revelle's β, and Mcdonald's ωH: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* **70**, 123–133.

104. Zinbarg RE, Yovel I, Revelle W, McDonald RP. 2006 Estimating Generalizability to a Latent

Variable Common to All of a Scale's Indicators: A Comparison of Estimators for ωh. *Appl. Psychol. Meas.* **30**, 121–144.

105. Reise SP. 2012 Invited Paper: The Rediscovery of Bifactor Measurement Models. *Multivariate Behav. Res.* **47**, 667–696.

106. Di Stefano C, Zhu M, Mîndrilă D. 2009 Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research & Evaluation* **14**.

107. Ruigrok ANV, Salimi-Khorshidi G, Lai M-C, Baron-Cohen S, Lombardo MV, Tait RJ, Suckling J. 2014 A meta-analysis of sex differences in human brain structure. *Neurosci. Biobehav. Rev.* **39**, 34–50.

108. Trabzuni D, Ramasamy A, Imran S, Walker R, Smith C, Weale ME, Hardy J, Ryten M, North American Brain Expression Consortium. 2013 Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* **4**, 2771.

109. Dosenbach NUF *et al.* 2010 Prediction of individual brain maturity using fMRI. *Science* **329**, 1358–1361.

110. Geerligs L, Renken RJ, Saliasi E, Maurits NM, Lorist MM. 2015 A Brain-Wide Study of Age-Related Changes in Functional Connectivity. *Cereb. Cortex* **25**, 1987–1999.

111. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. 2012 Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154.

112. Satterthwaite TD *et al.* 2013 An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256.

113. Tyszka JM, Kennedy DP, Paul LK, Adolphs R. 2014 Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cereb. Cortex* **24**, 1894–1905.

114. Hänggi J, Fövenyi L, Liem F, Meyer M, Jäncke L. 2014 The hypothesis of neuronal interconnectivity as a function of brain size-a general organization principle of the human connectome. *Front. Hum. Neurosci.* **8**, 915.

115. Smith SM, Nichols TE. 2018 Statistical Challenges in 'Big Data' Human Neuroimaging. *Neuron* **97**, 263–268.

116. Elam J. 2015 Ramifications of Image Reconstruction Version Differences. *HCP Data Release Updates: Known Issues and Planned fixes*. See https://wiki.humanconnectome.org/display/PublicData/Ramifications+of+Image+Reconstruction+Version+Differences.

117. Smith S *et al.* 2016 Second beta-release of the HCP Functional Connectivity MegaTrawl. See https://db.humanconnectome.org/megatrawl/HCP820_MegaTrawl_April2016.pdf.

118. Laumann TO *et al.* 2015 Functional System and Areal Organization of a Highly Sampled

Individual Human Brain. *Neuron* **87**, 657–670.

119. Yeo BTT *et al.* 2011 The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165.

120. Deary IJ, Cox SR, Ritchie SJ. 2016 Getting Spearman off the Skyhook: One More in a Century (Since Thomson, 1916) of Attempts to Vanquish g. *Psychol. Inq.* **27**, 192–199.

121. Mar RA, Spreng RN, Deyoung CG. 2013 How to produce personality neuroscience research with high statistical power and low additional cost. *Cogn. Affect. Behav. Neurosci.* **13**, 674–685.

122. Le H, Schmidt FL. 2006 Correcting for indirect range restriction in meta-analysis: testing a new meta-analytic procedure. *Psychol. Methods* **11**, 416–438.

123. Henrich J, Heine SJ, Norenzayan A. 2010 Most people are not WEIRD. *Nature* **466**, 29.

124. Westfall J, Yarkoni T. 2016 Statistically Controlling for Confounding Constructs Is Harder than You Think. *PLoS One* **11**, e0152719.

125. Satterthwaite TD, Ciric R, Roalf DR, Davatzikos C, Bassett DS, Wolf DH. 2017 Motion artifact in studies of functional connectivity: Characteristics and mitigation strategies. *Hum. Brain Mapp.* (doi:10.1002/hbm.23665)

126. Smith SM *et al.* 2015 A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* **18**, 1565–1567.

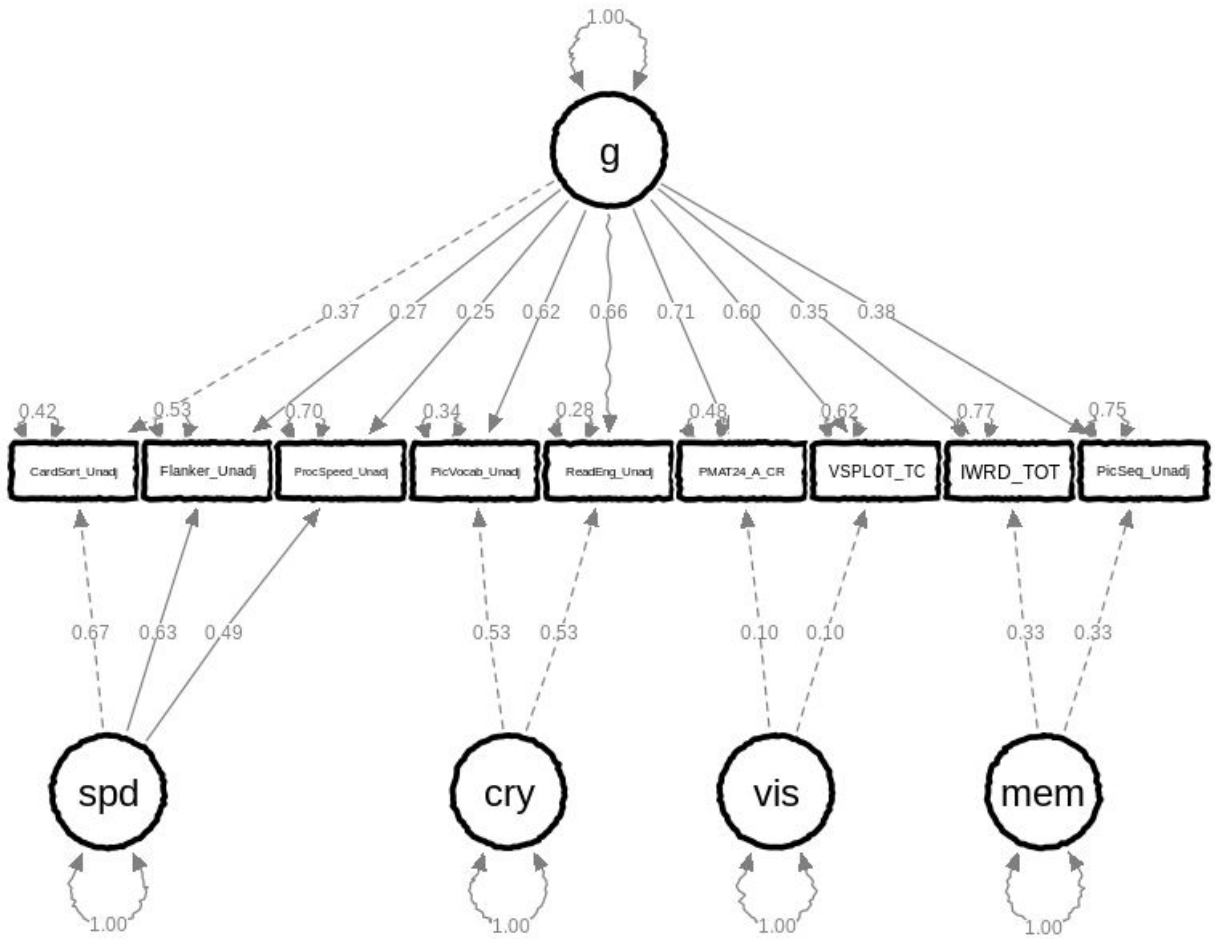
Supplementary material

Confirmatory factor analysis

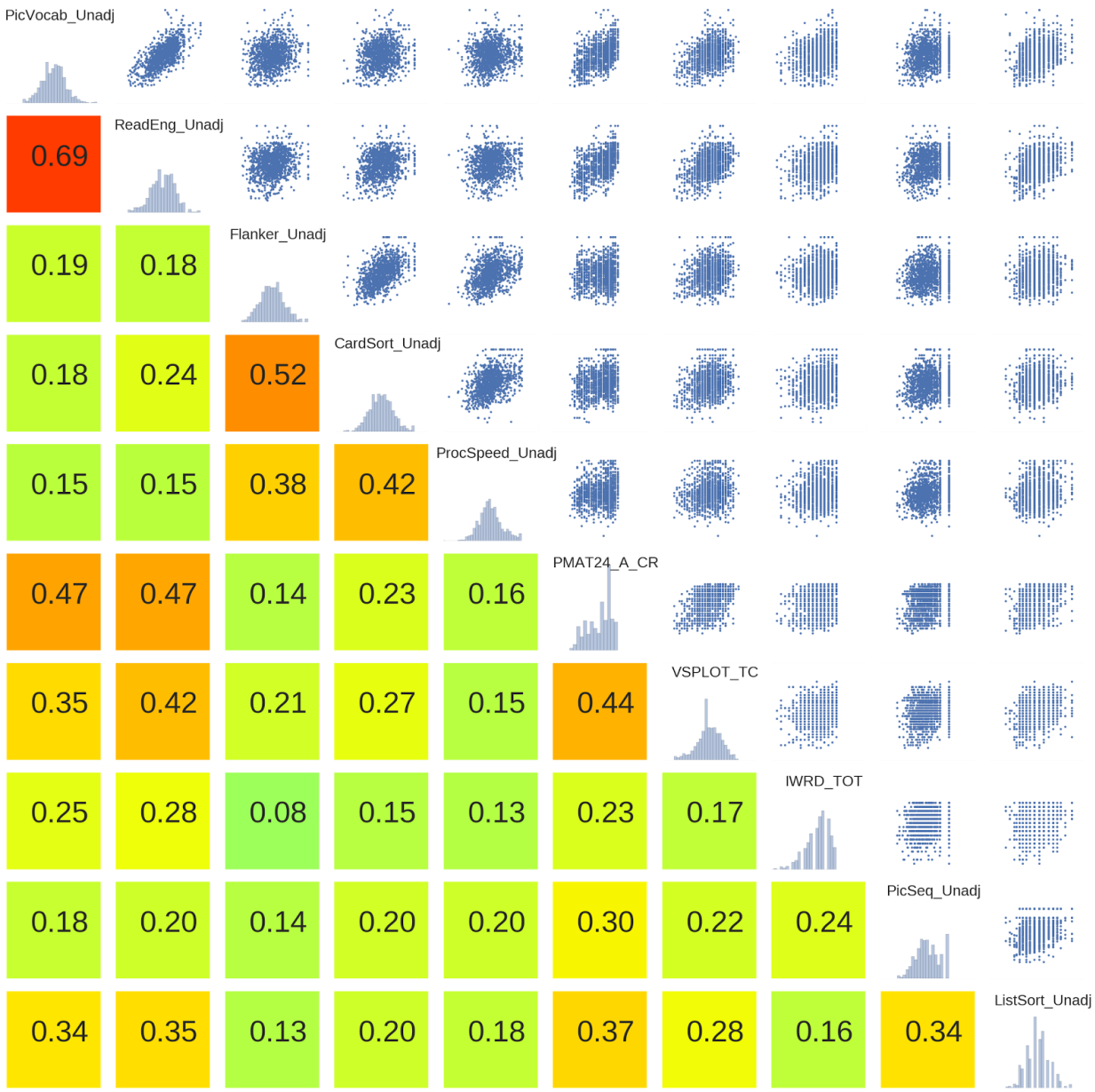
Using the EFA solution, we specify a bi-factor model, including a general factor (loading on all tasks) and four group factors (loading on subsets of tasks), in a confirmatory factor analysis. The model does not allow for any cross-loadings of a task on several factors, and group factors are orthogonal to one another and to the general factor. As some of the group factors are only defined by only two indicators, it is necessary to impose a constraint for the purposes of identification. We fix the unstandardized loadings for both tasks to 1.0 in this case. We initially find that the lavaan model does not converge, and identify that the issue lies with the ListSort_Unadj task score. We run the CFA without ListSort_Unadj, with the following lavaan syntax:

```
#g-factor
g      =~ CardSort_Unadj + Flanker_Unadj + ProcSpeed_Unadj + PicVocab_Unadj +
    ReadEng_Unadj + PMAT24_A_CR + VSPLOT_TC + IWRD_TOT + PicSeq_Unadj
#Domain factors
spd    =~ CardSort_Unadj + Flanker_Unadj + ProcSpeed_Unadj
cry    =~ 1*PicVocab_Unadj + 1*ReadEng_Unadj
vis    =~ 1*PMAT24_A_CR    + 1*VSPLOT_TC
mem =~ 1*IWRD_TOT      + 1*PicSeq_Unadj
#Domain factors are not correlated with g
g ~~ 0*spd
g ~~ 0*cry
g ~~ 0*vis
g ~~ 0*mem
#Domain factors are not correlated with one another
spd ~~ 0*cry
spd ~~ 0*vis
spd ~~ 0*mem
cry ~~ 0*vis
cry ~~ 0*mem
vis ~~ 0*mem
```

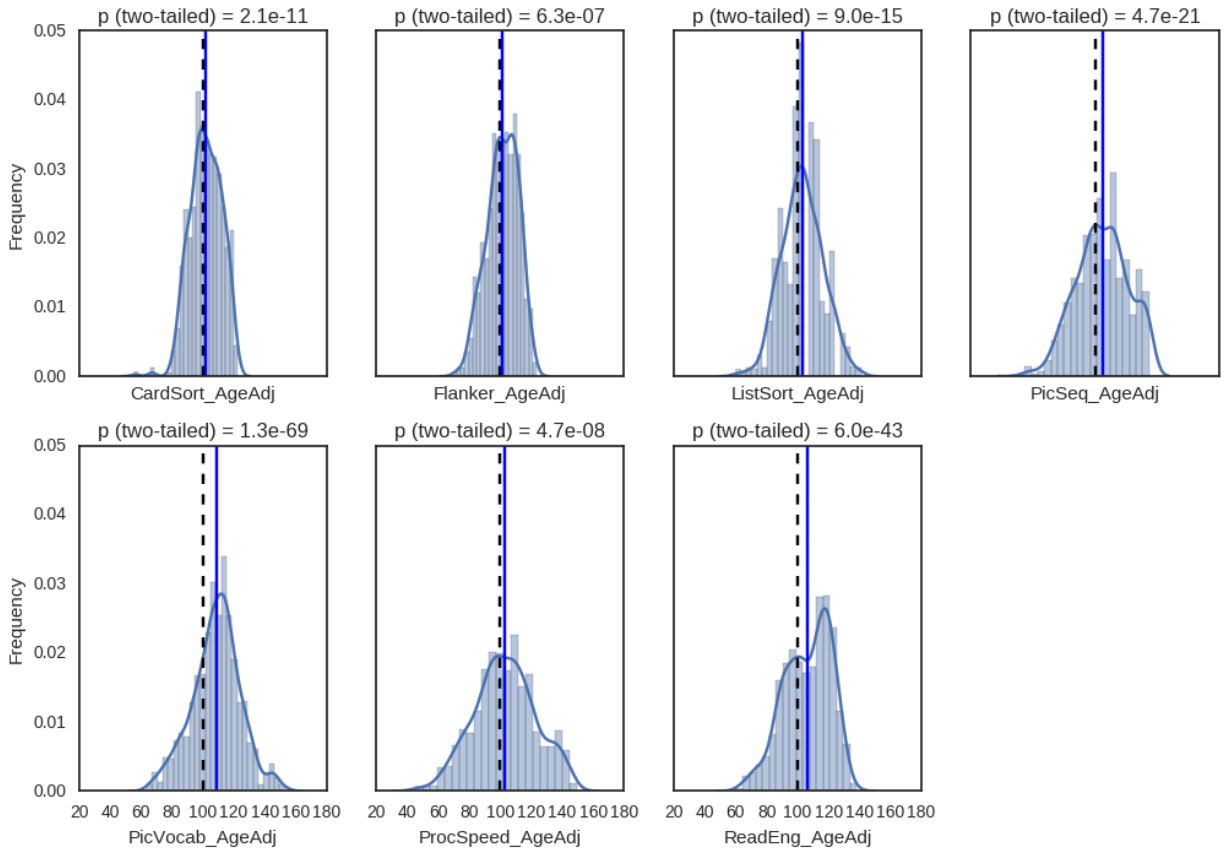
This CFA model converges after 49 iterations, and the fit is very good with CFI=0.974, RMSEA=0.052, SRMR=0.032, BIC = 27820.2. The standardized solution is depicted in **Supplementary Figure 1**. The general factor is found to account for 64.0% of the variance (coefficient omega_hierarchical ω_h), while group factors account for 17.2% of the variance.
We derive the factor scores for *g* using the regression method; we find that the scores derived from the CFA are almost perfectly correlated with the scores derived from the EFA (Figure 1), r=0.99.



Supplementary Figure 1. Standardized solution for our confirmatory factor analysis of the HCP cognitive task scores. The CFA omits ListSort_Unadj which prevent the model from converging.



Supplementary Figure 2. Correlations between (z-scored) HCP cognitive task scores. On the diagonal, the distribution of each of the 11 cognitive variables in shown. Below the diagonal, the Pearson correlation is displayed, together with a color visualizing the strength of the relationship. Above the diagonal, a scatter plot is displayed for each pair of variables, with x- and y- axes between -4 and 4 (standard deviation of all variables is 1 due to z-scoring).



Supplementary Figure 3. Distributions of the age-normed scores of HCP subjects on NIH-toolbox cognitive tasks. The blue line shows the mean score in our subject sample, which is greater than 100 for all tests, while the black dashed line shows the mean in the normative population. For all tests, a 1-sample Student's t-test indicates that the mean is significantly higher than 100.