Making Cognitive Latent Variables Manifest: Distinct Neural Networks for Fluid Reasoning and Processing Speed

Christian Habeck[1], Jason Steffener[1], Daniel Barulli[1], Yunglin Gazes[1], Qolamreza Razlighi[1], Danielle Shaked[1], Timothy Salthouse[2], and Yaakov Stern[1]

Abstract

■ Cognitive psychologists posit several specific cognitive abilities that are measured with sets of cognitive tasks. Tasks that purportedly tap a specific underlying cognitive ability are strongly correlated with one another, whereas performances on tasks that tap different cognitive abilities are less strongly correlated. For these reasons, latent variables are often considered optimal for describing individual differences in cognitive abilities. Although latent variables cannot be directly observed, all cognitive tasks representing a specific latent ability should have a common neural underpinning. Here, we show that cognitive tasks representing one ability (i.e., either perceptual speed or fluid reasoning) had a neural activation pattern distinct from that of tasks in the other ability. One hundred six participants between the ages of 20 and 77 years were imaged in an fMRI scanner while performing six cognitive tasks, three representing each cognitive ability. Consistent with prior research, behavioral performance on these six tasks clustered into the two abilities based on their patterns of individual differences and tasks postulated to represent one ability showed higher similarity across individuals than tasks postulated to represent a different ability. This finding was extended in the current report to the spatial resemblance of the task-related activation patterns: The topographic similarity of the mean activation maps for tasks postulated to reflect the same reference ability was higher than for tasks postulated to reflect a different reference ability. Furthermore, for any task pairing, behavioral and topographic similarities of underlying activation patterns are strongly linked. These findings suggest that differences in the strengths of correlations between various cognitive tasks may be because of the degree of overlap in the neural structures that are active when the tasks are being performed. Thus, the latent variable postulated to account for correlations at a behavioral level may reflect topographic similarities in the neural activation across different brain regions. ■

INTRODUCTION

General intelligence, in particular, its utility in predicting specialized competence and the challenge of its accurate measurement, has long stimulated enormous interest in psychological research (e.g., Marquart & Bailey, 1955; Bottorf, 1946; Embree, 1946; Kendler, 1946; Thomson, 1946; Willoughby, 1928). With the advent of fMRI imaging, neural substrates of general and fluid intelligence have been mapped in the brain, primarily to frontoparietal and temporal locations (Barbey, Colom, Paul, & Grafman, 2014; Hampshire, Thompson, Duncan, & Owen, 2011; Preusse, van der Meer, Deshpande, Krueger, & Wartenburger, 2011; Colom, Karama, Jung, & Haier, 2010; Woolgar et al., 2010). Furthermore, rigorous attempts at identifying general and specific components of intelligence by combining neuroimaging and use of latent variable modeling have been undertaken recently (Hampshire, Highfield, Parkin, & Owen, 2012). Cognitive psychologists often use the construct of a latent variable to best represent a hypothesized cognitive ability. The basis of latent variable modeling is the observation that performances on tasks that purportedly tap a specific underlying cognitive ability are strongly correlated, whereas performances on tasks that tap different cognitive abilities are less strongly correlated. Mathematically, the latent variable represents variance across individuals that are common to the constituent indicator variables and thus represents a "pure" measure that is free of any features that are specifically associated with any one task. This has made the latent variable approach valuable for cognitive studies. For example, when assessing cognitive aging, it has been repeatedly demonstrated (Salthouse & Ferrer-Caja, 2003) that multiple tests can be summarized into a number of latent or "reference" abilities that effectively capture the key features of age-related cognitive differences. This observation provides a basis for focusing on the key features of cognitive aging rather than on individual tasks that potentially have idiosyncratic features.

[1]Columbia University College of Physicians and Surgeons, [2]University of Virginia

Journal of Cognitive Neuroscience 27:6, pp. 1249–1258
doi:10.1162/jocn_a_00778

Although the statistical and theoretical basis for latent abilities is clear, by definition, they are not directly observable. Despite such difficulties of attribution, it is reasonable to hypothesize that (a) there are latent cognitive abilities, (b) any latent cognitive ability has a neural basis, and (c) constituent tasks tapping the same latent cognitive ability engage a common brain network or set of brain areas. Identifying such neural networks would provide support for the existence of the latent ability as well as a directly measurable and observable component to an otherwise theoretical construct, providing a mechanistic explanation for the observed behavioral covariances. We directly tested this hypothesis in an fMRI study by imaging participants performing a set of tasks associated with two putative latent abilities. Our prime objective was to confirm the existence of these a priori postulated abilities at the level of neuroimaging indicators without using behavioral performance (or any other guidance or constraints) to inform the analysis of the neuroimaging data to provide a possible explanation of the relations used to infer the existence of latent variables We employed simple correlational analysis to ascertain the construct validity (CV) of within-domain convergence and between-domain discrimination with regard to the prior postulated reference abilities. In contrast to Hampshire et al. (2012), we do not rely on an unsupervised multivariate decomposition of neuropsychological and imaging data with a post hoc rotation of factors. Post hoc rotations of factors are in a sense arbitrary and demand a compelling a priori motivation that should not be confounded with the final interpretation of the orthogonal or oblique relationships of the factor themselves. For this reason and the fact that the relationship of rotated factors was integral to the interpretation propounded in the article, the study by Hampshire et al. (2012) has encountered some criticism (Haier, Karama, Colom, Jung, & Johnson, 2014). We have fewer interpretational difficulties because we are confirming the labeling taxonomy suggested by Salthouse and Ferrer-Caja (2003). Figure 1 shows a schematic sketch for the illustration of our method of assessing CV.

In the figure, good CV of a latent variable model with strong loadings of the indicator variables onto their separate respective latent variables results in a block-diagonal structure of the correlation matrix observed for all indicator variables. This means indicators belonging to the same reference ability correlate strongly, while correlating only weakly across reference abilities. Poor CV, on the other hand, results in all off-diagonal elements in

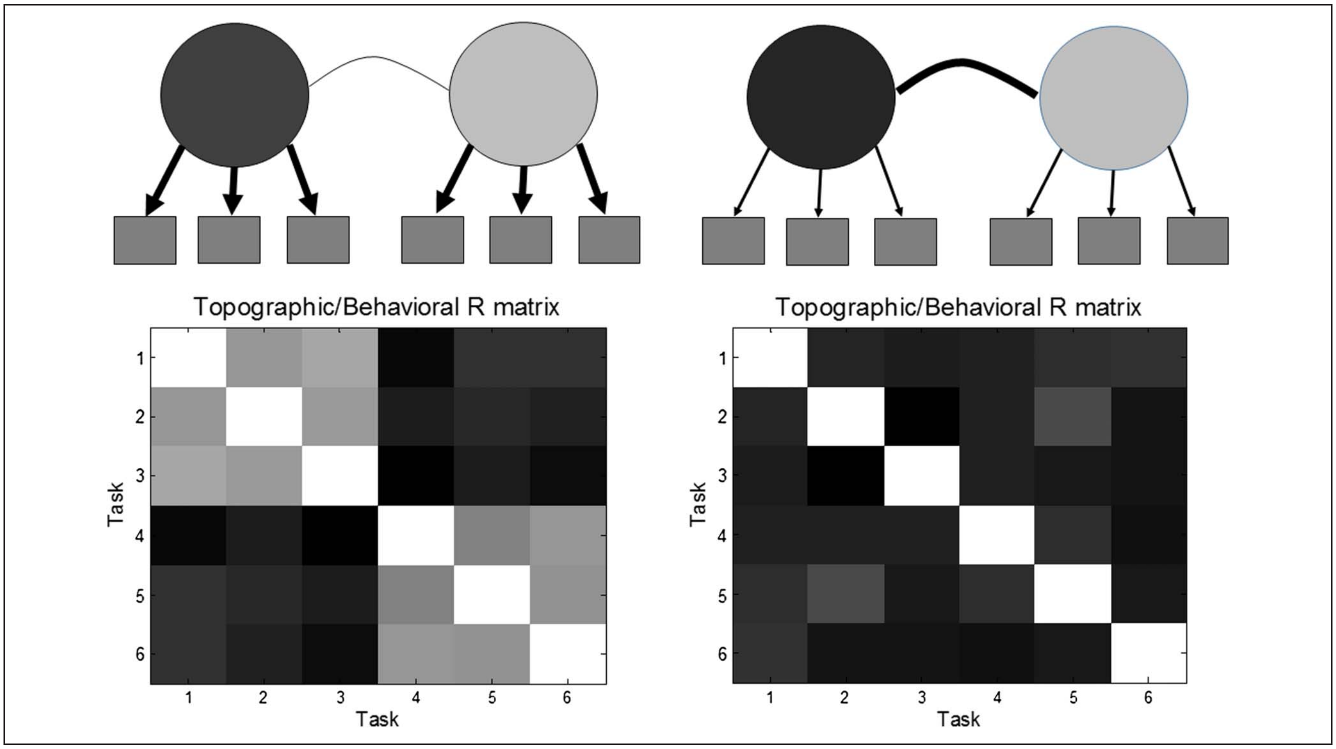
Topographic/Behavioral R matrix
Task
1
2
3
4
5
6
Task
1
2
3
4
5
6
Topographic/Behavioral R matrix
Task
1
2
3
4
5
6
Task
1
2
3
4
5
6

Figure 1. Explanation of the CV concept used extensively in the current article. The upper row shows different scenarios of how the latent reference abilities of FR and PS (round circles) can manifest themselves in the six indicator tasks (squares boxes). The lower row shows the topographic similarity of mean activation patterns or the behavioral similarity for any task pair in a 6 × 6 correlation matrix. (Left) Good CV, that is, the tasks load on strongly to their respective reference ability latent variable, although the correlation between latent reference abilities is low. This can be appreciated without latent variable modeling: The R matrix shows a block-diagonal structure, and the similarity of any two tasks belonging to the same reference ability is much higher than that of two tasks chosen from different reference abilities. (Right) The converse scenario of poor CV: Indicator tasks are only weakly loading onto their respective reference ability, and the residual correlation between the latent variables is strong. The correlation matrix has no block-diagonal structure, that is, any task's membership to a reference ability is irrelevant: All tasks share equally low topographic or behavioral similarity. Correspondingly, CV is low.

the correlation matrix to be of similar magnitude: There is no difference between correlation within or across reference abilities, and any clustering into reference abilities would thus be made on an arbitrary basis. The current study used this simple method of investigating the presence of two distinct reference abilities at the neural level, so that the CV emerges simply from the topographic similarity of the mean activation maps underlying a set of cognitive tasks. Reference-ability neural networks (RANNs) then were derived from region-wise analyses after removing neural activation common to all tasks, that is, activation that is not task specific, and this cannot be attributed to any particular reference ability.

Successful identification of RANNs has the potential of contributing to a paradigm shift in research on the neural bases of cognitive processes by emphasizing the broad and replicable neural components common to several tasks that assess the same general cognitive ability, without being hampered by task-specific "noise."

METHODS

Participants

One hundred six participants (age = 50.13 ± 16.53 years, range = 20–77 years) underwent behavioral testing and brain imaging. All participants were required to be native English speakers, to be strongly right handed, and to have at least a fourth-grade reading level. Participants were free of medical or psychiatric conditions that could affect cognition and were carefully screened to ensure that the elder participants did not meet criteria for dementia or mild cognitive impairment.

Procedure

The six cognitive activation tasks were administered in the context of a larger study. Both sessions included six tasks spread across two cognitive domains. For one session, perceptual speed (PS) and vocabulary (V) were the domains tested (task order: Synonyms [V], digit–symbol [PS], antonyms [V], letter comparison [PS], picture naming [V], pattern comparison [PS]), whereas for the other session, fluid reasoning (FR) and memory (M) were the domains tested (task order: paper folding [FR], logical memory [M], matrix reasoning [FR], word order recognition [M], letter sets [FR], paired associates [M]). The order of tasks within the two cognitive domains was not varied, but the order of the domains was varied across participants. Task stimuli were back-projected onto a screen located at the foot of the MRI bed using an LCD projector. Participants viewed the screen via a mirror system located in the head coil and, if needed, had vision corrected to normal using MR compatible glasses (manufactured by SafeVision, LLC, Webster Groves, MO). Responses were made on a LUMItouch response system (Photon Control Company, Burnaby, Canada). Task administration and collection of RT and accuracy data were controlled by E-Prime (Psychology Tools, Inc., Pittsburgh, PA) running on a PC computer. Task onset was electronically synchronized with the MRI acquisition computer.

Cognitive Tasks

In the scanner, participants receive a battery of six computerized tasks. Paper-and-pencil versions of these tasks have been found to have high internal consistency and test–retest reliabilities (i.e., .7 or greater) and moderate-to-large (i.e., .7 or greater) loadings on their respective factors. To accommodate testing in the scanner as well as to optimize power for image analysis, some changes in test structure and presentation were required. Before the scan session, the participants were familiarized with the tasks and allowed to practice each of them. For all tasks, responses were differential button presses. During training, responses were on the computer keyboard, and during scans, they were made on the LUMItouch response system. For the FR tasks, the primary dependent variable was accuracy, whereas for the speed of processing tasks, the primary dependent variable was RT. The tests used for the two cognitive domains were as follows.

PS: (1) Digit symbol: A code table is presented on the top of the screen, consisting of numbers 1 through 9, each paired with an associated symbol. Below the code table, an individual number/symbol pair is presented. Participants are asked to indicate whether the individual pair is the same as that in the code table using a differential button press. (2) Letter comparison (Salthouse & Babcock, 1991): Two strings of letters, each consisting of three to five letters "i," are presented alongside one another. Participants indicate whether the strings are the same or different using a differential button press. (3) Pattern comparison (Salthouse & Babcock, 1991): Two figures consisting of varying number of lines connecting at different angles are presented alongside one another. Participants indicated whether the figures were the same or different by a differential button press.

FR: (1) Paper folding (Ekstrom, French, Harman, & Derman, 1976): Participants select a pattern of holes that would result from a sequence of folds in a piece of paper, through which a hole is then punched. Participants choose from six options which unfolded paper best corresponds to the sequence given. (2) Matrix reasoning (adapted from Raven's 1962 Advanced Progressive Matrices): Participants discern a pattern from a series of figures displayed in a matrix. Participants are given a matrix that is divided into nine cells, in which the figure in the bottom right cell is missing. Below the matrix, they are given eight figure choices, and they are instructed to evaluate which of the figures would best complete the missing cell. (3) Letter sets

(Ekstrom et al., 1976): Participants are presented with five sets of letters, where four of the five sets have a common rule (i.e., have no vowels), with one of the sets not following this rule. Participants are instructed to select the unique set.

Image Acquisition Procedures

All MR images were acquired in a 3.0-T Philips (Andover, MA) Achieva Magnet. At each session, first, a scout, T1-weighted image was acquired to determine patient position. All scans use a 240-mm field of view. For EPI, the pulse sequence parameters were TE/TR = 20/2000, flip angle = 72°, in-plane resolution = 112 × 112 voxels, slice thickness/gap (mm) = 3/0, and 41 slices. A high-resolution structural image was also acquired for coregistration. A neuroradiologist reviewed each participant's MRI scan. Any significant findings were conveyed to the participant's primary care physician. Behavioral performance was recorded while participants executed the tasks in the scanner. *Z* scores were computed for all six behavioral variables.

Image Analysis Preprocessing Procedures

Each individual's six fMRI scans was preprocessed using the SPM8 software (Wellcome Trust Centre for Neuroimaging, London, UK) as implemented in MatLab in two basic steps: (a) standard initial processing steps including motion correction, transformation into standardized (Montreal Neurological Institute) space, and smoothing and (b) statistical analysis of individual participants' data. Statistical analysis of individual participants' data for the six cognitive tasks consisted of block-based time series analysis of each individual task as a whole for each participant via general linear modeling of the fMRI time series data. These data, along with the predictor variables, were high-pass filtered and prewhitened to explicitly correct for intrinsic autocorrelations in the data. The predictor variables comprising the first-level design matrix were composed of epochs representing each unique experimental task relative to the same duration of rest (i.e., no-performance blocks) and were convolved with a model of the hemodynamic response function. Our main interest in this article is the mean activation maps for each task. For these mean activation patterns, we can obtain a pattern score with a simple linear regression (without intercept term) across all voxels for each participant:

$$\mathbf{N}(s,t) = F\_mean(s,t)\mathbf{M}(t) + \varepsilon(s,t)$$

where $\mathbf{M}(t)$ denotes the mean activation pattern for task *t* and *F_mean(s,t)* is its pattern score of the mean pattern for participant *s*; $\boldsymbol{\varepsilon}(s,t)$ denotes subject-, task-, and voxel-specific residual activation that is orthogonal to the mean activation pattern and was not considered further in the current study.

CV Statistic

Behavioral data as well as activation patterns are expected to manifest both convergent as well as discriminant validity. To conveniently monitor these requirements, we summarized both validity requirements in one CV statistic. The CV statistic can be computed for topographic mean activation patterns, behavioral performance, or mean activation pattern scores according to the following procedure. Imagine we have a task pairing of task t1 and task t2:

(1) We can compute the topographic CV by considering the correlation across voxels of the mean activation patterns for the task pairing in question. For instance, for task pairing t1 and t2, we can compute the correlation $r(\mathbf{M}(\text{t1}),\mathbf{M}(\text{t2}))$ and then turn this into a Fisher *Z* coefficient according to $Z = 0.5 \log((1 + r)/(1 - r))$.
(2) We can compute the behavioral CV by considering the correlation across participants of the behavioral performance variables for the task pairing in question, with a subsequent conversion into Fisher *Z*.
(3) Finally, we can compute the CV of the mean pattern scores, that is, we consider the correlation of the mean pattern scores: $Z(F\_mean(:,\text{t1}), F\_mean(:,\text{t2}))$, where the colon denotes that the correlation is computed across participants, similar to the behavioral measures.

After clarifying how task similarity is established as a Fisher *Z* correlation coefficient for any task pairing, we can give the general formula for CV:

$$CV = \hat{Z}_{\text{within}} - \hat{Z}_{\text{between}}$$

Here, $\hat{Z}_{\text{within}}$ denotes the average Fisher *Z* correlation coefficient for all within-domain task pairings (=6), and $\hat{Z}_{\text{between}}$ denotes the average Fisher *Z* correlation coefficient for all between-domain task pairings (=9).

One complication arising in this computation is the fact that tasks pertaining to the two reference domains were obtained in different scanning sessions. It is thus possible that within-domain task pairings look more similar than between-domain pairings purely by virtue of having been recorded in the same session. Fortunately, we can estimate the potential increase in both topographic and behavioral CV because of session effects. All participants have six additional tasks available: Overall, there was one scanning session with three speed-of-processing tasks and three vocabulary tasks and a separate scanning session with three episodic memory tasks and three FR tasks. This means we can estimate 18 within-session similarities for task pairs coming from different reference domains and contrast these with 36 between-session similarities for task pairs coming from different reference domains to obtain an upper limit of the boost in task similarity for tasks that were recorded in the same session. The boost amounts can be estimated for all measures, that is, topographic mean activation patterns, behavioral performance, or mean activation pattern scores, and subtracted from the within-session and domain correlations that go into the construct validity computation according to

$$\hat{Z}_{\text{within}} \rightarrow \hat{Z}_{\text{within}} - \text{boost}$$

Significant construct validity statistics in all three of the aforementioned senses imply that (1) the mean activation patterns are topographically more similar, (2) behavioral performances between two tasks are more similar, and (3) the degrees of subject-specific deployments of mean task activation patterns are more similar, for a pairing of tasks within, rather than between, reference domains.

For the inferential judgment about CV, we perform a simple permutation test in which within-domain similarity is contrasted with between-domain similarity (as quantified by Fisher Z) in null data. There are six within-domain values and nine between-domain values, yielding 5,005 permutations of random group assignment, and we compute $CV = \hat{Z}_{\text{within}} - \hat{Z}_{\text{between}}$ for all of them. The p level is computed as the fraction of these 5,005 permutations that yielded construct validates larger than the point estimate value.

Split-half Validation Analyses

As an additional validation of the reference ability concept, we can use (the subject-and-task scores of) the mean activation patterns to predict which reference domain any particular task and subject activation map belong to. Split-half analyses are very helpful for this purpose. We randomly divide (all) the pool of all 6 × 106 = 636 subject-task activation maps into a training fold of 500 maps and a test fold of 136 maps. Two mean activation patterns were computed in the training sample for all FR and PS tasks, and for every map in the test fold, two scores can be computed quantifying the extent to which this particular map manifests both FR and PS patterns derived from the training set. The reference domain is predicted according to which pattern the test map loads onto more strongly. Using the gold standard knowledge of the true reference domain labels in the test fold, the classification accuracy can be recorded for all the test maps. The grouping of tasks into reference abilities can be changed and "diluted" by swapping one task between the FR and PS domains. Ideally, the prediction performance for the correct assignment of tasks into domain groupings should give (1) much better performance than chance level and (2) better performance than a "diluted" grouping.

RESULTS AND DISCUSSION

As described in the Methods, the two domains of FR and PS were investigated with three cognitive tasks each. For FR, we have (1) matrix reasoning, (2) letter sets, and (3) paper folding. For PS, we have (1) digit symbol, (2) letter comparison, and (3) pattern comparison.

Behavioral Performance

Performance was measured as accuracy of response for the FR tasks and RT for the PS tasks. Age-related performance decreases were noted for all tasks at an unadjusted p level of <.05; however, after a Bonferroni correction for a six age comparisons, letter sets and letter comparison no longer showed significant age differences.

Demonstration of CV of Reference Domains in Neuroimaging and Behavioral Data

We first examined the interrelationships of the performance data acquired from the six tasks administered in the scanner. To quantify this relationship, we computed a CV statistic, corresponding to the difference of the mean correlation between tasks within a domain and the mean correlations between domains. This CV statistic captures to what extent the correlation in performance of tasks within the same domain exceeds that of tasks that belong to different domains, and therefore, larger values indicate greater distinctiveness of the constructs. The behavioral CV statistic was 0.8945 (corrected mean within-domain $Z = 0.5602$, between-domain $Z = -0.3343$). A permutation test (see Methods) established that such a difference is unlikely to occur by chance ($p < .0002$). These results replicate numerous previous studies that have demonstrated that these six tasks can be inferred to represent two distinct cognitive abilities.

The similarity of the areas activated for any task pair can be quantified using a spatial correlation of the topographic composition of mean task activation patterns. The CV of the resulting spatial R matrix can be computed in the same way as for the behavioral data, contrasting the average within-domain Z correlation to the average between-domain Z value. We obtained a value of 0.6235 (corrected mean within-domain $Z = 1.7681$, between-domain $Z = 1.1446$), which was highly significant in the permutation procedure ($p = .0010$).

Finally, we computed a pattern score for each subject and task, quantifying to what extent a person in a particular task manifests the mean activation pattern for the task under consideration in their task activation map (see Methods). These subject and task scores can be treated in an identical manner to the behavioral variables. Again, we submitted the correlation matrix to the construct validity computation and obtained a value of 0.2505 (mean within-domain $Z = 0.5556$, between-domain $Z = 0.3052$; $p = .0006$, from permutation test). Figure 2 displays the correlation matrices for mean activation patterns and behavioral, subject, and task scores.

We then tested a more general mathematical relationship of CV concerning topographic composition of activation patterns and behavioral performance. Ideally, topographic similarity of activation patterns for any chosen pair of tasks should be linked with the corresponding similarity of behavioral performance, that is, tasks that

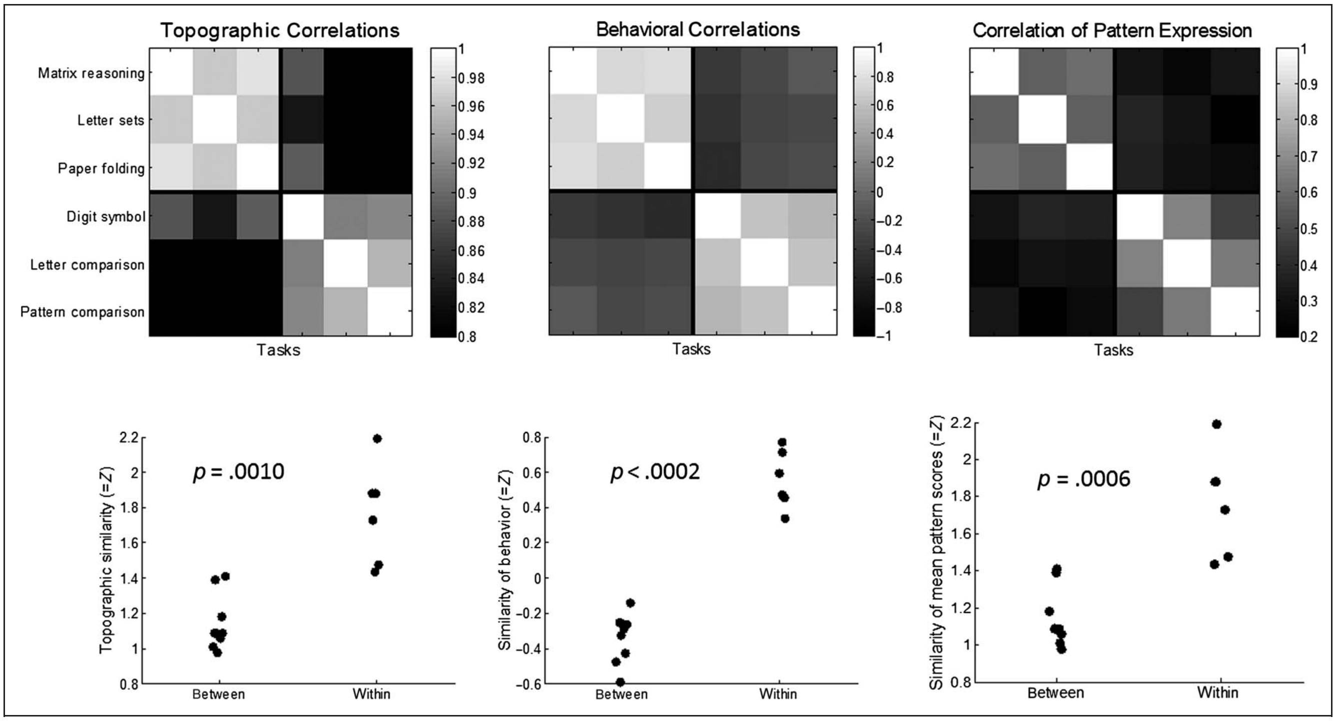

Figure 2. (Top) Correlation matrices to illustrate task clustering into reference abilities according to topographic similarity of mean task activation maps (left), behavioral similarity of task performance (middle), and usage similarity of the mean pattern scores (right). (Bottom) Corresponding scatter plots for the within-domain versus between-domain correlations (=Fisher Z), with p levels obtained from a permutation test. The similarity values are highest for the mean activation pattern topography and pattern score and much lower for behavior. The differences of between- versus within-domain similarities, however, were always highly significant, thus establishing good CV. Note also that within-domain Z values were corrected for session effects as explained in the manuscript.

activate similar brain regions across participants should also show a high correlation of behavioral performance across participants, and vice versa. As shown in Figure 3, the mean topographic similarity of the underlying activation patterns between any two tasks across participants was indeed strongly linked to the similarity of behavioral performance for that task pairing (r = .8289, p = .0001, obtained from Pearson correlation). Thus, tasks that elicit similar activation patterns are more likely to have similar performance levels for any participant, and both behavioral and neural similarity was higher for within-domain pairings.

Finally, to illuminate the plausibility of the concept of reference abilities and the particular assignment of tasks into the FR and PS reference abilities, furthermore, we conducted some split-half validation analyses (see Methods) with 1,000 iterations each. Every participant and task map gets assigned the label FR or PS according to reference ability the tasks belongs to. After splitting the total pool of subject-task maps randomly into a training set of 500 maps and a test set of 136 maps, we can compute two mean activation maps in the training sample: one mean activation map for all FR tasks and one for all PS tasks. In the test data set, we can compute to what extent every map loads onto these two mean activation patterns. This means every subject-task map gets assigned two numbers. The prediction of the reference domain in the test set is

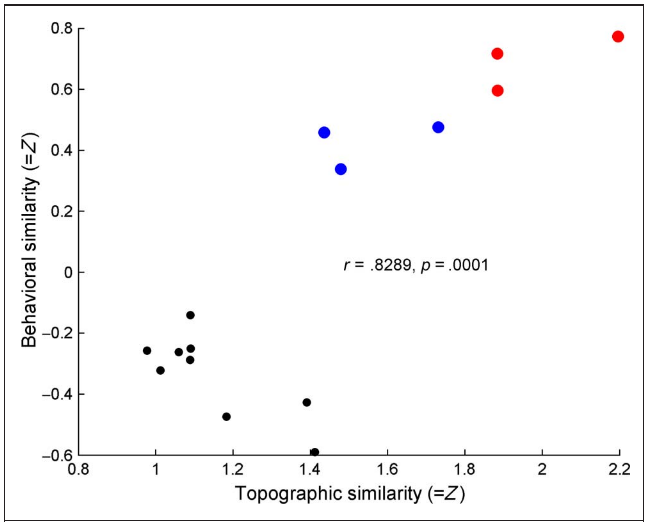

Figure 3. The topographic similarity of mean task-related patterns is closely related to the similarity of behavioral performances for any task pairing, as measured by Fisher Z. In the figure, the three red data points indicate task pairings within the domain of FR, and the three blue data points denote task pairings within the domain of PS. The remaining nine data points in black indicate all between-domain task pairings. The similarity mapping is statistically significant at r = .8289, p = .0001.

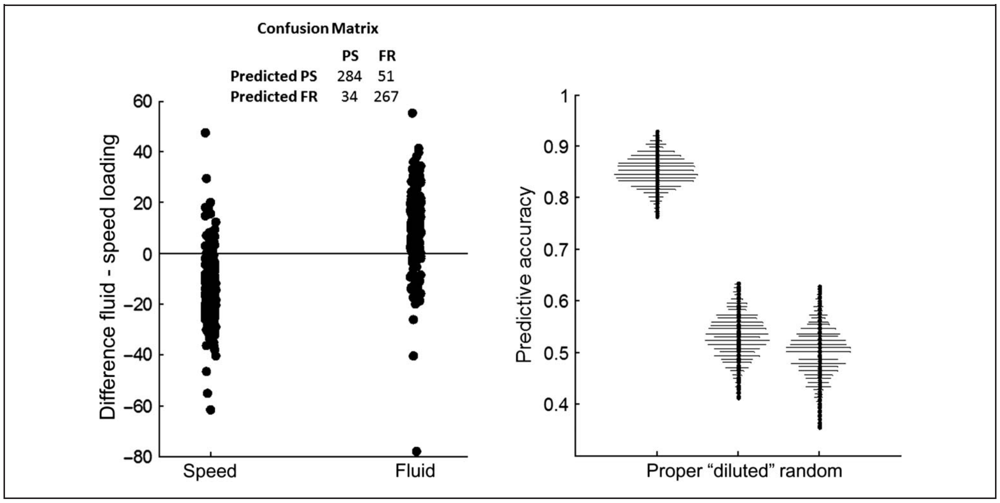

Figure 4. Split-half validation analysis with 1,000 iterations. Using the mean task pattern scores with respect to the RANN patterns derived from the training sample, we can predict the reference ability (FR or PS) of scans in the test sample with a simple difference of loadings. Shown in the left are the differences in loadings with respect to the mean FR and the mean PS activation patterns for the FR and PS scans. There is large overlap, but on average, FR maps load onto the FR pattern more strongly than PS maps, and vice versa. Imposing a decision threshold for the difference scores at threshold = 0 achieves a classification accuracy of 87%. The classification was also tried for held-out data (right) with a split-sample test of 1,000 iterations. Classification accuracies are shown with violin plots for a proper assignment of tasks into the two reference domains—FR = [matrix reasoning , letter sets, paper folding], PS = [digit symbol, letter comparison, pattern comparison]—and a “diluted” assignment where the domain label for paper folding and digit symbol had been changed to speed and fluid domains, respectively. Finally, we show the result obtained when randomly permuting the label assignments for each iteration, corresponding to chance performance. One can see that, for the proper assignment, the prediction success in the test sample is appreciably better than the diluted assignment, giving additional support to the concept of latent reference abilities with the proper assignment.

done simply by assigning the reference domain whose pattern score is higher; for instance, if a map’s pattern score with respect to the FR pattern is smaller than with respect to the PS pattern, we classify this map as “PS.” Predictive performance is recorded as classification accuracy by comparing the gold standard knowledge of the true labels with the predicted labels.

As Figure 4 shows, the prediction is successful, and the hit rates for both reference abilities are much better than chance level. Furthermore, to give additional credence to the reference ability concept, we swapped two tasks in their assignment to the reference abilities. The paper-folding task was mislabeled as PS, and the digit–symbol substitution task was mislabeled as FR.

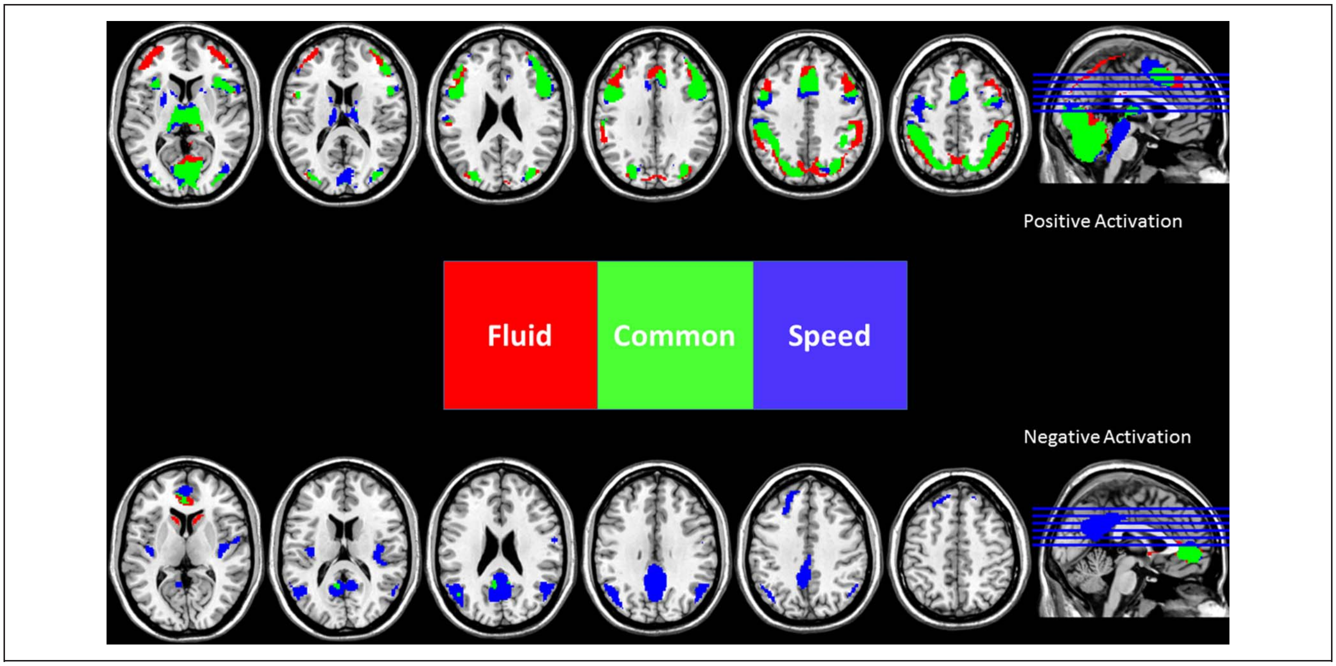

Figure 5. Select axial slices of the unique parts of the FR (red) and PS (blue) RANNs as well as common parts (green). Shown are areas of mean activation across the three tasks in each reference domain, which survive a significance threshold of $p < .05$, corrected for the number of voxels in our analysis. The upper row shows areas of activation, and the lower row shows areas of deactivation. One can appreciate that large parts of frontal, parietal, and occipital lobe activate in common, but only one frontal pole area deactivates in common.

These "diluted" reference domains were then used in a second split-half validation test in both training and test samples. Figure 4 shows for these latter reference domain assignments that the prediction in the test sample is still better than chance but substantially worse than for the correct reference domain assignment. This demonstrates that the assignment into reference domains as envisioned by Salthouse and Ferrer-Caja (2003) is optimal: An alternative grouping of tasks into reference domains not only produces worse construct validities but also hampers the prediction of reference domain labels in independent data. We checked all eight other possibilities of mislabeling one task in the reference ability assignment and found similarly poorer prediction performance in all of them.

Testing for Ability-specific Activation

Next, we turned to the topographic composition of the cognitive ability neural networks themselves. The representation of the neural networks associated with each of the two domains was derived by computing voxel-wise one-sample t contrasts obtained by averaging across all three tasks within both speed of processing and FR domains. The keys of common, significant activation for each domain are illustrated in Figure 5 and summarized in Tables 1–3.

For both networks, we found widespread bilateral parietal, frontal, temporal, and occipital activation, with additional frontal regions involved for the FR network. Inferior parietal activation has been found across a wide range of reasoning tasks (Hampshire et al., 2011; Woolgar et al., 2010; Kalbfleisch, Van Meter, & Zeffiro, 2007; Wright, Matlen, Baym, Ferrer, & Bunge, 2007; Lee et al., 2006; Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997) in the past, accompanied by lateral frontal activation as well. For PS, we found additional temporal, occipital, and thalamic areas. These areas are indicative of perceptual and working memory processing and have been retrieved before in neuroimaging explorations of the one constituent task of the PS domain that has been brought to the scanner repeatedly: the digit–symbol substitution task (Forn et al., 2009; Usui et al., 2009).

Table 1. Areas of Significant Activation and Deactivation that Are Unique for the FR Reference Domain

x	*y*	*x*	*lod(p)*	*AAL_label*
Positive activation				
−48	−48	46	17.5034	Parietal_Inf_L
−36	−60	44	15.767	Angular_L
52	−34	52	15.6967	Parietal_Inf_R
48	12	40	15.4119	Precentral_R
−38	54	2	15.1475	Frontal_Mid_L
34	−82	32	15.0633	Occipital_Mid_R
−4	−66	0	14.898	Lingual_L
34	56	2	14.623	Frontal_Mid_R
−48	16	36	14.6021	Frontal_Inf_Oper_L
−6	−62	54	14.5038	Precuneus_L
Negative activation				
−10	20	6	−14.6893	Caudate_L
12	42	−2	−14.2058	Frontal_Med_Orb_R
6	36	4	−14.0522	Cingulum_Ant_R
8	16	8	−13.5408	Caudate_R
−2	12	0	−10.6975	No AAL label
−28	10	−24	−10.4459	Temporal_Pole_Sup_L
−10	42	6	−10.0847	No AAL label
10	22	−2	−7.9925	Caudate_R

lod(p) denotes the sign-weighted decadic logarithm of the uncorrected p values, that is, a value of 3 corresponds to a positive correlation with $p < .001$, a value of −4 to a negative correlation with $p < .0001$, and so forth. For the correction by the number for voxels in the analysis, we have an uncorrected lod(p) with a magnitude of at least 7 with cluster size of >100 voxels. For positive activations, we only show the most significant 10 noncerebellar locations.

Conclusions

The observation that performance on multiple cognitive tests clusters into specific abilities that can be described by latent variables is long standing. The underlying assumption has been that these latent variables capture the common features of sets of tasks and are supported statistically but, by their nature latent abilities, are not observable. In this report, we sought neurobiological instantiation of these latent abilities. We showed that two sets of tests that have repeatedly been known to describe two different abilities, speed of processing and FR, have distinct neural substrates that replicate the logic of latent variable modeling: Each neural substrate is common to tests within a domain but is not applicable to the tests in the other domain. The brain areas associated with these two cognitive abilities are consistent with those reported in previous functional imaging studies of individual tasks.

We also demonstrated that the topographic similarity of the neural substrates underlying two tasks is related to the similarity of the performance on the two tasks. This observation directly links the commonality in cognitive performance to commonality of the underlying neural network. In summary, these observations empirically affirm that latent variables, which are unobservable, can be made manifest on the neural and behavioral level.

Our study offered a unique opportunity to observe whether task-related brain activation naturally clusters into some of the cognitive ability domains proposed by Salthouse and Ferrer-Caja. A stimulating recent debate (Haier et al., 2014) has highlighted the often ambiguous nature of solutions to multivariate data decompositions employed to discover latent variables. Multiple rotation schemes can be imposed, and to avoid any problems of "double dipping," attribution of neuroscientific meaning to any latent factors demands that the rationale rotation scheme has a strong a priori motivation, which is not confounded with the interpretation of the factor structure itself. Our study avoids such difficulties by employing Salthouse and Ferrer-Caja's assignment of tasks to reference abilities: After eliminating common areas of activation, the latent-ability neural networks were evident from direct observation of the patterns of activation of each task and their mutual correlation. Further validation was provided by split-sample analyses: For the preferred assignment of tasks into reference abilities according to Salthouse and Ferrer-Caja, the out-of-sample prediction of the reference ability label from a training data set was the best.

Table 2. Areas of Significant Activation (p(corrected) < .05) for the PS Reference Domain

x	*y*	*z*	*lod(p)*	*AAL_label*
Positive activation				
14	−86	−8	35.734	Lingual_R
−16	−84	−10	33.7132	Lingual_L
24	−84	−10	31.0423	Lingual_R
38	−70	−12	30.9997	Occipital_Inf_R
−14	−22	12	27.0871	Thalamus_L
32	30	0	26.9557	Insula_R
−42	−78	−2	26.0493	Occipital_Inf_L
38	−78	8	25.7111	Occipital_Mid_R
−40	−4	46	25.3444	Precentral_L
12	−92	10	23.6805	Cuneus_R
Negative activation				
−8	−54	30	−29.557	Precuneus_L
−50	−68	30	−29.2834	Angular_L
52	−64	32	−27.5507	Angular_R
10	−50	28	−26.5538	Precuneus_R
−8	−58	14	−20.0947	Precuneus_L
−6	−50	20	−18.1589	Precuneus_L
−36	−18	16	−14.2445	Insula_L
58	−60	14	−13.456	Temporal_Mid_R
40	−16	12	−12.6603	Insula_R
−16	42	46	−11.8694	Frontal_Sup_L

lod(p) denotes the sign-weighted decadic logarithm of the uncorrected p values, that is, a value of 3 corresponds to a positive correlation with $p < .001$, a value of −4 to a negative correlation with $p < .0001$, and so forth. For the correction by the number for voxels in the analysis, we have an uncorrected lod(p) with a magnitude of at least 7 with cluster size of >100 voxels. For both signs of activations, we only show the most significant 10 noncerebellar locations. "AAL_label" was obtained from the AAL template in the software package MRIcron.

Table 3. Areas of Significant Activation (p(corrected) < .05) Common for Both Domains

x	*y*	*z*	*lod(p)*	*AAL_label*
Positive activation				
18	−94	−4	29.3483	Calcarine_R
−40	−40	40	28.358	Parietal_Inf_L
−2	−82	−10	28.0189	Calcarine_L
−44	−66	−14	26.9939	Fusiform_L
−24	−66	40	26.512	Parietal_Sup_L
32	22	4	24.7064	Insula_R
30	−90	−8	24.6746	Occipital_Inf_R
−38	−80	−8	24.5743	Occipital_Inf_L
46	10	26	24.4257	Frontal_Inf_Oper_R
−16	−96	−8	24.2717	Occipital_Inf_L
Negative activation				
0	54	−6	−16.8982	Frontal_Med_Orb_L
0	36	−4	−14.7772	Cingulum_Ant_R

lod(p) denotes the sign-weighted decadic logarithm of the uncorrected p values, that is, a value of 3 corresponds to a positive correlation with $p < .001$, a value of −4 to a negative correlation with $p < .0001$, and so forth. For the correction by the number for voxels in the analysis we have an uncorrected lod(p) with a magnitude of at least 7 with cluster size of >100 voxels. For both signs of activations, we only show the most significant 10 noncerebellar locations. "AAL_label" was obtained from the AAL template in the software package MRIcron.

Age-related differences were not the main interest of this study; however, for completeness, we tested our CV measures for behavior and mean task activation patterns for relationships with age by way of permutation tests along a median age split with 1,000 iterations. Neither the topographic nor behavioral CV showed any difference as a function of age group.

Cognitive psychologists have long recognized that latent variables provide a more accurate description of an underlying cognitive ability than any single test. Similarly, we propose that these common neural networks can better capture the neural substrates of a cognitive ability than imaging observations based on a single task. This assumption departs from traditional cognitive neuroscience research, which focuses on discovering the neural concomitants of single tasks. The identification of latent ability neural networks can have great practical utility. For example, it has been observed many times that performance on the two cognitive abilities addressed in this report declines with aging. However, as we demonstrated here, both the structure of the latent variables for these constructs and the underlying neural networks remained the same: No significant differences along a median age split could be observed. However, with substantially boosted numbers of observations, the possible dependence of the latent ability networks' CV on age can be elucidated more completely, giving insight into the mechanisms of normal cognitive aging and possibly providing diagnostic utility for identifying pathological cognitive aging. Low behavioral performance scores might, in themselves, not yet constitute pathological aging if unaccompanied by a substantial reduction in the integrity and distinctiveness of latent ability neural networks. CV at the neural level, which, in contrast to behavioral performance, can be quantified for a single participant, could be an informative additional metric by which to judge participants' cognitive profiles. Our current report constitutes a first step in the better understanding of the neural underpinnings of cognitive aging and intelligence in general.

Acknowledgments

The authors thank two anonymous reviewers and gratefully acknowledge funding support through grant NIH/NIA R01 AG038465 - Exploring Cognitive Aging Using Reference Ability Neural Networks (PI: Stern)

Reprint requests should be sent to Christian Habeck, Taub Institute, Columbia University, 630 W 168th St., Box 16, New York, NY 10032, or via e-mail: ch629@columbia.edu.

REFERENCES

Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure & Function, 219,* 485–494.

Bottorf, E. A. (1946). A study comparing art abilities and general intelligence of college students. *Journal of Education & Psychology, 37,* 398–426.

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues in Clinical Neuroscience, 12,* 489–501.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Embree, R. B., Jr. (1946). Note on the estimation of college aptitude test scores from IQ's derived from group intelligence tests. *Journal of Education & Psychology, 37,* 502–504.

Forn, C., Belloch, V., Bustamante, J. C., Garbin, G., Parcet-Ibars, M. A., Sanjuan, A., et al. (2009). A symbol digit modalities test version suitable for functional MRI studies. *Neuroscience Letters, 456,* 11–14.

Haier, R. J., Karama, S., Colom, R., Jung, R., & Johnson, W. (2014). A comment on "fractionating intelligence" and the peer review process. *Intelligence, 46,* 323–332.

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating human intelligence. *Neuron, 76,* 1225–1237.

Hampshire, A., Thompson, R., Duncan, J., & Owen, A. M. (2011). Lateral prefrontal cortex subregions make dissociable contributions during fluid reasoning. *Cerebral Cortex, 21,* 1–10.

Kalbfleisch, M. L., Van Meter, J. W., & Zeffiro, T. A. (2007). The influences of task difficulty and response correctness on neural systems supporting fluid reasoning. *Cognitive Neurodynamics, 1,* 71–84.

Kendler, H. H. (1946). On the meaning of intelligence. *Bulletin of Military Clinical Psychologists, 1,* 48–52.

Lee, K. H., Choi, Y. Y., Gray, J. R., Cho, S. H., Chae, J. H., Lee, S., et al. (2006). Neural correlates of superior intelligence: Stronger recruitment of posterior parietal cortex. *NeuroImage, 29,* 578–586.

Marquart, D. I., & Bailey, L. L. (1955). An evaluation of the culture free test of intelligence. *Journal of Genetic Psychology, 86,* 353–358.

Prabhakaran, V., Smith, J. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive Psychology, 33,* 43–63.

Preusse, F., van der Meer, E., Deshpande, G., Krueger, F., & Wartenburger, I. (2011). Fluid intelligence allows flexible recruitment of the parieto-frontal network in analogical reasoning. *Frontiers in Human Neuroscience, 5,* 22.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory? *Developmental Psychology, 27,* 763–776.

Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging, 18,* 91–110.

Thomson, G. (1946). The trend of national intelligence. *The Eugenics Review, 38,* 9–18.

Usui, N., Haji, T., Maruyama, M., Katsuyama, N., Uchida, S., Hozawa, A., et al. (2009). Cortical areas related to performance of WAIS Digit Symbol Test: A functional imaging study. *Neuroscience Letters, 463,* 1–5.

Willoughby, R. R. (1928). The survival of intelligence. *Proceedings of the National Academy of Sciences, U.S.A., 14,* 892–894.

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., et al. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 107,* 14899–14902.

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2007). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience, 1,* 8.