Reference ability neural networks and behavioral performance across the adult life span

Christian Habeck [*], Teal Eich, Ray Razlighi, Yunglin Gazes, Yaakov Stern

Cognitive Neuroscience Division, Department of Neurology, Columbia University, New York, NY, 10032, United States

A B S T R A C T

To better understand the impact of aging, along with other demographic and brain health variables, on the neural networks that support different aspects of cognitive performance, we applied a brute-force search technique based on Principal Components Analysis to derive 4 corresponding spatial covariance patterns (termed Reference Ability Neural Networks –RANNs) from a large sample of participants across the age range. 255 clinically healthy, community-dwelling adults, aged 20–77, underwent fMRI while performing 12 tasks, 3 tasks for each of the following cognitive reference abilities: Episodic Memory, Reasoning, Perceptual Speed, and Vocabulary. The derived RANNs (1) showed selective activation to their specific cognitive domain and (2) correlated with behavioral performance. Quasi out-of-sample replication with Monte-Carlo 5-fold cross validation was built into our approach, and all patterns indicated their corresponding reference ability and predicted performance in held-out data to a degree significantly greater than chance level. RANN-pattern expression for Episodic Memory, Reasoning and Vocabulary were associated selectively with age, while the pattern for Perceptual Speed showed no such age-related influences. For each participant we also looked at residual activity unaccounted for by the RANN-pattern derived for the cognitive reference ability. Higher residual activity was associated with poorer brain-structural health and older age, but –apart from Vocabulary-not with cognitive performance, indicating that older participants with worse brain-structural health might recruit alternative neural resources to maintain performance levels.

Introduction

Cognitive aging can be described parsimoniously by a set of four reference abilities – Episodic Memory, Reasoning, Perceptual Speed, and Vocabulary – that serve as the "primitive types" of cognition in general (Salthouse and Ferrer-Caja, 2003). Our group has recently extended this line of research (Habeck et al., 2016; Stern et al., 2014) by collecting functional imaging data on a battery of tasks that tap each of the four reference abilities across the adult life span, in order to determine their neural correlates, i.e. Reference Ability Neural Networks (RANNs). Using a multivariate technique that married Principal Components Analysis (PCA) and Linear-Indicator Regression (Hastie et al., 2009), we previously derived spatial activation patterns that accurately classified the reference ability underlying each activation task. Indeed, even when these patterns were derived only in people below age 30, out-of-sample task classification performance of the four RANNs in people older than 30 was high, and did not decline with age, suggesting that these RANNs are age-invariant.

In this previous specification of the RANNs to their underlying reference ability, one crucial aspect that was not accounted for was behavioral performance: ideally, RANNs should not only be specific to the reference ability of the underlying cognitive process, but also account for behavioral performance. To reconcile this issue, we extended our previous findings, analyzing the data from 255 20–77 year old adults who underwent fMRI while performing three tasks for each of the four cognitive reference abilities, i.e. 12 tasks in total. In the current report, we combined PCA with a brute-force search that sought to maximize *both* the brain-behavioral correlation of the derived RANNs *and* cognitive specificity to the reference domain in question. Statistical inference was performed by resampling, and testing the prediction of cognitive process and behavioral performance in held-out data. This approach allowed us to investigate how well the RANNs could account for performance and cognitive specificity, and to understand how changes in performance concomitant with aging are reflected in these RANNs.

We stress that the current report is *not* methodological in focus: the approach we chose is conceptually simple, but algorithmically somewhat involved, lengthy and inelegant. We are not bringing a prime-time ready technique to the field that should be widely disseminated; rather, our focus here is on the results of our approach, i.e. the derived networks, their association with behavioral performance, demographics and brain structure across the adult life span.

---

* Corresponding author.
 E-mail address: ch629@columbia.edu (C. Habeck).

Methods

Participant sample and demographics

Analyses included data from 255 strongly right-handed, native English speaking healthy adults. Participants were recruited via random-market-mailing, and screened for MRI contraindications and hearing or visual impairment that would impede testing. Older adult participants were additionally screened for dementia and mild cognitive impairment prior to participating in the study, and participants who met criteria for either were excluded. Apart from these obvious cognitive exclusion criteria, we had a host of other health-related exclusion criteria including: myocardial infarction, congestive heart failure or any other heart disease, brain disorder such as stroke, tumor, infection, epilepsy, multiple sclerosis, degenerative diseases, head injury (loss of consciousness > 5 mins), mental retardation, seizure, Parkinson's disease, Huntington's disease, normal pressure hydrocephalus, essential/familial tremor, Down Syndrome, HIV Infection or AIDS diagnosis, learning disability/dyslexia, ADHD or ADD, uncontrolled hypertension, uncontrolled diabetes mellitus, uncontrolled thyroid or other endocrine disease, uncorrectable vision, color blindness, uncorrectable hearing and implant, pregnancy, lactating, any medication targeting central nervous system, cancer within last five years, Renal insufficiency, untreated neurosyphilis, any alcohol and drug abuse within last 12 months, recent non-skin neoplastic disease or melanoma, active hepatic disease, insulin dependent diabetes, any history of psychosis or ECT, recent (past five years) major depressive, bipolar, or anxiety disorder, objective cognitive impairment (dementia rating scale of <130), and subjective functional impairment (BFAS > 1). The prevalence of medication for hypertension, diabetes, and high cholesterol is as follows, respectively: 18%, 14%, and 7%. This compares favorably with CDC statistics for the adult US population at large (33.5%, 12.6%, and 12.1%, from www.cdc.gov/nchs/fastats). A complete description of the participants in terms of demographics and cortical thickness can be found in Table 1.

Procedure

FMRI data was acquired as participants performed 12 cognitive tasks, pertaining to the four reference abilities (Stern et al., 2014). In the remainder of the manuscript, we occasionally use the following short-hand notation for the reference abilities: episodic memory — MEM or just "Memory", reasoning — REASON, perceptual speed — SPEED or just "Speed", and vocabulary — VOCAB. We will refer to the Reference Ability Neural Networks (RANNs) as activation patterns, for brevity. Further, "cognitive reference ability" and "cognitive reference domain" will be used interchangeably.

Tasks were administered over the course of two 2-h scanning sessions, with six tasks administered in each scanning-session. One session presented three VOCAB tasks and three SPEED tasks interspersed in a fixed order: Synonyms, Digit-Symbol, Antonyms, Letter Comparison, Picture Naming, and Pattern Comparison; and the other session presented three MEM tasks and three REASON tasks, also interspersed in a fixed order: Logical Memory, Paper Folding, Word Order Recognition, Matrix Reasoning, Paired Associates, and Letter Sets. The order of tasks within session was not varied, but the order of the two sessions was counterbalanced across participants. Prior to each scan session, computerized

training was administered for the six tasks to be administered during that session. At the completion of training for each task, participants had the option of repeating the training. For all tasks except Picture Naming, responses were differential button presses. During training, responses were on the computer keyboard. During scans, they were made on the LUMItouch response system (Photon Control Company).

Stimulus presentation

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 Safe-Vision, LLC. Webster Groves, MO). Task administration and collection of reaction time (RT) and accuracy data were controlled by EPrime running on a PC computer. Task onset was electronically synchronized with the MRI acquisition computer.

Reference ability tasks

VOCAB tasks. The primary dependent variable for all VOCAB tasks was the proportion of correct responses.

Synonyms (Salthouse, 1993): Participants were instructed to match a given probe word to its synonym or to the word most similar in meaning. The probe word was presented in all capital letters at the top of the screen, and four numbered choices were presented below. Participants indicated which choice was correct.

Antonyms (Salthouse, 1993): Participants matched a given word to its antonym, or to the word most different in meaning. The probe word was presented in all capital letters at the top of the screen, and four numbered choices were presented below. Participants indicated which choice was correct.

Picture Naming: Participants verbally named pictures, adapted from the picture naming task of the WJ-R Psycho-Educational battery (Salthouse, 1998; Woodcock et al., 1989).

SPEED tasks. As accuracy for all three SPEED tasks was high, the primary dependent variable was reaction time (RT). For all tasks, participants were instructed to respond as quickly and accurately as possible.

Digit Symbol: A code table was presented on the top of the screen, consisting of nine number (ranging in value from 1 to 9)-symbol pairs. Below the code table, an individual number/symbol pair was presented. Participants indicated whether the individual pair was the same as that in the code table.

Letter Comparison (Salthouse and Babcock, 1991): Two strings of letters, each consisting of three to five letters, were presented alongside one another. Participants indicated whether the letter-strings were the same or different.

Pattern Comparison (Salthouse and Babcock, 1991): Two figures, consisting of varying numbers of lines connecting at different angles, were presented alongside one another. Participants indicated whether the figures were the same or different.

REASON tasks. The primary dependent variable for the REASON tasks was proportion of correct trials.

Paper Folding (Ekstrom et al., 1976): Participants selected which of

Table 1

Participant sample and demographics. Cortical thickness has strong negative linear age trends, p < .0001.

	<30	30–39	40–49	50–59	60–69	>70
N	35	44	38	40	62	36
NART-IQ	113 ± 9	112 ± 9	114 ± 9	115 ± 8	118 ± 9	120 ± 10
Education	15.3 ± 2.3	16.4 ± 2.5	15.9 ± 2.6	15.5 ± 2.3	16.0 ± 2.5	17.3 ± 2.5
Sex	23 F, 12 M	27 F, 17 M	17 F, 21 M	20 F, 20 M	32 F, 30 M	18 F, 18 M
DRS	140.4 ± 2.7	139.8 ± 2.3	139.1 ± 2.9	140.1 ± 3.3	139.7 ± 3.1	139.7 ± 3.0
Mean Cortical thickness	2.69 ± 0.11	2.66 ± 0.09	2.65 ± 0.09	2.59 ± 0.08	2.55 ± 0.11	2.51 ± 0.12

six options best represented the pattern of holes that would result from a sequence of folds in a piece of paper through which a hole was punched. The sequence was given on the top of the screen, and the six options were given across two rows (three options in each row) below. Response consisted of pressing one of six buttons corresponding to the chosen solution.

Matrix Reasoning (adapted from Raven (1962))**:** Participants were given a matrix that was divided into nine cells, in which the figure in the bottom right cell was missing. Participants were instructed to evaluate which of eight figure choices, presented below the matrix, would best complete the missing cell.

Letter Sets (Ekstrom et al., 1976**):** Participants were presented with five sets of letters, where four out of the five sets had a common rule (i.e. they have no vowels), with one of the sets not following this rule. Participants identified the unique set.

MEM tasks. The primary dependent variable for the MEM tasks was the proportion of correctly answered questions.

Logical Memory: Participants were asked to answer detailed multiple-choice questions about a story presented on the computer screen, with four possible answer choices.

Word Order Recognition: Participants were presented with 12 words sequentially, and were instructed to remember the order in which the words were presented. Following the word list, they were given a probe word at the top of the screen, and four additional word choices below. They were instructed to choose out of the four options the word that immediately followed the word given above in the word list.

Paired Associates: Participants were instructed to remember pairs of words presented sequentially on the screen. Following presentation of the pairs, participants were given a probe word at the top of the screen and four additional word choices below. Participants were asked to choose the word that was originally paired with the probe word.

Behavioral performance variables

Behavioral performance was recorded while participants executed each task in the scanner. When proportion of correct trials was calculated, trials where the participant did not respond (i.e. timed out) were not considered in the calculation. For the SPEED tasks, reaction time was reversed in sign, such that an increasing value of the behavioral score implied better performance.

Image acquisition procedures

All MR images were acquired on a 3.0 Tesla Philips Achieva Magnet. There were two, 2-h MR imaging sessions to accommodate the 12 fMRI activation tasks as well as the additional imaging modalities, described below. At each session, a scout, T1-weighted image was acquired to determine participant position. Participants underwent a T1-weighted MPRAGE scan to determine brain structure, with a TE/TR of 3/6.5 ms and Flip Angle of 8°, in-plane resolution of 256 × 256, field of view of 25.4 × 25.4 cm, and 165–180 slices in axial direction with slice-thickness/gap of 1/0 mm. All scans used a 240 mm field of view. For the EPI acquisition, the parameters were: TE/TR (ms) 20/2000; Flip Angle 72°; In-plane resolution (voxels) 112 × 112; Slice thickness/gap (mm) 3/0; Slices 41. In addition, MPRAGE, FLAIR, DTI, ASL, and a 7-min resting BOLD scan were acquired. A neuroradiologist reviewed each participant's scans. Any significant findings were conveyed to the participant's primary care physician.

Image-processing procedures

Structural Imaging - Two structural indices were included in our calculations: (1) gray matter volume and (2) thickness by region of interest (ROI).

Each participant's structural T1 scans were reconstructed using FreeSurfer v5.1 (http://surfer.nmr.mgh.harvard.edu/). The accuracy of FreeSurfer's subcortical segmentation and cortical parcellation (Fischl et al., 2002, 2004) has been reported to be comparable to manual labeling. Each participant's white and gray matter boundaries, as well as gray matter and cerebral-spinal-fluid boundaries, were visually inspected slice by slice, and manual control points were added in the case of any visible discrepancy. Boundary reconstruction was repeated until we reached satisfactory results for every participant. The subcortical structure borders were plotted by *freeview* visualization tools and compared against the actual brain regions. In the case of discrepancy, they were corrected manually. Finally, we computed mean values for 68 cortical ROIs for cortical thickness and cortical volume for each participant to be used in group-level analyses.

Functional neuroimaging - Each participant's 12 task-activation fMRI scans were pre-processed in FSL (Smith et al., 2004) using the following steps: (1) within-participant histogram computation for each participant volume to identify noise (FEAT); (2) participant-motion correction (MCFLIRT); (3) slice-timing correction; (4) brain-mask creation from first volume in participant's fMRI data; (5) high-pass filtering (T = 128 s); (6) pre-whitening; (7) General-Linear-Model (GLM) estimation with equally temporally filtered regressors and double-gamma hemodynamic response functions; and (8) registration of functional and structural images with subsequent normalization into MNI space (FNIRT).

General linear models (GLM) for each participant and each task consisted of block-based time-series analysis for SPEED and VOCAB tasks, and event-related modeling for REASON and MEM tasks. For MEM, only the recognition phase of the trial was analyzed. Contrary to usual voxel-wise FSL practice, we obtained standardized contrast images of task > fixation for every participant and task to perform group-level multivariate analysis (described in the next section). Other than for MEM, there was no separation of stimulus presentation or behavioral response in our task design. For all fitted neural responses, whether event-related or block-designed, incorrect and correct responses were *not* separated, but analyzed indiscriminately, in the GLM-fitting process.

Brute-force search pattern derivation with optimization of indicator and behavioral correlation

We repeat that the main focus of our paper was *not* methodological. This means that the techniques we used to derive activation patterns were not optimized for elegance or computational speed, neither were they intended to be disseminated for widespread use; rather, they were conceptually simple and with repeated executions to deal with inherent stochastic variability. To explore and maximize the relationship between RANN pattern scores and behavioral performance, a simple brute-force computational approach was adopted. This approach is slower and less elegant than alternative parametric approaches, like regularized least squares, but it pays equal attention to both cognitive specificity and brain-behavioral correlation, rather than maximizing one aspect while treating the other merely as a boundary condition. Conceptually, our approach is very simple. Fig. 1 below gives a schematic birds-eye view of our approach.

Fig. 1 illustrates the goals of our multivariate technique for deriving an activation pattern that is both specific to one particular reference domain (in the figure the example is shown for memory), and correlates with performance within that reference domain. Such an activation pattern is only derived from the task-maps pertaining to the reference domain of interest; however, for validating the specificity of the pattern to this reference domain, the pattern has to be prospectively applied to the task-maps from all other domains too. Pattern scores in the task-maps of these other reference domains should be significantly lower than for the reference domain of interest. We will refer to the fulfillment of this requirement as "indicator correlation" for the remainder of the manuscript.

We give the full algorithmic recipe below, and occasionally list the linear algebra involved in Matlab notation. The parameters listed below were chosen to ensure that (1) resampling non-parametric tests for the
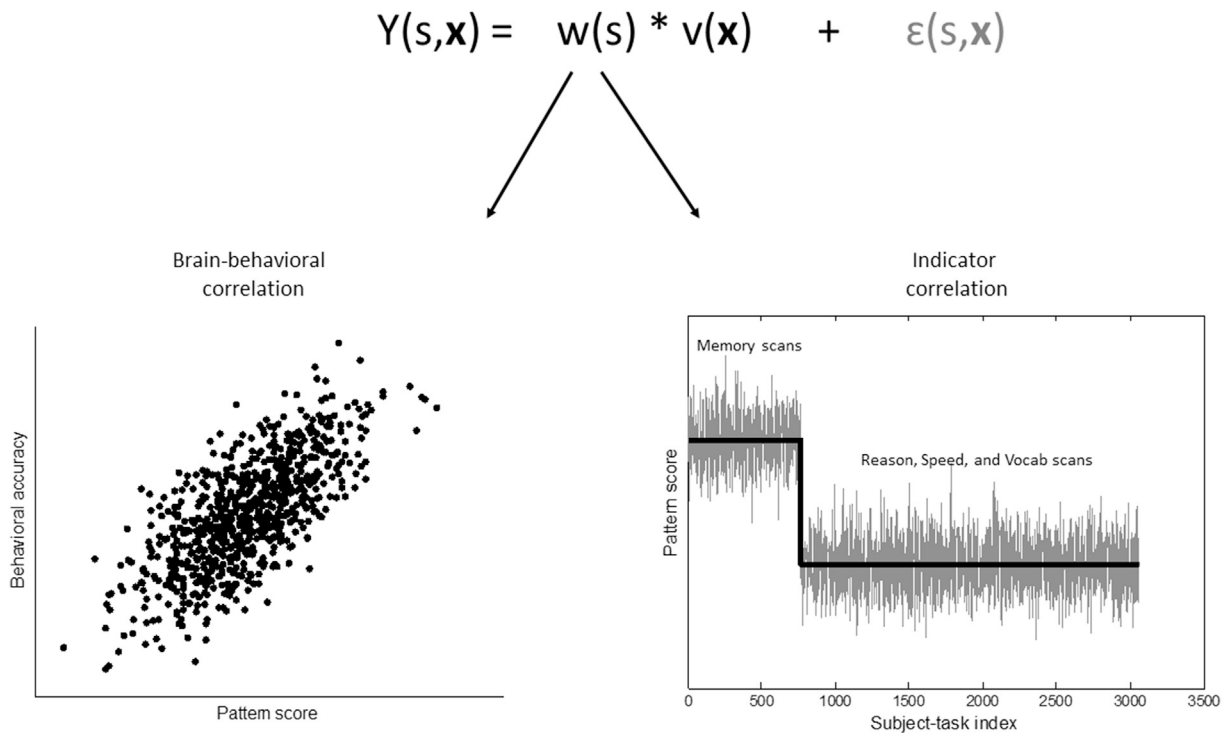
$$Y(s,\mathbf{x}) = w(s) * v(\mathbf{x}) + \varepsilon(s,\mathbf{x})$$

Fig. 1. Main illustration of the goals of our analytic technique with mock data. The top equation restates the generic form for any multivariate decomposition of a data array $Y(s,\mathbf{x})$ which captures all subject and task information (s) in the pattern score $w(s)$, while the group-invariant pattern $v(\mathbf{x})$ contains all spatial information with regard to voxel location $\mathbf{x}$, while irrelevant information is discarded into the residual error term $\varepsilon(s,\mathbf{x})$. For the pattern score $w(s)$ the reader can now appreciate the purpose of the optimization: simultaneous maximization of the brain-behavioral correlation in the reference domain of interest (left lower panel), and the indicator correlation across all task maps (right lower panel). The example shown in this schematic is for the Memory domain. Our technique imposes no prior spatial constraints on the pattern $v(\mathbf{x})$.

point-estimate computation could be performed in reasonable execution times, and (2) there would be sufficient reproducibility of the derived topographic images for the point estimate. Because the method we used has inherent stochastic variability even when executed on identical data, we demanded that spatial correlations of covariance patterns derived from identical data should exceed R = 0.95 for repeated execution of the search routine. We give a listing of pseudo-code that demonstrate the algorithm for the Memory domain. For the other domains, it would be executed similarly.

1. Collect all Memory task maps, $N =$ number of maps; ($N = 3 \times 200 = 600$ for our purposes since the derivation is only executed in a training sample of 200 of the participants)
2. Perform PCA on these N maps;
3. Retain all PCs with an Eigen value that is bigger than 1/600 of the sum of all Eigen values, i.e. 0.17% of the total variance; number of retained PCs = P;
4. Compute pattern scores (cf. section Pattern expression and residual activation) of these P PCs in ALL task maps, even the ones *not* belonging to the MEMORY domain;
5. Execute "inner loop" with 1000 iterations;
   Inner loop:
   a. Assign a "best estimate pattern" of zeros as voxel loadings, and with brain-behavioral and indicator-correlation values of zero; this is just a placeholder for initialization and will be overwritten in the course of the inner loop;
   b. Generate P random coefficients for P retained PCs;
   c. Compute the linear-combination pattern with these P coefficients by using them as weights in a weighted sum; the computation is executed according to **pattern = v(:,1:P) *b**, where **pattern** and **v** have as many rows as voxels, and **b** is the column vector of the P random coefficients;

   d. Compute the pattern score of the linear-combination pattern in all task maps (see section Pattern expression and residual activation. for an explanation of pattern score computation), not only the Memory task maps since all task maps are needed for the indicator correlation;
   e. Compute the brain-behavioral correlation of the pattern score and performance variable for the Memory task maps;
   f. Compute the indicator correlation of the pattern score across all maps, i.e. correlate the pattern score in all maps with an indicator variable that assigns values of 1 to all Memory maps, but zero to all non-Memory maps;
   g. Check if both indicator and behavioral correlations are higher than the values of best current estimate; if so, retain the current pattern as the best current estimate with its correlation values, and discard the previously stored values;
   h. Repeat steps b.-g. with 200 repetitions;
   i. Outcome of inner loop: one activation pattern with a corresponding correlation strength, which is the average of the associated indicator and brain-behavioral correlations;
      ➔Outcome of the execution of 1000 inner loops: 1000 activation patterns with corresponding correlation strengths
6. Compute a weighted average of the 1000 patterns, where the correlation strengths serve as weights;
   ➔End result: one single activation pattern for the Memory domain

Fig. 2 shows a mock display of data resulting from the execution for the inner loop, where 100 PCs were retained at the start.

The figure displays the 100 random PC weights for each of the 200 iterations, the behavioral and indicator correlation. In the example, iterations 4, 46, 73, and 164 resulted in updates of the best-estimate pattern. For this run of the inner loop, the final pattern retained would be the pattern obtained in iteration 164.
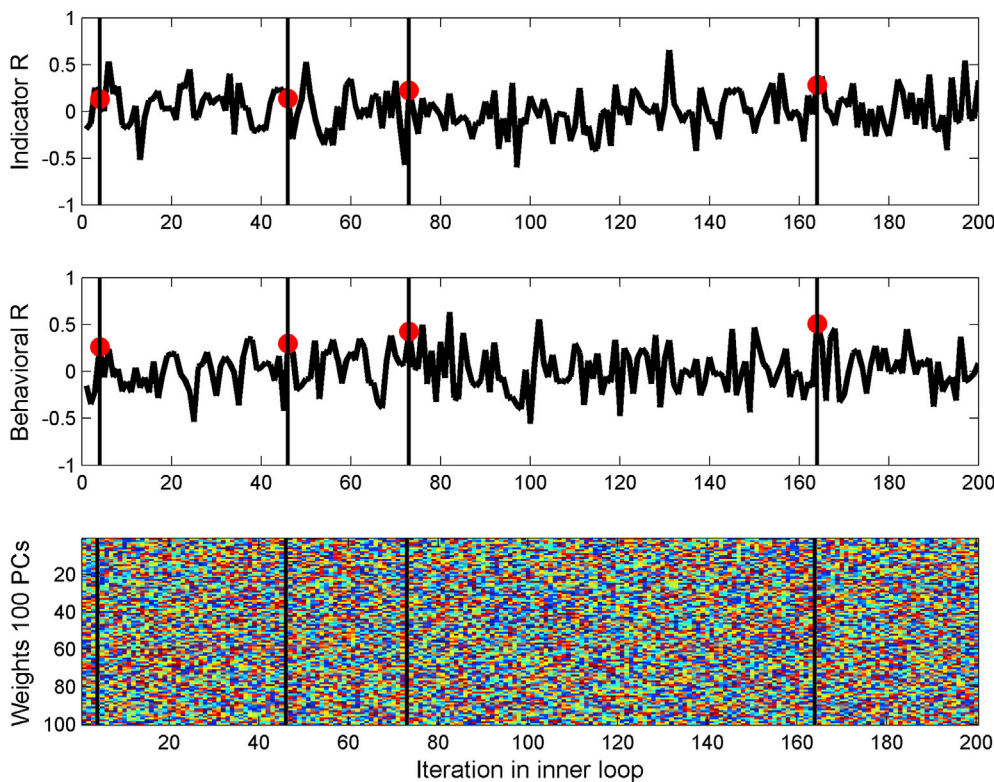
Fig. 2. Illustration of the 'inner loop' in the pattern derivation process with mock data. Execution of the inner loop supposes that PCA was performed on the data and 100 PCs have been retained according to the 'Eigen value > 1' criterion. Now 200 random co-efficients for the PCs are generated and the corresponding linear-combination pattern is computed. Next we compute the indicator correlation, i.e. selectivity of the pattern scores to the reference ability in question, and the brain-behavioral correlation of the pattern scores. If both values exceed the values of the previous best estimate, we accept the pattern as the current best estimate. In the figure below, the indicator correlation is plotted in the top pane, and the brain-behavioral correlation in the middle pane, while the corresponding randomly generated 100-component weight vectors are plotted in the bottom pane. Simultaneous best estimates are found sequentially at iterations 4, 46, 73, and 164, marked with vertical lines and red dots in the correlation values. The estimate at iteration 164 is taken as the final best estimate for this inner-loop execution.

Pattern expression and residual activation

Once a covariance pattern has been derived, it can be projected onto any data set, for computation of the pattern-expression scores. Such scores can then be used to validate brain-behavioral correlations in quasi replication samples (see section Split sample simulations and statistical inference on pattern loadings). If we assume pattern v is a column vector

with as many rows as voxels, and matrix $\mathbf{Y}$ is a data set with as many rows as voxels and as many columns as observations, the pattern score can be computed

$$\mathbf{w} = \mathbf{Y}' \, \mathbf{v}$$

where the prime symbol ' indicates matrix transposition, and the pattern

Fig. 3. Illustration of pattern w and residual activation *ress* for a mock example of 2-dimensional brain maps. Each map can be represented by a point in a two dimensional space. The two maps could denote the same subjects, but different tasks, or different tasks and subjects. Computing the pattern score mounts to performing the inner (or 'dot') product between the brain maps and the pattern v. Geometrically, this is a simple orthogonal projection. In the example shown, the pattern score is the same for both brain map 1 and 2, but the residual activation is different. Map 1 has larger residual activation, i.e. both voxel 1 and 2 have larger parts unaccounted for by the covariance pattern v than for map 2.

score vector **w** is a column vector with as many rows as observations.

In addition to quantifying with a pattern score to what degree a brain map expresses pattern **v**, we can also quantify the residual activation in any brain map *not* accounted for by **v**. The residual activation with respect to any covariance pattern **v** can be computed according to

$$R = Y (1 - v \ v')$$

where **1** denotes the voxel-by-voxel unit matrix, and **R** accordingly has the same data format as **Y**. We squared the residual matrix and average within map across observations to obtain a mean Residual-Sum-Of-Squares (RESS) vector. In Matlab notation, this would be written as

$$ress = mean(R.2,1)';$$

Ress has the same format as **w**, i.e. it is a column vector with as many rows as observations, which are tasks and subjects in this case, i.e. *not* voxels. In our regression analyses, the three *ress* values for a single participant in each reference ability were averaged across the three tasks for that reference ability to produce one scalar value per reference ability per participant.

Fig. 3 illustrates pattern score and residual activation for a sample case of two voxels. However, the computation readily generalizes to any higher dimension.

In addition to illustrating how pattern and residual activation scores are derived, Fig. 3 further shows that *a priori* there is no relationship that is forced upon these two quantities. Across participants, they can be found to be negatively correlated, positively correlated, or uncorrelated. In particular, a higher pattern score for any RANN-pattern in an individual does *not* necessitate a lower residual activation score; such a negative relationship would imply that the total amount of activation in any task map is conserved *across participants*, which is not true in the absence of particular normalizations.

Split sample simulations and statistical inference on pattern loadings

To ensure inferential robustness of the RANNs, we applied our search technique repeatedly to resampled data for all four reference abilities, using a Monte-Carlo 5-fold cross-validation in which 80% of the data is used as a training set and 20% of the data is used as the test set. Under this procedure, training and test sets are non-overlapping within each iteration (but are overlapping between different iterations). Specifically, we randomly split our data repeatedly (=1000 times) into a derivation sample of 200 participants and a replication sample of 55 participants. Within each iteration, all task images for the 200 participants in the training sample (12 × 200 = 2400 in total) were used to derive the four RANNs, and all task maps for the 55 participants in the replication sample (55 × 12 = 660 in total) were used to test the out-of-sample generalization. Age was not used as covariate for the pattern derivation. Brute-force optimization (see section Brute-force search pattern derivation) was performed in the 200-participant derivation sample, and four RANNs were derived, leading to 1000 estimates for all four reference abilities. The expressions of the derived patterns in the replication sample were then correlated (1) with the particular indicator variable for the corresponding reference ability and (2) with behavioral performance. These out-of-sample correlations were computed for descriptive purposes and to test whether our pattern-derivation technique yielded not only reliable pattern estimates, but also reliable brain-behavioral correlations. Out-of-sample correlation was quantified with the lodP value according to

$$lodP = -sign(R) * log10(P)$$

where *R* is the correlation, either with the behavioral variable in the tasks pertaining to the reference abilities in question, or with the indicator variable for that reference ability across all tasks, and *P* is the p-value associated with this correlation. Both *lodP(behavior)* as well as *lodP(indicator)* were computed. LodP was used because it enables better visualization than the p-level alone, and it incorporates the sign of the

correlation. (Low p-levels with the wrong sign of the association would not qualify as a satisfactory out-of-sample prediction; the lodP statistic therefore distinguishes the sign of the correlation and provides the level of significance in one statistic.) We stress that our split-sample simulations, technically described as Monte-Carlo 5-fold cross validation, cannot be regarded as true replications since our derivation and replication samples have overlap between iterations, and each derived pattern and out-of-sample prediction does not originate from truly independent data samples. Our simulations are thus probably under-estimating the true out-of-sample prediction error; however, for relative statements and comparing the reference domains, our methods are still suitable. The term "replication" for the remainder of the manuscript should be understood with this caveat in mind.

To compute the robustness of pattern loadings, the mean and standard deviation of the loadings across all 1000 estimates for each RANN were computed. From this, a Z-value was computed as the inverse coefficient of variation for every voxel loading:

$$Z (loading) = mean \ value \ (loading) / STD \ (loading)$$

Bootstrap loadings in all four activation patterns were thresholded according to $|Z| > 4.75$, which under a standard-normal assumption corresponds to a two-tailed *p*-level of 0.05, where the *p*-level has been Bonferroni corrected for the number of voxels in the analysis (=24,596). Local maxima that were more than three voxels (=9 mm) apart and a cluster-size threshold of 100 were computed. For visualization, we adopted a more lenient threshold of $|Z| > 3$ and cluster-size > 50. The Automated Anatomical Labeling (AAL) template provided by the software package MRIcron (www.mricron.com by Chris Rorden, Version 11) was used to locate and annotate local maxima.

Relation of pattern scores, residual-activity scores, demographics and brain structure

After the derivation of the four RANN-patterns, we sought to derive full bivariate correlations and partial correlations of several quantities, separately for each reference ability. The quantities were: (1) behavioral performance, (2) RANN-pattern score, (3) residual activity, (4) age, (5) education, (6) NART-IQ, (7) mean cortical thickness, and (8) mean cortical volume.

Results

Topographic composition and replication of the reference-ability neural networks

Topographic composition of the four RANNs, derived according to the steps described in the Methods section, is shown in Fig. 4.

The quasi-replication performance regarding both brain-behavioral and indicator correlations out of sample is shown in Fig. 5. In total we ran 1000 iterations of our routine. The indicator correlation was significant for every iteration; the brain-behavioral correlation, on the other hand, was significant at $p < .01$: 100% of the time for MEM, 99% of the time for REASON, 43% of time for SPEED, and 75% of the time for VOCAB. Despite these variations, it is important to note that all of these values are highly significant: Chance replication at $p < .01$ would be expected only 1% of the time. Thus, even for the worst performing RANN (SPEED), chance performance is exceeded more than 40-fold.

By design, all of the RANN patterns correlated positively with behavioral performance, such that as performance increases (=gets better), regions with positive loadings increase in activation and regions with negative loadings decrease in activation.

All RANN patterns showed regions with both positive and negative loadings. The MEM pattern consisted of positive loadings in the right supplementary motor area, and a widespread bilateral pattern of frontal, occipital and parietal areas; negative loadings were lateralized to the
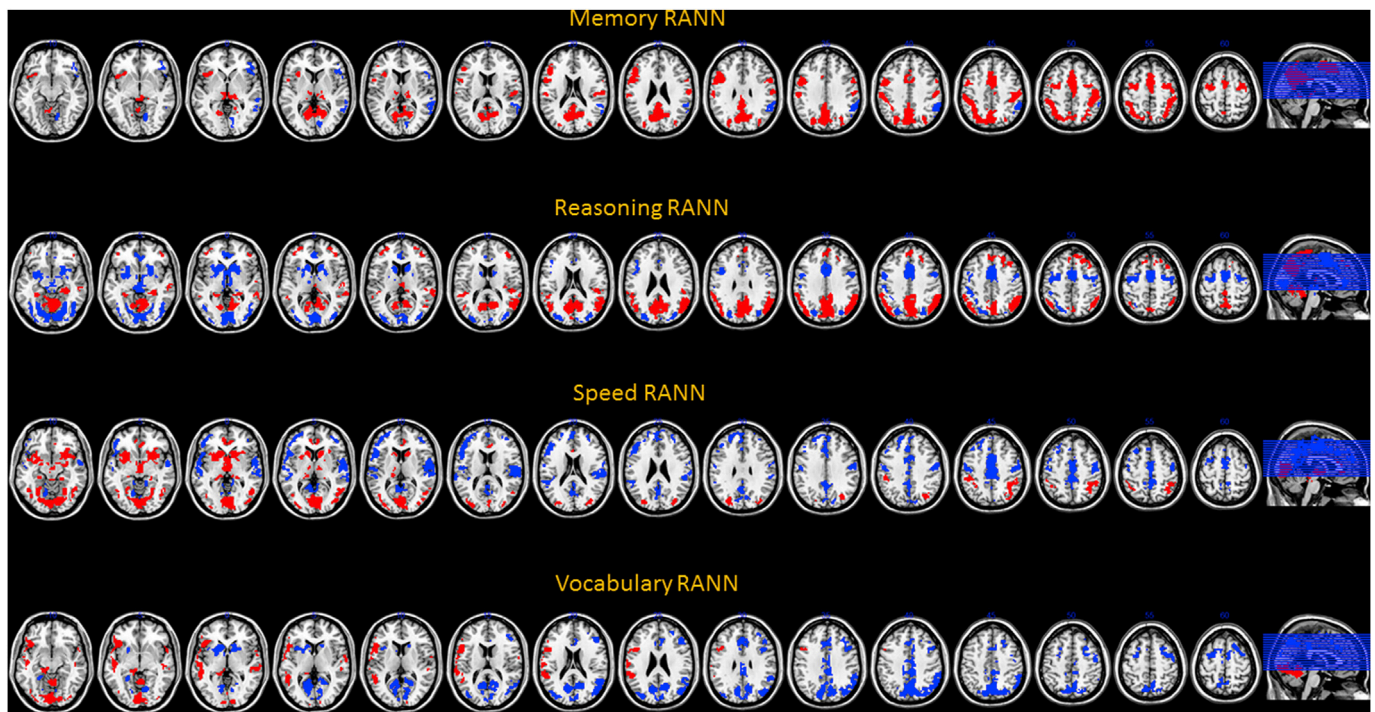
Fig. 4. Topographic composition of 4 derived RANNs. Red corresponds to activation, blue to de-activation. The statistical threshold |Z| was chosen as 3, with a minimum cluster size of 50. These settings were more liberal than in Tables 2–5, to allow a more complete visual appreciation.
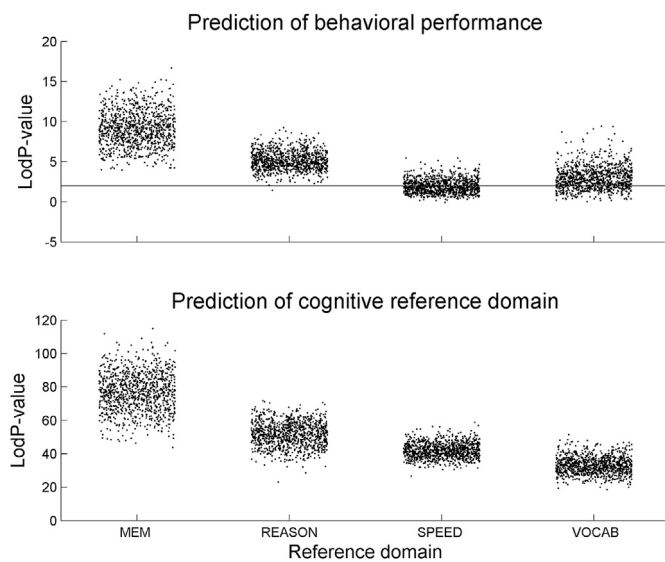

Fig. 5. Replication performance for behavioral correlation out of sample. The RANNs that were derived in the derivation sample of 200 study participants by optimizing both behavioral and indicator correlation were subsequently tested for behavioral and indicator correlation in the left-out data for 55 study participants. **Top panel:** prediction of behavioral performance out of sample. The black line indicates the lodP = 2, i.e. p < .01. The number of iterations (out of 1000) yielding replications with lodP>2 are 1000 for MEM, 999 for REASON, 426 for SPEED, and 750 for VOCAB. **Bottom panel:** quasi replication performance for indicator correlation; all replications were significant at p < .01.

right hemisphere and found in right temporal and inferior frontal areas. For the REASON pattern, positive loadings consisted mainly of bilateral parietal and occipital with only scant frontal involvement; negative loadings were found in bilateral lingual gyri, caudate, putamen, and the

supplementary motor area with a prominent focus of de-activation in the anterior cingulate. For SPEED, regions with positive loading included mainly basal ganglia and occipital areas, while negative loadings were found mainly in bilateral frontal and parietal areas. The VOCAB-pattern was highly left-lateralized, as expected: positive areas included left-fronto-temporal areas, while negative loadings included the right parietal and occipital areas.

Detailed listings of the super-threshold areas with anatomical labels can be found in Tables 2–5. We also checked forward-inference maps from Neurosynth for Memory, Reason and Speed to convey to the reader how many of our activation loci have been covered in previous papers with the search terms "Episodic Memory", "Reasoning" and "Speed", respectively, and how many loci appear as novel.

Relation of RANN patterns to age, education, verbal intelligence, and brain structure

For each participant, we obtained the pattern score of each RANN in the three constituent task-activation maps and averaged the scores within participant across all three tasks. Pattern scores and mean-squared residual activations were then considered together with age, education, NART-IQ, mean cortical thickness, and mean cortical volume. We computed both bivariate as well as partial correlations, for each reference ability. Below we display the full bivariate and partial correlations, and briefly comment on the results for cognitive performance and network scores for all reference abilities.

Memory

We direct our attention towards row 1 in Table 6, to glean the unique associations that remain with Memory performance after partialling out all other influences. Pattern expression of the Memory-RANN, education and NART-IQ remain uniquely associated with Memory performance. The unique variance contributed to Memory performance by the RANN pattern was 19.4%, with a total joint variance contribution by all measures of 37.4%. In the lower triangular matrix, the first column in Table 6 indicates that all measures apart from residual activation show bivariate

Table 2

Most prominent loadings for the Memory RANN. Shown are local maxima that fulfilled |Z|>4.75, CS > 100, and were at least 3 voxels (=9 mm) apart. Apart from the AAL label, we also show the Z-value obtained from a forward-inference (=P (Activation | Function)) from the meta-analytic website http://www.neurosynth.org/, where "episodic memory" was input and 270 studies were obtained.

MNI-X	MNI-Y	MNI-Z	CS	Z	AAL_label	"Episodic memory"
Positive loadings						
39	−21	48	356	11.9253	Postcentral_R	0
42	−15	57	356	11.5455	Precentral_R	0
33	−9	63	356	10.7692	Precentral_R	0
0	0	54	225	10.7326	Supp_Motor_Area_L	0
45	−30	48	356	10.3542	Postcentral_R	0
−54	−21	33	988	10.0158	Postcentral_L	0
−39	−39	48	988	9.9248	Parietal_Inf_L	0
51	−21	42	356	9.9025	Postcentral_R	0
51	−24	54	356	9.8786	Postcentral_R	0
36	−36	54	356	9.8197	Postcentral_R	0
−51	9	30	281	9.6498	Frontal_Inf_Oper_L	3.7202
−12	−69	27	988	9.5252	Cuneus_L	0
0	−66	33	988	9.5086	Precuneus_L	0
3	6	48	225	9.2556	Supp_Motor_Area_R	0
−42	6	27	281	8.7155	Frontal_Inf_Oper_L	7.8370
−45	−30	39	988	8.6612	Parietal_Inf_L	0
−54	3	39	281	8.6584	Precentral_L	0
−51	−30	48	988	8.3301	Parietal_Inf_L	0
−42	18	27	281	8.3246	Frontal_Inf_Tri_L	8.2944
15	−63	12	988	8.2741	Calcarine_R	0
−30	−57	48	988	7.9746	Parietal_Inf_L	7.8370
0	−57	27	988	7.6562	Precuneus_L	6.0073
6	−18	48	225	7.6444	Cingulum_Mid_R	0
15	−60	21	988	7.5801	Cuneus_R	0
−24	−72	51	988	7.5179	Parietal_Sup_L	0
−21	−66	57	988	7.487	Parietal_Sup_L	0
−27	−78	33	988	7.2738	Occipital_Mid_L	0
−36	−54	57	988	7.2029	Parietal_Inf_L	0
−36	−48	48	988	7.1261	Parietal_Inf_L	5.5499
−54	6	15	281	6.983	Rolandic_Oper_L	0
51	−18	18	356	6.8531	Rolandic_Oper_R	0
3	−66	21	988	6.6697	Cuneus_R	0
−9	−69	15	988	6.5713	Calcarine_L	0
−3	−54	39	988	6.5515	Precuneus_L	0
−51	30	18	281	6.516	Frontal_Inf_Tri_L	6.9221
63	−18	27	356	6.4028	SupraMarginal_R	0
−24	−66	39	988	6.3517	Occipital_Mid_L	7.8370
−3	−39	24	988	6.299	Cingulum_Post_L	0
−3	−30	27	988	6.2677	No label	0
−3	−42	33	988	6.0313	Cingulum_Post_L	0
−12	−66	6	988	5.9293	Calcarine_L	0
−12	−69	51	988	5.8469	Precuneus_L	0
39	−48	57	356	5.7896	Parietal_Sup_R	0
6	−45	18	988	5.4151	Precuneus_R	0
−3	21	45	225	5.2257	Supp_Motor_Area_L	6.9221
−9	12	39	225	5.0226	Cingulum_Mid_L	3.7202
Negative loadings						
54	−48	42	372	−9.3197	Parietal_Inf_R	0
57	−42	48	372	−7.9422	Parietal_Inf_R	0
48	27	3	118	−7.7577	Frontal_Inf_Tri_R	0
57	−54	3	372	−7.5337	Temporal_Mid_R	0
57	−45	30	372	−7.3346	SupraMarginal_R	0
9	−72	−3	103	−7.266	Lingual_R	0
45	33	−3	118	−7.1925	Frontal_Inf_Orb_R	0
54	15	−9	118	−7.0586	Temporal_Pole_Sup_R	0
12	−78	3	103	−6.8006	Calcarine_R	0
63	−39	9	372	−6.7591	Temporal_Mid_R	0
63	−39	27	372	−6.7543	SupraMarginal_R	0
63	−33	0	372	−6.6477	Temporal_Mid_R	0
63	−48	9	372	−6.5084	Temporal_Mid_R	0
51	−42	15	372	−6.3491	Temporal_Sup_R	0
15	−63	−9	103	−6.3189	Lingual_R	0
57	15	6	118	−6.2341	Frontal_Inf_Oper_R	0
66	−36	36	372	−6.0834	SupraMarginal_R	0
12	−87	6	103	−6.0626	Calcarine_R	0
48	−54	30	372	−5.9234	Angular_R	0
60	−57	18	372	−5.7677	Temporal_Mid_R	0
12	−57	−21	103	−5.6877	Cerebellum_6_R	0
45	−60	45	372	−5.1836	Angular_R	0
51	−42	3	372	−5.1714	Temporal_Mid_R	0

Table 2 (*continued*)

MNI-X	MNI-Y	MNI-Z	CS	Z	AAL_label	"Episodic memory"
51	−66	39	372	−5.0247	Angular_R	0

relationships with performance in the expected directions. Unique associations with pattern expression (row 2 in Table 6) were only shown by age and the residual activation with negative and positive directionality, respectively. (We repeat that apriori the residual activation cannot be expected to show any particular relationship to the pattern score across participants. Empirically, all manifestations – negative correlation, no correlation, positive correlation – are possible.) Bivariate correlations with the pattern score in the expected directions were obtained for most measures apart from NART-IQ: more intact brain structure, lower age and higher education were all associated with higher levels of pattern expression.

Reasoning

Reasoning performance was uniquely associated with pattern expression, age, education and NART-IQ in the expected directions. The unique variance contributed to Reasoning performance by the RANN pattern was 16.8%, with a total joint variance contribution by all measures of 42.0%. The full bivariate correlations showed associations of performance with all measures in the expected direction: positive associations with brain structure, NART-IQ and education, negative associations with age and residual activation. In full bivariate correlations, pattern expression was associated positively with education, and negatively with residual activation; under partial correlations, the residual activation retained its negative relationship to pattern expression, and age was associated *positively* with the pattern score (see Table 7).

Speed

Speed performance (=negative reaction time) showed unique associations with the pattern score, age and NART-IQ in partial correlations with expected directionality. The unique variance contributed to Speed performance by the RANN pattern was only 4.8%, with a total joint variance contribution by all measures of 35.9%. In full bivariate correlations, additional associates were shown by residual activation and brain structure with expected directionality. No unique associations with pattern score were found in partial correlations apart from a positive association with residual activation; however, in bivariate correlations additional negative and positive associations with age and brain structure, respectively, appeared (see Table 8).

Vocabulary

Of all reference abilities, Vocabulary presents the most number of associations with performance in partial correlations. Further, these partial correlations show a positive association of performance with age. Beyond age, residual activation was associated *negatively* with performance, while pattern expression, education and NART-IQ was associated positively with performance. Pattern expression showed partial correlations with residual activation, age and mean cortical volume. The unique variance contributed to Vocabulary performance by the RANN pattern was 7.8%, with a total joint variance contribution by all measures of 53.7%. For the full bivariate correlations, residual activation was associated negatively with performance, while age, education and NART-IQ showed positive associations (see Table 9).

Collinearity of behavioral performance and pattern scores

Our analysis paid no heed to possible collinearities of our pattern scores between the reference domains. Ideally, the pattern score pertaining to the reference domains should be independent of one another. However, even behavioral performance in the four domains was collinear. For behavioral performances, the only pairing that was *not* significant was R (SPEED, VOCAB) = 0.11, p = .07. All other pairings are

Table 3

Most prominent loadings for the Reasoning RANN. Shown are local maxima that fulfilled |Z|>4.75, CS > 100, and were at least 3 voxels (=9 mm) apart. Apart from the AAL label, we also show the Z-value obtained from a forward-inference (=P (Activation | Function)) from the meta-analytic website http://www.neurosynth.org/, where "reasoning" was input and 147 studies were obtained.

MNI-X	MNI-Y	MNI-Z	CS	Z	AAL_label	"Reasoning"
Positive loadings						
51	−63	39	505	10.9877	Angular_R	0
45	−60	33	505	10.8976	Angular_R	0
9	−54	27	887	10.0959	Precuneus_R	0
0	−45	−24	887	10.0552	No label	0
51	−51	36	505	9.5218	Angular_R	0
48	−51	45	505	9.52	Parietal_Inf_R	3.8672
57	−60	24	505	9.3614	Angular_R	0
−45	−63	30	218	8.9327	Angular_L	4.5394
48	−57	21	505	8.9022	Temporal_Mid_R	0
45	−72	39	505	8.8835	Angular_R	0
12	−60	−6	887	8.7543	Lingual_R	0
6	−69	36	887	8.6382	Precuneus_R	0
−6	−63	30	887	8.6188	Precuneus_L	0
60	−51	33	505	8.2465	Angular_R	4.5394
12	−63	18	887	7.9755	Calcarine_R	0
3	−48	−15	887	7.8744	Vermis_4_5	0
−6	−66	18	887	7.5784	Calcarine_L	0
3	−78	45	887	7.499	No label	0
−6	−57	21	887	7.3586	Precuneus_L	0
−6	−81	30	887	7.281	Cuneus_L	0
−54	−54	36	218	7.155	Angular_L	0
12	−42	36	887	6.8965	Cingulum_Mid_R	0
−15	−66	15	887	6.7833	Calcarine_L	0
−9	−63	39	887	6.6493	Precuneus_L	0
−45	−57	51	218	6.3247	Parietal_Inf_L	0
−12	−75	21	887	6.2848	Cuneus_L	0
−18	−66	0	887	6.1497	Lingual_L	0
9	−48	3	887	5.9608	Lingual_R	0
54	−48	18	505	5.7993	Temporal_Sup_R	3.8672
−36	−78	42	218	5.7083	Parietal_Inf_L	0
−63	−42	33	218	5.6282	SupraMarginal_L	0
3	−54	−3	887	5.5999	Vermis_4_5	0
18	−63	−21	887	5.5868	Cerebellum_6_R	0
3	−69	−9	887	5.541	Vermis_6	0
−12	−42	−15	887	5.2856	Cerebellum_4_5_L	0
12	−48	−21	887	5.1616	Cerebellum_4_5_R	0
Negative loadings						
15	−84	−12	1220	−12.8391	Lingual_R	0
−27	−81	21	1220	−12.2673	Occipital_Mid_L	0
33	−87	6	1220	−11.9318	Occipital_Mid_R	0
−21	−84	−12	1220	−11.557	Lingual_L	0
−27	−93	6	1220	−11.0113	Occipital_Mid_L	0
−6	−90	3	1220	−10.445	Calcarine_L	4.5394
15	−90	0	1220	−10.4129	Calcarine_R	0
−27	−75	30	1220	−10.3594	Occipital_Mid_L	0
−12	−84	−9	1220	−10.1431	Lingual_L	3.8672
30	−78	24	1220	−10.1321	Occipital_Mid_R	0
36	−81	12	1220	−10.007	Occipital_Mid_R	0
21	6	−9	419	−9.893	No label	0
27	−90	18	1220	−9.6788	Occipital_Mid_R	0
−33	−54	−15	1220	−9.417	Fusiform_L	0
−18	12	0	419	−9.3854	Putamen_L	0
−6	9	3	419	−9.0603	Caudate_L	0
21	18	−6	419	−8.9707	Putamen_R	0
9	9	0	419	−8.9481	No label	0
33	−54	−15	1220	−8.7846	Fusiform_R	0
−33	−78	−15	1220	−8.6847	Fusiform_L	0
6	3	48	266	−8.4921	Supp_Motor_Area_R	0
−6	3	48	266	−8.4297	Supp_Motor_Area_L	4.5394
−21	−75	39	1220	−8.2544	Occipital_Sup_L	0
−18	3	−9	419	−8.2343	No label	0
−39	−78	−6	1220	−7.9954	Occipital_Inf_L	0
−6	0	60	266	−7.9589	Supp_Motor_Area_L	0
42	−63	−12	1220	−7.9157	Occipital_Inf_R	0
−21	0	3	419	−7.8126	Pallidum_L	0
−27	−93	−9	1220	−7.7419	Occipital_Inf_L	0
−42	−69	−6	1220	−7.6888	Occipital_Inf_L	0
−33	−63	−15	1220	−7.6843	Fusiform_L	0
6	6	57	266	−7.6087	Supp_Motor_Area_R	0
39	−72	−12	1220	−7.3593	Occipital_Inf_R	0
30	−63	−12	1220	−7.2268	Fusiform_R	0
6	3	39	266	−7.1928	Cingulum_Mid_R	0

Table 3 (*continued*)

MNI-X	MNI-Y	MNI-Z	CS	Z	AAL_label	"Reasoning"
6	12	39	266	−7.1104	Cingulum_Mid_R	4.5394
−18	−6	−9	419	−7.0838	No label	0
30	−75	33	1220	−6.7055	Occipital_Mid_R	0
15	−93	18	1220	−6.5997	Occipital_Sup_R	0
−9	21	−3	419	−6.5434	Caudate_L	0
42	−78	−3	1220	−6.4869	Occipital_Inf_R	0
−6	9	36	266	−6.008	Cingulum_Mid_L	0
−27	0	−12	419	−5.9485	Amygdala_L	0
27	−90	−9	1220	−5.8054	Occipital_Inf_R	3.8672
12	6	12	419	−5.8013	Caudate_R	0
−24	−66	48	1220	−5.4537	Parietal_Sup_L	6.5561
24	15	−18	419	−5.3344	Insula_R	0
−33	−3	−6	419	−5.0504	No label	0
24	3	−21	419	−5.0347	Amygdala_R	0

Table 4

Most prominent loadings for the Speed RANN. Shown are local maxima that fulfilled |Z|>4.75, CS > 100, and were at least 3 voxels (=9 mm) apart. Apart from the AAL label, we also show the Z-value obtained from a forward-inference map from the meta-analytic website Neurosynth, where "speed" was input and 162 studies were obtained.

MNI-X	MNI-Y	MNI-Z	CS	Z	AAL_label	"Speed"
Positive loadings						
9	−84	0	471	10.2884	Calcarine_R	0
0	−78	3	471	9.6483	Lingual_L	0
−21	−78	−9	471	8.8408	Lingual_L	0
27	−78	−9	471	8.7919	Fusiform_R	0
18	−78	−9	471	8.616	Lingual_R	0
18	18	0	207	7.2603	Putamen_R	0
−6	−15	−3	207	7.2594	No label	4.1714
3	−9	−6	207	7.2138	No label	0
−27	9	−9	145	7.0775	No label	0
−33	−3	−15	145	7.0474	No label	0
27	−57	−12	471	7.0394	Fusiform_R	0
21	12	−12	207	6.9869	No label	0
36	−57	−12	471	6.9696	Fusiform_R	0
−24	18	−6	145	6.9228	Putamen_L	0
−18	−78	−18	471	6.9205	Cerebellum_6_L	0
−27	0	−21	145	6.5311	Amygdala_L	0
9	−93	9	471	6.2369	Calcarine_R	0
15	3	−3	207	6.2097	Pallidum_R	0
33	9	−6	207	6.0466	No label	0
−12	−78	6	471	5.9167	Calcarine_L	0
−15	9	−3	145	5.9102	Pallidum_L	0
30	21	6	207	5.5987	No label	6.2769
27	−66	−6	471	5.4255	Fusiform_R	0
30	18	−3	207	5.1694	No label	0
Negative loadings						
3	−15	48	227	−8.6955	Cingulum_Mid_R	0
51	−12	15	104	−8.254	Rolandic_Oper_R	0
−51	39	0	166	−7.7443	Frontal_Inf_Tri_L	0
−42	48	0	166	−7.559	Frontal_Mid_Orb_L	0
−27	57	3	166	−7.5183	Frontal_Sup_L	0
3	−6	48	227	−7.3561	Cingulum_Mid_R	0
60	−15	15	104	−7.1207	Rolandic_Oper_R	0
−39	48	9	166	−6.8602	Frontal_Mid_L	0
0	−30	48	227	−6.8448	Cingulum_Mid_L	0
6	−45	63	227	−6.6972	Precuneus_R	0
−48	33	18	166	−6.6617	Frontal_Inf_Tri_L	0
−3	3	63	227	−6.2572	Supp_Motor_Area_L	0
6	−36	51	227	−6.2428	Cingulum_Mid_R	0
−54	21	21	166	−6.2293	Frontal_Inf_Tri_L	0
−42	33	27	166	−6.1212	Frontal_Inf_Tri_L	4.1714
9	6	45	227	−6.0324	Cingulum_Mid_R	0
3	−3	60	227	−5.8752	Supp_Motor_Area_R	0
66	−12	9	104	−5.8694	Temporal_Sup_R	0
−6	−3	48	227	−5.7517	Cingulum_Mid_L	4.8733
−54	24	9	166	−5.749	Frontal_Inf_Tri_L	0
3	9	39	227	−5.5305	Cingulum_Mid_R	0
−9	3	39	227	−4.9079	Cingulum_Mid_L	0
3	−48	54	227	−4.8875	Precuneus_R	0

Table 5

Most prominent loadings for the Vocabulary RANN. Shown are local maxima that fulfilled |Z|>4.75, CS > 100, and were at least 3 voxels (=9 mm) apart. No suitable meta-analytic maps could be obtained from Neurosynth for comparison.

MNI-X	MNI-Y	MNI-Z	CS	Z	AAL_label
Positive loadings					
0	−90	−9	714	9.3887	Calcarine_L
24	−54	−27	714	9.007	Cerebellum_6_R
−18	−78	−18	714	8.4283	Cerebellum_6_L
3	−60	−24	714	8.1938	Vermis_6
−9	−81	−15	714	7.7469	Cerebellum_6_L
21	−69	−24	714	7.5566	Cerebellum_6_R
18	−63	−18	714	7.4147	Cerebellum_6_R
9	−69	−21	714	7.1698	Cerebellum_6_R
−48	27	−9	102	7.1204	Frontal_Inf_Orb_L
−48	12	21	102	7.0958	Frontal_Inf_Oper_L
12	−87	−15	714	6.9094	No label label
9	−81	−21	714	6.8237	Cerebellum_Crus1_R
6	−54	−6	714	6.6599	Vermis_4_5
−54	12	−3	102	6.6148	Temporal_Pole_Sup_L
−24	−51	−27	714	6.563	Cerebellum_6_L
−54	18	9	102	6.5326	Frontal_Inf_Tri_L
−36	−45	−18	714	6.5202	Fusiform_L
−15	−63	−21	714	6.1343	Cerebellum_6_L
−6	−72	−15	714	6.0599	Cerebellum_6_L
−42	−54	−21	714	5.9555	Fusiform_L
−36	−75	−18	714	5.8894	Cerebellum_Crus1_L
36	−63	−27	714	5.3858	Cerebellum_Crus1_R
−9	−57	−27	714	5.2374	No label label
−45	−69	−15	714	5.1494	Occipital_Inf_L
−12	−96	0	714	5.0924	Calcarine_L
Negative loadings					
−12	−72	39	495	−8.2775	Precuneus_L
18	−72	39	495	−8.1227	Cuneus_R
9	−66	36	495	−7.4914	Precuneus_R
−18	−60	15	495	−7.4507	No label label
21	−63	24	495	−7.2701	Precuneus_R
18	−78	48	495	−7.2486	Cuneus_R
9	−84	24	495	−7.059	Cuneus_R
18	−54	9	495	−6.5976	Calcarine_R
9	−84	15	495	−6.5446	Calcarine_R
−12	−51	9	495	−6.3994	Calcarine_L
−9	−69	24	495	−6.1759	Calcarine_L
27	−66	39	495	−5.8724	Occipital_Sup_R
−15	−81	42	495	−5.8237	Occipital_Sup_L
6	−78	39	495	−5.7069	Cuneus_R
6	−54	42	495	−5.6951	Precuneus_R
−9	−60	9	495	−5.5549	Calcarine_L
18	−75	30	495	−5.1257	Cuneus_R
9	−63	51	495	−4.9296	Precuneus_R

Table 6
Results for the Memory Domain. *p < .05, **p < .01, ***p < .001, ****p < .0001. Upper triangular part: partial correlations; lower triangular part: full bivariate correlations. Of particular interest is the first row, which shows unique contributions to Memory performance. The unique part of the Memory pattern accounts for $(0.44)^2 = 19.4\%$ of the variance in Memory performance. The total amount of variance accounted for jointly by all independent variables in Memory performance is 37.4%. Listed in bold face are correlations that have a p-value smaller than 0.05.

	MEM-perf	MEM-pattern	RES	AGE	EDU	NART-IQ	mean-THX	mean-VOL
MEM-perf		**0.44******	−0.08	−0.13	**0.18****	**0.15***	0.02	0.06
MEM-pattern	**0.54******		**0.20****	**−0.21****	0.02	0.00	0.05	0.00
RES	−0.03	**0.13***		−0.02	−0.02	0.00	−0.09	−0.05
AGE	**−0.27******	**−0.38******	−0.02		0.10	**0.33******	**−0.36******	**−0.32******
EDU	**0.30*****	**0.13***	−0.05	**0.14***		**0.41******	−0.01	0.06
NART-IQ	**0.23*****	0.04	−0.03	**0.28******	**0.52******		0.06	**0.17****
mean-THX	**0.21****	**0.25*****	−0.05	**−0.53******	0.00	−0.02		**0.26******
mean-VOL	**0.30******	**0.26******	−0.06	**−0.46******	**0.13***	0.11	**0.43******	

Table 7
Results for Reasoning, *p < .05, **p < .01, ***p < .001, ****p < .0001. Upper triangular part: partial correlations; lower triangular part: full bivariate correlations. The first row shows unique contributions to Fluid-Reasoning performance. The unique part of the REASON pattern accounts for $(0.41)^2 = 16.8\%$ of the variance in performance. The total amount of variance accounted for jointly by all independent variables in Reasoning performance is 42.0%. Listed in bold face are correlations that have a p-value smaller than 0.05.

	REASON-perf	REASON-pattern	RES	AGE	EDU	NART-IQ	mean-THX	mean-VOL
REASON-perf		**0.41******	0.05	**−0.28******	**0.18****	**0.24*****	0.12	0.11
REASON-pattern	**0.41******		**−0.34******	**0.18****	0.02	−0.11	−0.08	0.07
RES	**−0.13***	**−0.34******		0.09	−0.04	−0.05	−0.06	0.10
AGE	**−0.30******	0.04	0.04		0.09	**0.36******	**−0.34******	**−0.31******
EDU	**0.31******	**0.15***	−0.10	**0.14***		**0.39******	−0.01	0.05
NART-IQ	**0.28*****	0.10	−0.06	**0.28******	**0.52******		0.04	**0.15***
mean-THX	**0.28******	−0.02	−0.06	**−0.53******	0.00	−0.02		**0.26*****
mean-VOL	**0.39******	0.11	−0.01	**−0.46******	**0.13***	0.11	**0.43******	

Table 8
Results for Speed, *p < .05, **p < .01, ***p < .001, ****p < .0001. Upper triangular part: partial correlations; lower triangular part: full bivariate correlations. The first row shows unique contributions to Speed performance. The unique part of the SPEED-pattern accounts for $(0.22)^2 = 4.8\%$ of the variance in Speed performance. The total amount of variance accounted for jointly by all independent variables in performance is 35.9%. Speed performance was defined as negative reaction time such that higher values imply better performance. Listed in bold face are correlations that have a p-value smaller than 0.05.

	SPEED-perf	SPEED-pattern	RES	AGE	EDU	NART-IQ	mean-THX	mean-VOL
SPEED-perf		**0.22******	−0.05	**−0.46******	−0.13	**0.28******	−0.11	−0.01
SPEED-pattern	**0.27******		**0.27******	0.01	−0.03	0.02	0.08	0.15
RES	**−0.22*****	**0.20****		**0.30******	0.03	−0.08	−0.03	0.00
AGE	**−0.51******	−0.13	**0.42******		−0.04	**0.41******	**−0.38******	**−0.29******
EDU	−0.01	0.00	0.04	**0.14***		**0.48******	−0.01	0.08
NART-IQ	0.09	0.06	0.06	**0.28******	**0.52******		0.10	**0.17***
mean-THX	**0.23*****	**0.16***	**−0.23*****	**−0.53******	0.00	−0.02		**0.25*****
mean-VOL	**0.31******	**0.23*****	**−0.17****	**−0.46******	**0.13***	0.11	**0.43******	

Table 9
Results for Vocabulary, *p < .05, **p < .01, ***p < .001, ****p < .0001. Upper triangular part: partial correlations; lower triangular part: full bivariate correlations. The first row shows unique contributions to Vocabulary performance. The unique part of the VOCAB-pattern accounts for $(0.28)^2 = 7.8\%$ of the variance in VOCAB performance. The total amount of variance accounted for jointly by all independent variables in performance is 57.3%. Listed in bold face are correlations that have a p-value smaller than 0.05.

	VOCAB-perf	VOCAB-pattern	RES	AGE	EDU	NART-IQ	mean-THX	mean-VOL
VOCAB-perf		**0.28******	**−0.20****	**0.17***	**0.16***	**0.55******	0.07	−0.01
VOCAB-pattern	**0.36******		**0.22*****	**0.24*****	−0.11	−0.07	0.11	**0.19****
RES	**−0.16***	0.12		0.03	0.06	0.00	**−0.17****	0.01
AGE	**0.32******	**0.25*****	0.10		0.03	0.13	**−0.41******	**−0.38******
EDU	**0.45******	0.09	−0.07	0.14		**0.29******	0.01	0.09
NART-IQ	**0.70******	**0.21****	−0.11	**0.28******	**0.52******		−0.01	**0.14***
mean-THX	−0.03	−0.01	**−0.20****	**−0.53******	0.00	−0.02		**0.24*****
mean-VOL	0.08	**0.13***	−0.09	**−0.46******	**0.13***	0.11	**0.43******	

significantly positively correlated at p < .0001.

The derived activation patterns were less collinear. We found three (of six) correlations between pattern scores that are significant: R (MEM, SPEED) = 0.29, p < .0001; R (MEM, VOCAB) = 0.22, p = .0004; and R (SPEED, VOCAB) = 0.56, p < .0001.

Discussion

In this paper, we derived four reference ability neural network (RANN) activation patterns, representing the cognitive domains of Memory, Reasoning, Speed, and Vocabulary in a large sample of adults across the age-range. Our approach, which combined Principal Components Analysis with a subsequent brute-force search, optimized not only brain-behavioral correlations, but also cognitive specificity of the RANNs. Our approach achieved quasi out-of-sample indicator and brain-behavioral correlations that far exceeded chance. Indeed, replication success produced significant correlations (p < .01) with reference ability-specific indicator variables for all 1000 iterations of a Monte-Carlo 5-fold

cross-validation procedure, while brain-behavioral correlations showed some variability and gave significant correlations with behavior for Memory in 100% of the 1000 iterations, 99% for Reasoning, 43% for Speed, and 75% for Vocabulary.

Success of quasi out-of-sample replication tracked the amount of unique variance accounted for in behavioral performance by the derived RANN patterns, and confirmed the following descending order: Memory, Reasoning, Vocabulary, and Speed. This ranking speaks to the success of achieving a unifying neural account of both domain specificity and behavioral performance. Memory and Reasoning allow such an account. In fact, for these two reference abilities adopting approaches that focus exclusively on cognitive domain specificity (without explicit consideration of brain-behavioral correlations) achieves passable brain-behavioral correlation already by itself (data not shown). For Speed and Vocabulary this is not possible, however. Brain-behavioral correlation needed to be addressed explicitly, and out-of-sample replication was still worse than for Memory and Reasoning. Our data revealed that the neural implementation underlying task performance is *not* as distinct from other cognitive processes for Speed and Vocabulary as for Memory and Reasoning.

It bears mentioning here that our guiding assumptions for this report, i.e. simultaneous indicator and brain-behavioral correlation of the derived activation patterns, are very strong. While they are suitable starting points for an investigation, considerations other than the extent of empirical cross-sectional validation in the current report might come into play in the future, and modify or constrain these assumptions, leading to possibly different analytic strategies on longitudinal data to produce RANNs with better biological plausibility or better predictive utility. More data would also allow sub-setting by age with a focus on the possible topographic re-organization of these networks across the adult life span.

Few variables showed independent influences on RANN pattern utilization, and several influences on behavioral performance remain beyond pattern utilization

For all four reference abilities, we examined possible drivers that influence RANN pattern scores and cognitive performance in partial correlations. Demographic variables including age, education and NART-IQ, as well as brain-health variables including mean cortical thickness and mean cortical volume, were probed for their associations with RANN pattern scores and behavioral performance in full bivariate and partial correlations. We confined our remarks in the following to the more rigorous partial-correlation results only. The picture that emerged was quite consistent, and only varied minimally by reference ability. All RANN pattern scores showed associations with residual activation, albeit with varying signs: positive for Memory, Speed, and Vocabulary, negative for Reasoning. We stress again that such across-participant associations do not follow *a priori* from the manner in which residual activation is computed on a single-participant basis. Residual activation had few other influences: for Memory and Reasoning there were none, for Speed residual activation was correlated positively with age, and for Vocabulary, it was negatively correlated with mean cortical thickness. Residual activation thus was selectively correlated with negative health indicators, i.e. higher age or lower cortical thickness. Only for Vocabulary was residual activation also independently associated with *worse* performance, otherwise it appeared as neutral. Notably absent in the partial correlations were independent influences on performance by the included brain structure measures. Otherwise performance was influenced by age, IQ and education, with slight variations across reference abilities. For Vocabulary, age was *positively* correlated with performance, which fits well with prior cross-sectional and longitudinal findings of a *positive* influence of age on VOCAB performance (Albert et al., 1988; Botwinick and Siegler, 1980). .

Connection to wider literature

The results of the current report shed new light on the neural basis of cognition, and the influence of demographics including age, IQ and level of education, and brain structure on this neural basis. For each of the four reference abilities we derived a unique pattern. It is notable that compared to our previous analysis which did not account for behavioral performance (Habeck et al., 2016), the new activation patterns presented in this report showed much more topographically extensive activation, with much bigger robustness as measured in our bootstrap Z-loadings. This was to be expected: despite the additional brain-behavioral constraint used in the pattern derivation which ordinarily should reduce activation, our earlier work focused only on participants in the age range 20–30, whereas the current report utilized more extensive data from the whole adult life span of ages 20–80. In both the MEM and REASON reference abilities, better behavioral performance was related to greater activation in the left lateralized rostral and caudal loci of the inferior parietal lobule (IPL), highly interconnected areas implicated in multiple cognitive abilities including complex mental arithmetic (Chochon et al., 1999) as well as bottom-up recollection processes involved in the "spontaneous retrieval of episodic details (Cabeza et al., 2008). The MEM ability also loaded positively on the left precuneus, an area shown in non-human primate studies to be highly connected to both cortical and subcortical structures (Leichnetz, 2001), the left superior parietal area, implicated in cognitive control processes that allow for goal-directed, top-down episodic memory retrievals (Cabeza et al., 2008, p. 618), as well a sub-region of the Ventral Lateral Prefrontal cortex –the pars triangularis, which has been consistently implicated in episodic memory retrieval (Badre and Wagner, 2007).

Likewise, the SPEED and VOCAB RANNs also include several notable brain areas. Several components of the limbic system, including the amygdala, orbital frontal cortex and thalamus, have positive loadings in the SPEED RANN. The limbic system has been typically associated with emotional processing, motivation and fear conditioning. For example, Ono and colleagues (Ono et al., 2000) suggested the limbic system motivated eating behaviors, operationalized as neuronal single-cell recordings in monkeys during a food-nonfood discrimination task: "the amygdala might be important in ongoing recognition of the affective significance of complex stimuli". Given the positive relationship found between activity in this set of brain areas, and decreased latencies to respond correctly on the SPEED tasks, a motivational account may fit with the data presented here, such that faster responses are subserved by the same systems implicated in motivated, often times unconscious, responses to external stimuli. The VOCAB-RANN, on the other hand, was associated with activity in the superior temporal lobe, including Brodmann areas 21 and 22 (Wernicke's area). These areas are key for speech and language generation (Friedman et al., 1998), the processing of speech and text (Giraud et al., 2004), and the semantic meaning of words (Chou et al., 2006; McDermott et al., 2003). Indeed, dysfunction in the left STG has been implicated in language processing deficits that are a marked symptom of autism (Bigler et al., 2007). Additionally, behavioral performance across the three VOCAB tasks was also associated with increased activation in the cerebellum, an area now recognized to be involved in linguistic processing (Smet et al., 2013).

Summary

Future research might revise our derived networks with longitudinal data, advanced analytic techniques, age-stratification and other requirements beyond brain-behavioral and indicator correlations across the adult life span. However, for now, we have derived activation pattern for four different cognitive reference abilities that could predict the appropriate reference-domain label and cognitive performance out of sample, far exceeding chance level. Participant expression of these patterns displayed associations with age, education and IQ with consistent directionality: older, less educated and less intelligent participants with

poorer brain-structural health manifested these patterns to a lesser degree. Residual activation orthogonal to the patterns also increased with age and poorer brain-structural health, and for Vocabulary was associated with worse performance as well. While the RANN patterns were derived to give significant accounts of cognitive performance and were revealed to be influenced by age, education, IQ, these variables mostly retained significant associations with performance beyond the influence of the patterns.

Funding support

References

Albert, M.S., Heller, H.S., Milberg, W., 1988. Changes in naming ability with age. Psychol. Aging 3, 173–178.

Badre, D., Wagner, A.D., 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia 45 (13), 2883–2901. https://doi.org/10.1016/j.neuropsychologia.2007.06.015.

Bigler, E.D., Mortensen, S., Neeley, E.S., Ozonoff, S., Krasny, L., Johnson, M., Lu, J., Provencal, S.L., McMahon, W., Lainhart, J.E., 2007. Superior temporal gyrus, language function, and autism. Dev. Neuropsychol. 31 (2), 217–238.

Botwinick, J., Siegler, I.C., 1980. Intellectual ability among the elderly: simultaneous cross-sectional and longitudinal comparisons. Dev. Psychol. 16 (49–53).

Cabeza, R., Ciaramelli, E., Olson, I.R., Moscovitch, M., 2008. The parietal cortex and episodic memory: an attentional account. Nat. Rev. Neurosci. 9 (8), 613–625. https://doi.org/10.1038/nrn2459.

Chochon, F., Cohen, L., van de Moortele, P.F., Dehaene, S., 1999. Differential contributions of the left and right inferior parietal lobules to number processing. J. Cognit. Neurosci. 11 (6), 617–630.

Chou, T.L., Booth, J.R., Bitan, T., Burman, D.D., Bigio, J.D., Cone, N.E., Cao, F., 2006. Developmental and skill effects on the neural correlates of semantic processing to visually presented words. Hum. Brain Mapp. 27 (11), 915–924.

Ekstrom, R.B., French, J.W., Harman, H.H., Dermen, D., 1976. Manual for Kit of Factor-referenced Cognitive Tests. Princeton.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B.,

Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. Cerebr. Cortex 14 (1), 11–22. https://doi.org/10.1093/cercor/bhg087.

Friedman, L., Kenny, J.T., Wise, A.L., Wu, D., Stuve, T.A., Miller, D.A., Jesberger, J.A., Lewin, J.S., 1998. Brain activation during silent word generation evaluated with functional MRI. Brain Lang. 64 (2), 231–256.

Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., Kleinschmidt, A., 2004. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cerebr. Cortex 14 (3), 247–255.

Habeck, C., Gazes, Y., Razlighi, Q., Steffener, J., Brickman, A., Barulli, D., Salthouse, T., Stern, Y., 2016. The Reference Ability Neural Network Study: life-time stability of reference-ability neural networks derived from task maps of young adults. Neuroimage 125, 693–704. https://doi.org/10.1016/j.neuroimage.2015.10.077.

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York.

Leichnetz, G.R., 2001. Connections of the medial posterior parietal cortex (area 7m) in the monkey. Anat. Rec. 263 (2), 215–236.

McDermott, K.B., Petersen, S.E., Watson, J.M., Ojemann, J.G., 2003. A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. Neuropsychologia 41 (3), 293–303.

Ono, T., Nishijo, H., Nishino, H., 2000. Functional role of the limbic system and basal ganglia in motivated behaviors. J. Neurol. 247 (Suppl. 5), V23–V32.

Raven, J.C., 1962. Advanced Progressive Matrices, Set II. H.K. Lewis, London, UK.

Salthouse, T., Ferrer-Caja, E., 2003. What needs to be explained to account for age-related effects on multiple cognitive variables? Psychol. Aging 18, 91–110.

Salthouse, T.A., 1993. Speed and knowledge as determinants of adult age differences in verbal tasks. J. Gerontol. 48 (1), P29–P36.

Salthouse, T.A., 1998. Independence of age-related influences on cognitive abilities across the life span. Dev. Psychol. 34 (5), 851–864.

Salthouse, T.A., Babcock, R.L., 1991. Decomposing adult age differences in working memory. Dev. Psychol. 27, 763–786.

Smet, H.J.D., Paquier, P., Verhoeven, J., Mariën, P., 2013. The cerebellum: its role in language and related cognitive and affective functions. Brain Lang. 127 (3), 334–342.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 (Suppl. 1), S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051.

Stern, Y., Habeck, C., Steffener, J., Barulli, D., Gazes, Y., Razlighi, Q., Shaked, D., Salthouse, T., 2014. The Reference Ability Neural Network Study: motivation, design, and initial feasibility analyses. Neuroimage 103, 139–151. https://doi.org/10.1016/j.neuroimage.2014.09.029.

Woodcock, R.W., Johnson, M.B., Mather, N., 1989. Woodcock-Johnson Psycho-Educational Battery–Revised. DLM Teaching Resources.