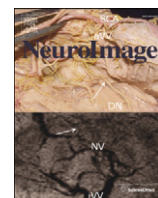


Contents lists available at [ScienceDirect](http://ScienceDirect.com)

## NeuroImage

journal homepage: [www.elsevier.com/locate/ynimg](http://www.elsevier.com/locate/ynimg)

## The Reference Ability Neural Network Study: Motivation, design, and initial feasibility analyses



Yaakov Stern <sup>a,\*</sup>, Christian Habeck <sup>a,1</sup>, Jason Steffener <sup>a</sup>, Daniel Barulli <sup>a</sup>, Yunglin Gazes <sup>a</sup>, Qolamreza Razlighi <sup>a</sup>, Danielle Shaked <sup>a</sup>, Timothy Salthouse <sup>b</sup>

<sup>a</sup> Cognitive Neuroscience Division, Department of Neurology and Taub Institute for Research on Alzheimer's Disease and The Aging Brain, Columbia University College of Physicians and Surgeons, 630 W 168th St, P&S Box 16, New York, NY 10032, USA

<sup>b</sup> Department of Psychology, University of Virginia, 102 Gilmer Hall, PO Box 400400, Charlottesville, VA 22904, USA

### ARTICLE INFO

#### Article history:

Accepted 13 September 2014

Available online 20 September 2014

#### Keywords:

Cognitive aging

fMRI

Convergent/discriminant validity

### ABSTRACT

We introduce and describe the Reference Ability Neural Network Study and provide initial feasibility data. Based on analyses of large test batteries administered to individuals ranging from young to old, four latent variables, or reference abilities (RAs) that capture the majority of the variance in age-related cognitive change have been identified: episodic memory, fluid reasoning, perceptual speed, and vocabulary. We aim to determine whether spatial fMRI networks can be derived that are uniquely associated with the performance of each reference ability. We plan to image 375 healthy adults (50 per decade from age 20 to 50; 75 per decade from age 50 to 80) while performing a set of 12 cognitive tasks. Data on 174 participants are reported here. Three tasks were grouped a priori into each of the four reference ability domains.

We first assessed to what extent both cognitive task scores and activation patterns readily show convergent and discriminant validity, i.e. increased similarity between tasks within the same domain and decreased similarity between tasks between domains, respectively. Block-based time-series analysis of each individual task was conducted for each participant via general linear modeling. We partialled activation common to all tasks out of the imaging data. For both test scores and activation topographies, we then calculated correlations for each of 66 possible pairings of tasks, and compared the magnitude of correlation of tasks within reference ability domains to that of tasks between domains. For the behavioral data, globally there were significantly stronger inter-task correlations within than between domains. When examining individual abilities, 3 of the domains also met these criteria but memory reached only borderline significance. Overall there was greater topographic similarity within reference abilities than between them ( $p < 0.0001$ ), but when examined individually, statistical significance was reached only for episodic memory and perceptual speed.

We then turned to a multivariate technique, linear indicator regression analysis, to derive four unique linear combinations of Principal Components (PC) of imaging data that were associated with each RA. We investigated the ability of the identified PCs to predict the reference domain associated with the activation of individual subjects for individual tasks. Median accuracy rates for associating component task activation with a particular reference ability were quite good: memory: 82%; reasoning: 87%; speed: 84%; vocabulary: 77%.

These results demonstrate that even using basic GLM analysis, the topography of activation of tasks within a domain is more similar than tasks between domains. The follow-up regression analyses suggest that all tasks with each RA rely on a common network, unique to that RA. Our ultimate goal is to better characterize these RA neural networks and then study how their expression changes across the age span. Our hope is that by focusing on these networks associated with key features of cognitive aging, as opposed to task-related activation associated with individual tasks, we will be able to advance our knowledge regarding the key brain changes that underlie cognitive aging.

© 2014 Elsevier Inc. All rights reserved.

### Introduction

There is a divergence of methods in the study of cognitive aging, with different approaches pursued by researchers within cognitive neuroimaging and cognitive psychologists interested in the structure of cognition. In an attempt to understand the neural underpinnings of age-related differences within a specific cognitive domain, neuroimaging researchers

\* Corresponding author.

E-mail addresses: [ys11@columbia.edu](mailto:ys11@columbia.edu) (Y. Stern), [ch629@columbia.edu](mailto:ch629@columbia.edu) (C. Habeck), [js2746@columbia.edu](mailto:js2746@columbia.edu) (J. Steffener), [djb2168@columbia.edu](mailto:djb2168@columbia.edu) (D. Barulli), [y12107@columbia.edu](mailto:y12107@columbia.edu) (Y. Gazes), [qr2108@columbia.edu](mailto:qr2108@columbia.edu) (Q. Razlighi), [ds3203@columbia.edu](mailto:ds3203@columbia.edu) (D. Shaked), [ts4v@virginia.edu](mailto:ts4v@virginia.edu) (T. Salthouse).

<sup>1</sup> Shared first authorship. These authors contributed equally to the manuscript.

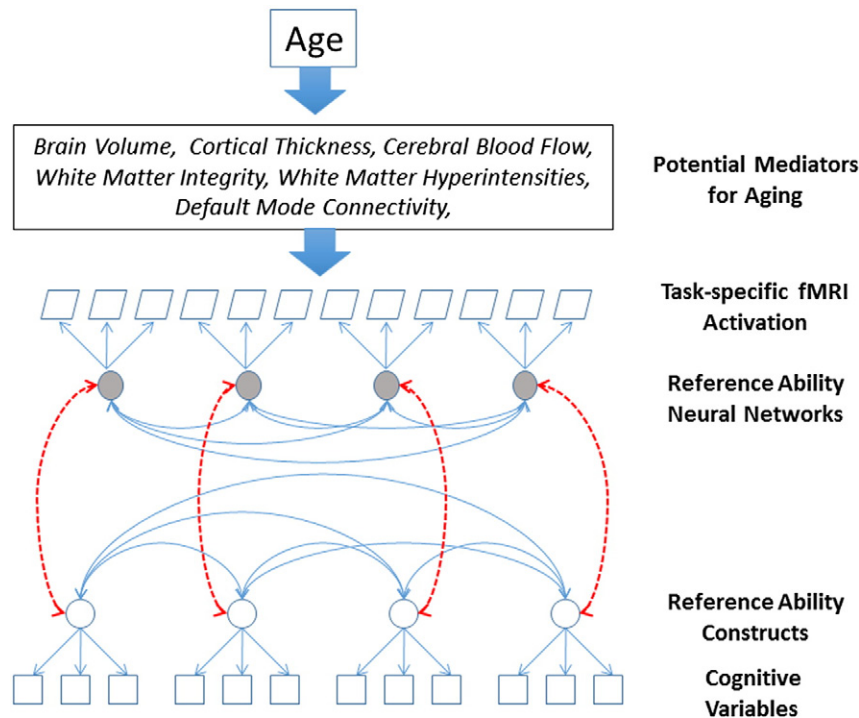
typically evaluate age differences in neural activation associated with the performance of a selected task that purportedly taps that domain. In contrast, cognitive psychologists have repeatedly demonstrated the advantages of parsimoniously representing batteries of diverse cognitive tests by sets of latent variables that each summarize performance across multiple tests. For example, based on analyses of large test batteries administered to individuals ranging from young to old, Salthouse and colleagues have consistently identified latent variables, or reference abilities (RAs) that capture the majority of the variance in age-related cognitive change across four domains: episodic memory, fluid reasoning, perceptual speed, and vocabulary (Salthouse, 2005, 2009; Salthouse et al., 2008). Based on these findings, Salthouse et al. have argued that a productive and efficient approach to cognitive aging research is to try to understand how aging impacts performance of this small set of RAs, rather than on specific tasks (Salthouse and Ferrer-Caja, 2003). Similarly, we suggest that a more productive approach to cognitive neuroimaging research in aging is to shift the emphasis from evaluating age-related differences in brain activation during performance on an individual task to studying a set of task-independent neural networks postulated to underlie the reference abilities for the four cognitive domains. This would allow us to more reliably explore the neural basis of aging's influence on key cognitive abilities.

In this report we introduce the Reference Ability Neural Network (RANN) Study. The RANN Study is designed to identify networks of brain activity uniquely associated with performance across adulthood of each of the four reference abilities described above. Using analytic approaches that parallel those used to derive latent variables from cognitive, psychometric data, we aim to determine whether spatial fMRI networks can be derived that serve as the neural substrate for the latent cognitive structure of the reference abilities as derived from behavioral data. The desired structure of relationships between reference abilities and their neural substrates is shown in Fig. 1.

Just as a reference ability can be considered to represent the common cognitive ability underlying disparate tasks, “reference ability

neural networks” (RANNs) are hypothesized to represent the common neural network that underlies the performance of each of the RAs. As in the analysis of the psychometric data, we expect the RANNs to demonstrate both convergent validity (for tasks within each cognitive domain) as well as discriminant validity (i.e. limited expression during performance of tasks in other cognitive domains). Ideally these RANNs would consist of a unique network of brain areas that are activated for each RA. Further, it would be compelling if the degree of expression or activation of a RANN was correlated with performance on the constituent tasks in that RA. Successful identification of these RANNs has the potential of contributing to a paradigm shift in research on the neural bases of age differences in cognition by emphasizing the broad and replicable aspects common to several tasks rather than the possibly idiosyncratic features of individual tasks.

The RAs that we are studying were identified in studies of several thousand individuals across the adult lifespan who have been administered extensive batteries of cognitive tasks (Salthouse, 2009). They have been modeled as latent constructs defined in terms of the common variance in four different sets of cognitive tasks representing: 1) episodic memory, 2) fluid reasoning, 3) perceptual speed, and 4) vocabulary. Although performance on the tasks that make up these RAs, and thus the levels of the abilities, differ across age, the latent constructs have been found to remain invariant across the adult lifespan. Further, although the modeled RAs are derived from performance of a limited set of tasks, these latent constructs have been found to account for the age-related performance differences on numerous other cognitive tasks (Salthouse, 2009; Salthouse and Davis, 2006; Salthouse et al., 2008). In the RANN study, we selected 12 tasks, three from each domain, that have reliably been associated with that RA. In turn, we hope to identify four RANNs, one for each cognitive ability. Once the RANNs are identified, we can investigate whether they too remain invariant over age despite different levels of performance. We can then explore the common and unique effects associated with chronological age on each of the RANNs. We also intend to examine



**Fig. 1.** Schematic sketch of brain-behavioral framework for the Reference Ability Neural Networks study: quantities given in rectangular or trapezoidal shapes are recorded in the study, while round shapes indicate both reference abilities and reference ability neural networks. The figure illustrates that not only are meaningful latent constructs expected on the basis of common patterns of fMRI activation across tasks representing the cognitive ability, but that those constructs are hypothesized to be related to the corresponding constructs defined on the basis of the correlations among behavioral performance in those same tasks. The upper boxes indicate that we expect that the effect of age on cognitive performance might be mediated by specific measures of brain aging. Further, brain changes may influence expression of the reference ability neural networks, yielding observed age-related cognitive changes.

the potential influence of several measures of structural and functional age-related brain changes on the integrity of the RANNs across the age range.

This paper introduces the RANN study, reviewing its design and implementation. We present behavioral data derived from the 12 cognitive tasks as modified for use in the MRI scanner and examine their construct validity, i.e. both (1) convergent validity, implying high correlation within reference domains, and (2) discriminant validity, implying low correlation between reference domains. We then used a straightforward GLM approach to test whether the topographies of the task-related activation of the 12 tasks also cluster by RA. Finally, we used linear-indicator regression to generate unique spatial brain maps, or RANN, for each cognitive ability, and tested the predictive utility of these RANNs for accurately classifying activation maps into the appropriate RA domain in left-out data.

## Material and methods

### Subjects

In the RANN study we plan to image 375 healthy adults (50 per decade from age 20 to 50; 75 per decade from age 50 to 80) while performing a set of 12 cognitive tasks representing four RAs (3 per cognitive domain): episodic memory, fluid reasoning, perceptual speed, and vocabulary. In this report, we present data from 174 healthy adults who have completed the study procedures. Demographic features of these participants are summarized in Table 1.

All subjects were required to be native English speakers, strongly right-handed, and have at least a fourth grade reading level. Subjects were screened for MRI contraindications and hearing or visual impairment that would impede testing. Subjects were free of medical or psychiatric conditions that could affect cognition. Careful screening ensured that the elder subjects did not meet criteria for dementia or Mild Cognitive Impairment (MCI). A score greater than 135 was required on the Mattis Dementia Rating Scale (Mattis, 1988). Further, performance was required to be within age-adjusted normal limits on a list-learning test, and participants were required to have no or minimal complaints on a functional impairment questionnaire (Blessed et al., 1968).

### Procedure

All subjects completed screening for dementia or MCI prior to participating in the remainder of the study. They participated in two 2.5-hour scanning sessions. Six tasks were administered in each session in the context of fMRI studies. One session presented three Vocabulary tasks and three Perceptual Speed tasks interspersed in a fixed order: Synonyms, Digit-Symbol, Antonyms, Letter Comparison, Picture Naming, and Pattern Comparison; and the other session presented three Episodic Memory tasks and three Fluid Reasoning tasks, also interspersed in a fixed order: Logical Memory, Paper Folding, Word Order Recognition, Matrix Reasoning, Paired Associates, Letter Sets. The

order of tasks within session was not varied, but the order of the two sessions was counterbalanced across subjects, with equal numbers having each order. We used a fixed task order because we intend to focus on individual differences in several analyses. The fixed order avoids confounding what the individual brings to the testing situation with how he/she is treated. In other words, if the order of tasks were to vary then it would be difficult to distinguish order effects from individual differences. The activation tasks were supplemented with other imaging procedures. At a separate session subjects completed a battery of neuropsychological tests as well as a set of questionnaires. These will not be discussed in the current report.

### 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 SafeVision, LLC, Webster Groves, MO). Responses were made on a LUMItouch response system (Photon Control Company). 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

In the scanner, participants performed a battery of twelve computerized tasks based on the cognitive tasks that have been used to derive the RAs that are addressed in this report. 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 RAs. In order to accommodate testing in the scanner, as well as to optimize power for image analysis, some changes in test structure and presentation were required. Prior to the scan session, computerized training was administered for the six tasks to be administered in that session. At the completion of training for each task, participants had the option of repeating the training. Each of the tasks and its implementation for use in the scanner is described below. For all tasks, except picture naming, responses were differential button presses. During training, responses were on the computer keyboard and during scans they were made on the LUMItouch response system.

### Vocabulary tests

**Synonyms (Salthouse and Kersten, 1993).** This task requires subjects to match a given word to its synonym, or to the word most *similar* in meaning. The probe word is presented in all capital letters at the top of the screen, and four numbered choices are presented below. Subjects are instructed to respond as quickly and accurately as possible. Pre-training consisted of two blocks, with two and three questions in each block, respectively. Response consisted of pressing one of four buttons corresponding to the word of choice. The entire scan was six minutes and 26 s long, with three items in each of 5 blocks for fifteen total items. The task began with a 36-second fixation cross, had five 42-second task blocks interspersed with 28-second fixation blocks. Each item was presented for 13.5 s with a 500 ms interstimulus interval (ISI) between items. The primary dependent variable was proportion of correctly identified synonyms.

**Antonyms (Salthouse and Kersten, 1993).** This task requires participants to match a given word to its antonym, or to the word most *different* in meaning. The probe word is presented in all capital letters at the top of the screen, and four numbered choices are presented below. The training, structure and timing of this task was identical to that for synonyms. The primary dependent variable was proportion of correctly identified antonyms.

**Picture naming.** This task requires subjects to verbally name pictures. Target pictures consisted of 40 colored bitmap images, adapted from the picture naming task of the WJ-R Psycho-Educational battery (Salthouse,

**Table 1**  
Participant demographics.

Decade	N	% F	Education	DRS	AMNART IQ
20–29	38	63	15.29 ( $\pm 1.96$ )	139.61 ( $\pm 2.96$ )	114.51 ( $\pm 8.34$ )
30–39	33	64	15.91 ( $\pm 2.47$ )	139.39 ( $\pm 2.65$ )	114.88 ( $\pm 10.60$ )
40–49	23	48	15.70 ( $\pm 2.40$ )	139.67 ( $\pm 2.61$ )	115.83 ( $\pm 7.57$ )
50–59	24	42	16.10 ( $\pm 1.50$ )	139.83 ( $\pm 3.41$ )	115.37 ( $\pm 8.19$ )
60–69	34	44	16.03 ( $\pm 2.69$ )	140.15 ( $\pm 2.57$ )	117.44 ( $\pm 9.69$ )
70–79	22	41	17.86 ( $\pm 2.59$ )	139.29 ( $\pm 2.66$ )	121.60 ( $\pm 7.01$ )

Note. %F = Percent female. DRS = score on the Mattis Dementia Rating Scale. AMNART IQ = IQ score estimated based on the American National Adult Reading Test. Education is measured in years. Values for EDU, DRS, and AMNART IQ are the mean ( $\pm$  standard deviation).

1998; Woodcock et al., 1989). Pre-training consisted of 2 blocks, with three and two pictures in each block, respectively. For the test phase, subjects were given VisuaStim digital headphones, and their responses in the scanner were recorded. Subjects were instructed to speak loudly and clearly, and to produce a response for every picture. The entire scan was 6 min and 16 s long, containing five 40-second blocks, with eight pictures in each block. The task began with a 36-second fixation-cross and there was a 28-second fixation time between blocks. Each picture was presented for 4.5 s with a 500-ms ISI between items. Audio recordings were filtered using a custom adaptive noise filtering procedure, and then transcribed and scored by researchers. A strict criterion for correct responses was used, with each response categorized as correct or incorrect, based on fixed criteria (Salthouse, 1998). There were two exceptions to this rule. First, a response received credit if the correct word was included in the subject's response; for example, if the subject said "printing press" for the word "press". Second, a slight mispronunciation of the correct word also received credit (e.g. "cumberbund" for cummerbund). The primary dependent variable was proportion of correctly named pictures.

#### Perceptual speed tests

**Digit symbol.** In this task, a code table is presented on the top of the screen, consisting of numbers one through nine, each paired with an associated symbol. Below the code table an individual number/symbol pair is presented. Subjects are asked to indicate whether the individual pair is the same as that in the code table using a differential button press. Subjects are instructed to respond as quickly and accurately as possible. Pre-training consisted of 2 blocks, with three and two items in each block, respectively. The entire scan was 7 min and 4 s long, containing 5 blocks, with eighteen items in each block for a total of ninety items. The task began with a 36-second fixation-cross and there was a 28-second fixation time between blocks. Each item was presented for 2.5 s with a 250-ms ISI between items. Accuracy for all three perceptual speed tasks was quite high so most of the performance variance is in the RT.

**Letter comparison (Salthouse and Babcock, 1991).** In this task, two strings of letters, each consisting of three to five letters, are presented alongside one another. Subjects indicate whether the strings are the same or different using a differential button press. Subjects are instructed to respond as quickly and accurately as possible. Pre-training consisted of 2 blocks, with three and two items in each block, respectively. The entire scan was six minutes and 26 s long, containing 5 blocks (42-s each), with twelve items in each block for a total of sixty items. The task began with a 36-second fixation-cross and there was a 28-second fixation time between blocks. Each item was presented for 3 s with a 500-ms ISI between items. The primary dependent variable is RT.

**Pattern comparison (Salthouse and Babcock, 1991).** In this task, two figures consisting of varying numbers of lines connecting at different angles are presented alongside one another. Subjects indicated whether the figures were the same or different by a differential button press. Pre-training consisted of two blocks, with three and two items in each block, respectively. Following the practice trial, subjects were given the option of undergoing the same practice trial again. The timing and item structure was the same as the letter comparison task. The primary dependent variable is RT.

#### Fluid reasoning tests

**Paper folding (Ekstrom et al., 1976).** This fluid reasoning task requires subjects to select from five options a pattern of holes that would result from a sequence of folds in a piece of paper, through which a hole is then punched. The sequence is given on the top of the screen, and the six options are given across two rows (three options in each row) below. Response consisted of pressing 1 of 6 buttons corresponding to the chosen solution. Before undergoing the pre-training, the researcher reviewed a practice problem with the participant and explained the task

and the correct solution as needed. Pre-training consisted of a minimum of four additional practice problems, with an option of answering two extra problems for more practice. Following each practice problem, the correct answer was presented on the screen, along with a description explaining why it was the correct answer. Additionally, if a subject incorrectly answered a practice problem, the researcher demonstrated the problem with a physical piece of paper and a hole punch. The entire scan was 14 min and 26 s long, containing a variable number of questions, depending on rate of task performance. Each trial began with a 24-second fixation-cross, followed by the stimulus. If a response was made in the first 11 s, the stimulus terminated at the 11 second mark. However, if a response was made after 11 s, the stimulus was terminated immediately after the response. If no response was made, the stimulus terminated after 85 s. Time-outs were coded as no response. The minimum number of trials presented was seven, which occurred when the participant required 85 s per trial, or a time-out occurred for every trial. The maximum number of trials presented was 18, which occurred when the participant required 11 or less seconds per trial. There was a 35-second ISI between items. This timing protocol allowed participants to solve these complex problems at their own rate, leading to a variable number of trials per participant. It does, however, place lower and upper bounds on the RTs to ensure an acceptable balance of task engagement and inter-trial interval rest times across the experiment. As the main objectives of the neuroimaging aspect of this study were the neural underpinnings of cognitive engagement, it was imperative to keep subjects engaged, especially the fast responders. The primary dependent variable is number of correct trials completed.

**Matrix reasoning (adapted from (Raven, 1962)).** This task requires subjects to discern a pattern from a series of figures displayed in a matrix. Subjects 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. Training procedures were the same as for the paper folding task (except for the physical paper example). Timing and item structure were the same as the paper folding task. The primary dependent variable is number of correct trials completed.

**Letter sets (Ekstrom et al., 1976).** In this task, subjects select a set of letters that do not fit with the others. Subjects are presented with five sets of letters, where four out of the five sets have a common rule (i.e. have no vowels), with one of the sets not following this rule. Subjects are instructed to select the unique set. Response consisted of pressing the one of five buttons corresponding to the chosen set. Training procedures were the same as the other fluid reasoning tasks, however before undergoing the pre-training, the researcher reviewed two practice problems with the subject as opposed to one. Timing and item structure were the same as the paper folding task. The primary dependent variable is number of correct trials completed.

**Episodic memory tests.** Note that for the episodic memory tasks, both the study and test phases were imaged.

**Logical memory.** This task requires subjects to remember specific details from stories presented on the computer screen. The subject is asked to answer detailed multiple-choice questions about the story, with four possible answer choices. Response consisted of pressing one of four buttons corresponding to the chosen answer. Pre-training consisted of one story, with two probe questions following the presentation of the story. Subjects were instructed to respond as quickly and accurately as possible. The entire scan was seven minutes long, containing two stories, with ten questions per story. The task began with a 30-second fixation-cross and there was a 30-second fixation time between stories. The story was divided into three 1 to 2 sentence sections, with each section displayed for 10-seconds. Ten seconds after completion of the story, the questions were presented for 10 s with a 2.5-second ISI between items. The primary dependent variable is proportion of correctly answered questions.



**Word order recognition.** In this task a list of twelve words is presented one at a time on the screen, and subjects are instructed to remember the order in which the words are presented. Following the word list they are given a probe word at the top of the screen, and four additional word choices below. They are instructed to choose out of the four options the word that *immediately* followed the word given above. Pre-training consisted of two lists of words, with the first list containing six words followed by five probe questions and the second list containing twelve words followed by ten probe questions. The entire scan was seven minutes and 2 s long, containing two word lists, with ten questions following each list. The task began with a 30-second fixation-cross and there was a 30-second fixation time between lists. Following the fixation-cross the first word list was presented, with each word lasting 4 s, with a varying ISI between words, ranging from 700 ms–11.4 s. Following the word list, was a 10-second fixation break, which was then followed by the questions. Each question was terminated after 6 s and at the end of the response period was a 2-second ISI between items. The primary dependent variable is the proportion of correctly answered questions.

**Paired associates.** In this task, six pairs of words are presented, one at a time, on the screen, and subjects are instructed to remember the pairs. Following the pairs, they were given a probe word at the top of the screen and four additional word choices below. Subjects were asked to choose the word that was originally paired with the probe word. Subjects were instructed to respond as quickly and accurately as possible. Pre-training consisted of one list of six pairs of words, followed by two probe questions. The entire scan was 3 min and 24 s long, containing two lists of pairs, with six probe questions for each list. The task began with a 30-second fixation-cross and there was a 10-second fixation time between lists. Following the fixation-cross the six pairs of words were presented, with each pair lasting 2 s, and with a varying ISI between words, ranging from 200 ms–5.6 s. Following the list of pairs, was a 10-second fixation break, which was then followed by the questions. Each question was terminated after 5 s and at the end of the response period was a 2-second ISI between items. The primary dependent variable is proportion of correctly answered questions.

#### Image acquisition procedures

All MR images were acquired on a 3.0 T Philips Achieva Magnet. There were two, 2-hour MR imaging sessions to accommodate the twelve fMRI tasks as well as the additional imaging modalities. At each session, first a scout, T1-weighted image was acquired to determine patient position. All scans used a 240 mm field of view. For the EPI acquisition, the parameters were: TE/TR (ms) 20/2000; Flip Angle (deg) 72°; In-plane resolution (voxels) 112 × 112; Slice thickness/gap (mm) 3/0; Slices 41. In addition, MPRAGE, FLAIR, DTI, ASL and a 7 minute resting BOLD scan were acquired. A neuroradiologist reviewed each subject's scans. Any significant findings were conveyed to the subject's primary care physician.

#### Behavioral performance variables

Behavioral performance was recorded while subjects executed the tasks in the scanner. When proportion correct trials was calculated for each task, trials where the participant did not respond (i.e. timed out) were not considered in the calculation. To ensure that we included data only from instances where subjects were performing the task, we eliminated data from any task where the participant's performance was at chance or lower. For the perceptual speed tasks, we required accuracy of 75% or greater because the focus was on the speed of performance as represented by reaction time. Z-scores were computed for all twelve behavioral variables based on the entire study group. For the perceptual speed tasks, the behavioral Z-scores were reversed in sign, such that an increasing value of the behavioral Z-score implied better performance.

#### Image analysis pre-processing procedures

Each individual's 12 fMRI scans was pre-processed in the same manner using the SPM8 software (Wellcome Trust Centre for Neuroimaging) before inclusion in the analyses described below. The processing of the functional imaging data involved two basic steps: 1) standard initial processing steps including slice timing correction, realignment, transformation into standardized (MNI) space, and smoothing and 2) statistical analysis of individual subjects' data.

Statistical analysis of individual subjects' data for the twelve cognitive tasks consisted of block-based time-series analysis of each individual task as a whole for each subject via general linear modeling of the fMRI time-series data. These data, along with the predictor variables, were high-pass filtered and pre-whitened 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 rest (i.e. no-performance, fixation periods) and were convolved with a double-gamma model of the hemodynamic impulse response function.

Confound regression for motion parameters was not performed at the subject level; however, for our RANN activation patterns we checked at the group level that pattern expression in any task did not correlate with subject motion.

#### Construct validity of behavioral and imaging data

Construct validity (CV) quantifies how behavioral data and activation patterns associated with each of twelve tasks manifest both convergent and discriminant validity. Convergent validity is indicated by strong relations among measures hypothesized to represent the same RA. Discriminant validity is high when the relations among measures hypothesized to represent different RAs, across domains, are low. A measure of CV summarizes both validity requirements as:

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

Here  $\hat{Z}_{\text{within}}$  denotes the average Fisher-Z correlation coefficient for all within-domain task pairings and  $\hat{Z}_{\text{between}}$  denotes the average Fisher-Z correlation coefficient for all between-domain task pairings. For behavioral data, task performance values were correlated between tasks, across subjects. For the neuroimaging data, mean task activation maps, explained below, were correlated between tasks, across voxels. In this manner, we derived *behavioral* and *topographic* CV measures.

We first calculated this CV across all twelve tasks, comparing the 12 within-domain correlations to the 54 between domain correlations ("global construct validity"). Subsequently, we tested the CV-difference for each cognitive domain separately ("domain-specific construct validity"). In that case, we compared the 3 within-domain correlations to the 27 between-domain correlations.

We illustrate the CV difference with a mock example for 3 tasks with each of 2 RA domains in Fig. 2.

In order to assess the significance of CV coefficients, we performed a permutation test. For this test, task assignments were permuted randomly, but subject assignments were left intact. This means that the behavioral and topographic correlation matrix for any task pair undergoes a random re-assignment of rows and columns with a subsequent computation of the CV-difference on this permuted Z-matrix. This was done for 10,000 iterations. The p-level was calculated as the fraction of the 10,000 null (i.e. permuted) CV values that exceed the CV value found for the point estimate, i.e. the value obtained with the correct task assignment.

### Analysis of functional neuroimaging data

**Evaluation of topographic similarity.** This analysis was intended to derive the topography of activation for each task in the most direct manner, in order to subject these topographies to the CV analysis derived above. This approach assesses whether task-related activation is inherently more similar within than between domain. We used straightforward GLM derive a pattern of activation for each task. Because we are interested in identifying patterns of activation that are unique to each of the four RAs, we began by attempting to identify and eliminate activation that was common to all 12 tasks. Such common activation could stem from similarities across tasks such as visual stimulation, noise, and motor demands, as well as from deactivation of networks active at rest. In order to identify common activation and de-activation across tasks we computed the 12 mean, across-subject, task-related activation patterns, i.e. one task activation map per task. That is, for every subject and task, a first-level General Linear Model regression computed a parametric activation map. These maps were then averaged across subjects for every task. The 12 subject-averaged task-activation maps were submitted to a spatial Principal Component (PC) Analysis. This means that PCA was performed on the data array  $\mathbf{Y}$ , where the number of rows equals the number of voxels  $n$  and the number of columns equals the number of tasks (or subject-averaged task-activation maps = 12). Since the true data rank of  $\mathbf{Y}$  is 12, there will be 12 Principal Components in total. Each PC is a weighted linear combination of all 12 original inputs,  $\mathbf{V}_{12 \times n} = \mathbf{W}_{12 \times 12} \times \mathbf{Y}_{12 \times n}$ , where each column of the weighting matrix  $\mathbf{W}$  is an eigenvector of the covariance matrix of the data array  $\mathbf{Y}$ .

The influence of the first PC,  $\mathbf{V}_1$  (the first row of the  $\mathbf{PC}$  matrix) is removed as follows:

$$\mathbf{Y}_{par} = (\mathbf{I} - \mathbf{V}_1^t \mathbf{V}_1) \mathbf{Y}$$

where  $\mathbf{I}$  denotes the  $n \times n$  unit matrix, and  $\mathbf{V}_1$  is the first component that has all-positive/all-negative loading on all the subject-averaged task-activation maps in the weighting matrix  $\mathbf{W}$  thus considered here as the common activation/deactivation map across all 12 tasks.  $\mathbf{V}_1^t \cdot \mathbf{V}_1$  is an  $n \times n$  projection matrix that only retains the component along  $\mathbf{PC}_1$  of any vector it is applied to.  $(\mathbf{I} - \mathbf{V}_1^t \cdot \mathbf{V}_1)$  is thus a residual projection matrix that removes from all mean images in  $\mathbf{Y}$  the component alongside  $\mathbf{V}_1$ .

We then calculated topographic correlations for each pair of tasks on the residualized task mean activation maps by calculating the correlation value across voxel locations. For a mean task activation map  $Y_1$  for task 1 and  $Y_2$  for task 2, which are columns vectors with as many rows as voxels in the brain, the simple Pearson correlation is computed as between both vectors across loadings. For instance, for a simplistic brain with just 3 voxel locations and two maps  $Y_2 = [1, -1, 1]$  and  $Y_1 = [1, 1, -1]$ , we would obtain a spatial correlation coefficient of  $R = (Y_1, Y_2) = 0.50$ . This correlation captures the spatial similarity of the activation pattern for two tasks.

We also tested for age-related differences in CV in both topographic and behavioral data. For age comparisons, the permutation test mixed up age-group membership along a median age-split, while leaving task-domain assignments intact. The p-level was calculated as the fraction of the 10,000 null (i.e. permuted) CV-difference values that exceed the CV-difference between the unpermuted age groups for the point estimate.

**Derivation of RANNs with linear-indicator regression.** The subsequent analysis intended to use a multivariate approach to derive 4 RANNs that were best associated with the 3 tasks in each RA. Consistent with our initial goal to use a method that used as few constraints as possible to derive RANNs, we used a linear-indicator regression approach (Hastie et al., 2009). This approach decomposes activation in each task to a set of PCs and then derives the optimal combination of PCs that

discriminates between the 3 tasks in a RA and the other 9 tasks. First, a PC Analysis was run on all  $174 * 12 = 2088$  subject-and-task parametric maps, and the individual pattern scores, or Subject Scaling Factors (SSF), for the first 500 PCs were obtained by an inner product of all 500 PCs with the 2088 subject and task maps. Concretely, the pattern score matrix  $\mathbf{SSF}$  is computed with the following multiplication,

$$\mathbf{SSF}(i, k) = \mathbf{Y}_i \mathbf{V}_k^t$$

where  $i$  denotes the subject-task index runs from 1 to 2088,  $\mathbf{Y}_i$  now represents the activation maps (= row vector) for each particular task and subject, and  $k$  indicates the PC index, running from 1 to 500. The SSFs were then used as independent variables in a subsequent linear-indicator regression (Hastie et al., 2009) to predict an indicator matrix  $\mathbf{I}$ .  $\mathbf{I}$  had 2088 rows and 4 columns and places a value of 1 in the appropriate column depending on the RA domain that the scan indexed by the row position belongs to. Each column in indicator matrix  $\mathbf{I}$  thus contained  $3 * 174 = 522$  unit entries on different rows, and each row contained only a single unit entry. Summing over all entries in  $\mathbf{I}$  recovers the total number of scans in the analysis: 2088. The regression equation can be written as

$$\mathbf{I} = [\mathbf{SSF}(:, 1 : k) \mathbf{1}] \mathbf{B} + \text{error}$$

where  $\mathbf{SSF}(:, 1:k)$  is the array of pattern scores for the first  $k$  PCs, and  $\mathbf{1}$  denotes an intercept term.  $\mathbf{B}$  is an array of regression weights of format  $(k + 1) \times 4$ . The corresponding 4 RANNs were constructed by applying the regression weights to the PCs according to

$$\mathbf{RANN} = \mathbf{V}(:, 1 : k) \mathbf{B}(:, 1 : 4)$$

where  $\mathbf{V}$  contains the PCs as column vectors and has as many rows as voxels in the analysis.

To select an optimum set of PCs, i.e. the best number  $k$  of included PCs, we used a goodness-of-fit measure, the AIC criterion (Burnham and Burnham, 2002), computed for each of the 4 dependent variables in the indicator matrix to arrive at an average value for each set of PCs. AIC picks an optimal bias-variance tradeoff and minimizes the residual sum of squares, while keeping the number of parameters in the model at a minimum. We picked  $k$  according to the AIC criterion, i.e. we varied  $k$  from 1 to 500, running the linear-indicator regression each time, and chose the optimal for  $k$  according to the result of the regression. For the case that several very similar local minima in the AIC curve were present, we had decided beforehand to take the set with the minimum number of PCs, to keep the variance contribution in the data as large as possible.

Once  $k$  was determined, we performed the linear-indicator regression for the full sample and computed the RANNs. To determine the robustness of RANN voxel-loadings, we performed a semi-parametric bootstrap resampling procedure (Efron and Tibshirani, 1993) with 500 iterations, which resampled from the full 2088 scans with replacement, each time performing the derivation of the RANNs. The variability of the voxel loadings in the bootstrap resampling procedure around the point estimate values can be approximated as a Z value at voxel location  $x$  according to the formula

$$Z(x) = \mathbf{RANN}(x) / \text{bootstrap-STD}(x)$$

Robust loadings fulfill  $|Z| > 2$  and are visualized in the four RANN images.

Next, we investigated the ability of the derived RANNs to predict the underlying RA domain for any individual subject's activation out of sample, using repeated randomized 10-fold cross-validation (Hastie et al., 2009). In this procedure, the reference-domain label is predicted for a portion of the data (10%) that have been left out of the derivation of the 4 RANNs in a training subset that contains 90% of the data. If scan

**y** was left out of the RANN derivation, a prediction of the reference label can be made according to

$$\mathbf{L} = [\mathbf{yV}(1 : k, :)'] \mathbf{B}$$

**L** is  $1 \times 4$  row vector and contains the degree to which the scan loads onto each RANN, while **V** and **B** were determined from the training-data fold. The biggest loading determines the predicted reference-domain label. The procedure was repeated 100 times for different random assignments of training and test data. Two confusion matrices are computed: (1) a domain-wise  $4 \times 4$  confusion matrix for which rows indicate the gold-standard reference-domain label and for which columns denote the predicted label; (2) a  $12 \times 4$  confusion matrix where rows indicate the gold-standard knowledge of task identity and columns, again, indicate the predicted reference-domain label. Both types of confusion matrix are helpful: (1) gives a global view of the level and type of misprediction for each reference domain; (2) gives such information at the task-level and can identify particularly problematic tasks.

## Results

### Behavioral results

The number of participants whose behavioral data was either missing or below chance accuracy was as follows: Logical Memory—3; Word Order—16; Pair Associates—11; Matrix Reasoning—21; Letter Sets—3; Paper Folding—16; Digit Symbol—9; Letter Comparison—26; Pattern Comparison—4; Antonyms—15; Picture Naming—15; Synonyms—10.

There were strong age-relationships for all RAs. Examining the mean Z-score for tests within each cognitive domain, there was a significant decline with age on performance of episodic memory (correlation with age,  $r = -0.34$ ,  $p < 0.0001$ ) and fluid reasoning tasks ( $r = -0.25$ ,  $p = 0.0027$ ). Similarly RT on the perceptual speed tasks increased with age ( $r = 0.58$ ,  $p < 0.0001$ ). In contrast performance on the vocabulary tasks increased with age ( $r = 0.40$ ,  $p < 0.0001$ ). These results are compatible with previous observations (Salthouse and Davis, 2006).

Fig. 3 summarizes the inter-task correlations of the behavioral data. The global CV value averaged over all RAs was 0.6221, indicating stronger within- than between-domain correlations. In a permutation test for 10,000 iterations, this CV was highly significant at  $p < 0.0001$ , indicating that, overall, the clustering of behavioral performance into RA domains is very unlikely to have occurred by chance. We then examined domain-

specific CV for each individual ability comparing the three within-domain correlations to the 27 across-domain values. From the visual inspection of the Z-matrix, the episodic memory domain displays substantial overlap with the fluid reasoning domain, suggesting poor discriminant validity. This was born out in the permutation test, which gave the following CV values and associated p-levels: episodic memory: 0.31 ( $p = 0.0790$ ); fluid reasoning: 0.63 ( $p = 0.0059$ ); perceptual speed: 0.9629 ( $p < 0.0002$ ); vocabulary: 0.59 ( $p = 0.0102$ ).

In order to determine whether behavioral CV was present in both younger and older subjects, we divided the subjects into two groups based on median age and assessed CV within and between the two age groups. The results are summarized in Table 2. Global CV was present in both groups, but the coefficient was stronger in young than in old. There were some differences between young and old in the strength of within-domain CVs for individual RAs. This was most notable for perceptual speed and fluid reasoning, which were both stronger in the younger group.

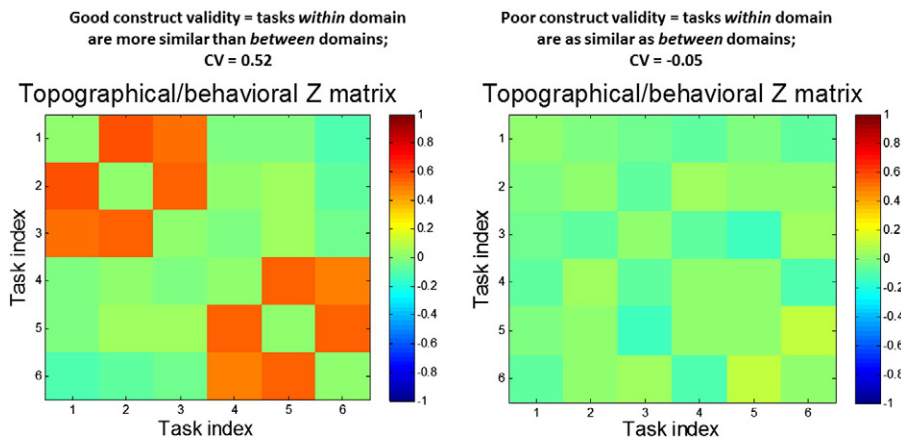
### Imaging analyses

#### GLM analysis

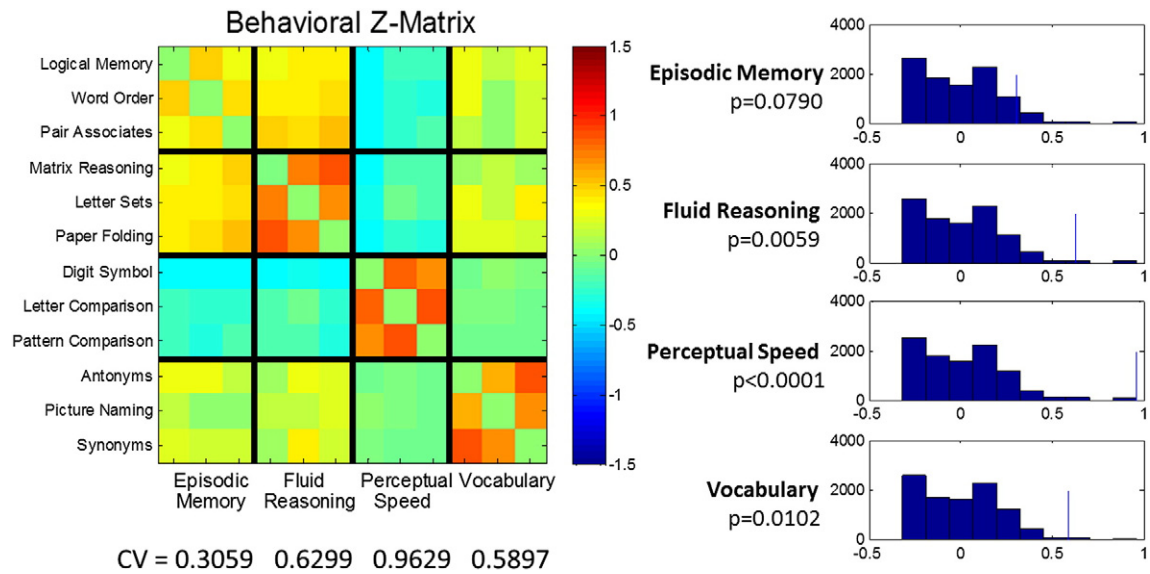
As described above, we first used GLM analyses to derive 12 subject-averaged activation maps, 1 for each task. These 12 maps were then submitted to PCA that produced 12 PCs. We removed the influence of the first PC from all subject and task imaging data, surmising that task-related activation and de-activation common to all tasks would be removed from the data in a non-differential manner. Details of task-timing and the proportion of time per block that occupies task performance should not influence the topographic of the first principal components, but might influence the level of pattern manifestation.

After removing the influence of the first PC from all subject and task maps, the residual maps can be averaged across subjects within task to produce 12 mean images with the common activation removed. We then computed the topographic correlations across all possible task-pairs, and these correlations summarize the spatial similarity of task-related activation between any pair of tasks. These topographic correlations are summarized in Fig. 4.

The CV that summarizes global CV yielded a value of 0.6297, indicating greater topographic similarity within RAs than between them. A permutation test with 10,000 iterations using CV as a test statistic yielded highly significant results ( $p < 0.0001$ ). We then examined comparing the 3 within-domain correlations to the 27 across-domain values. The domain-specific CV values were as follows: episodic memory: 0.92 ( $p = 0.01$ ); fluid reasoning: 1.01 ( $p = 0.005$ ); perceptual



**Fig. 2.** Schematic demonstration of good and bad Construct Validity (CV) for 2 reference domains, each described by 3 tasks. Left: very good global CV, performance of tasks within a reference domain are more similar than between-reference domains. The Z-matrix reveals a block-diagonal structure. The observed CV statistic, the difference of mean Z-values within a reference ability to mean Z-values between reference ability values quite high. Right: poor global CV, pairs of tasks within and between reference domains show equal similarity, thus yielding CV value close to zero. The diagonal elements in every Z matrix have been set to zero since for perfect correlation Z becomes infinite.



**Fig. 3.** Investigation of behavioral Construct Validity (CV). The matrix of inter-task performance correlations ( $=Z$ ) is plotted on the left. Note that for the most part correlations are stronger between tasks within a reference ability than across tasks in different reference abilities. The observed CV statistics, the differences of mean Z-values *within* a reference domain to *between*-reference domains are displayed below the Z-matrix. Null-histograms for a permutation tests with 10,000 iterations are displayed on the right, with the point estimates marked by a thin blue line. All domains show significantly higher- within versus between-domain correlations, apart from the episodic memory domain. One can appreciate cross-talk from the Z-matrix; the permutation test bears this out with a marginal p-value of 0.0571.

speed: 0.30 ( $=0.15$ ); vocabulary: 0.29 ( $p = 0.17$ ). The lack of significant topographic CV for the speed and vocabulary domain is evident visually from the Z-matrix: both domains have poor CV and with large correlations between the Letter Comparison and Pattern Comparison tasks on one side, and the Antonyms and Synonyms tasks on the other side.

In order to determine whether topographic CV was present in both younger and older subjects, we again divided the subjects into two groups based on median age and assessed CV within and between the two age groups. The results are summarized in Table 2. Global CV was present in both groups. Also there were no significant differences between young and old for domain-specific CV.

#### Linear-indicator regression analysis

**RANN derivation.** Linear-indicator regression analysis was used to derive a PC associated specifically with each of the four RAs. The first 174 PCs (out of 2088) were chosen to construct the four RANN patterns because they yielded a local minimum of the Akaike Information Criterion, while securing good performance in the randomized 10-fold cross-validation procedure. The four RANNs are illustrated in Fig. 5 and described in Tables 3–6.

**Prediction of RA membership.** We next investigated the ability of the identified RANNs to predict the RA associated with the activation of individual subjects for individual tasks. These data were derived from a

10-fold cross validation procedure in which RANNs were derived from 90% of all subject-task data, and used to predict the RA associated with the remaining 10% of subject-task data. Median accuracy rates for associating component task activation with a particular RA were as follows: episodic memory: 82%; fluid reasoning: 87%; speed: 84%; and vocabulary: 77%. We also examined the breakdown of RA attribution for each individual task. That is, the percentage of time that activation for an individual task was labeled with the appropriate RA, versus the percentage of time labeled with one of the other three RAs. These findings are summarized in Table 7. Individual scans were associated with the correct RA between 74 and 87% of the time. Classification appeared to be most consistent for the fluid reasoning domain. It is of interest to observe the patterns of misclassifications. For example, fluid reasoning misclassifications did not systematically fall in any other domain. In contrast, when episodic memory tasks were misclassified, they were most often mislabeled as fluid reasoning. In addition vocabulary tasks were most often misidentified as perceptual speed. To some degree the perceptual speed tasks were also misidentified most often as belonging to the vocabulary domain.

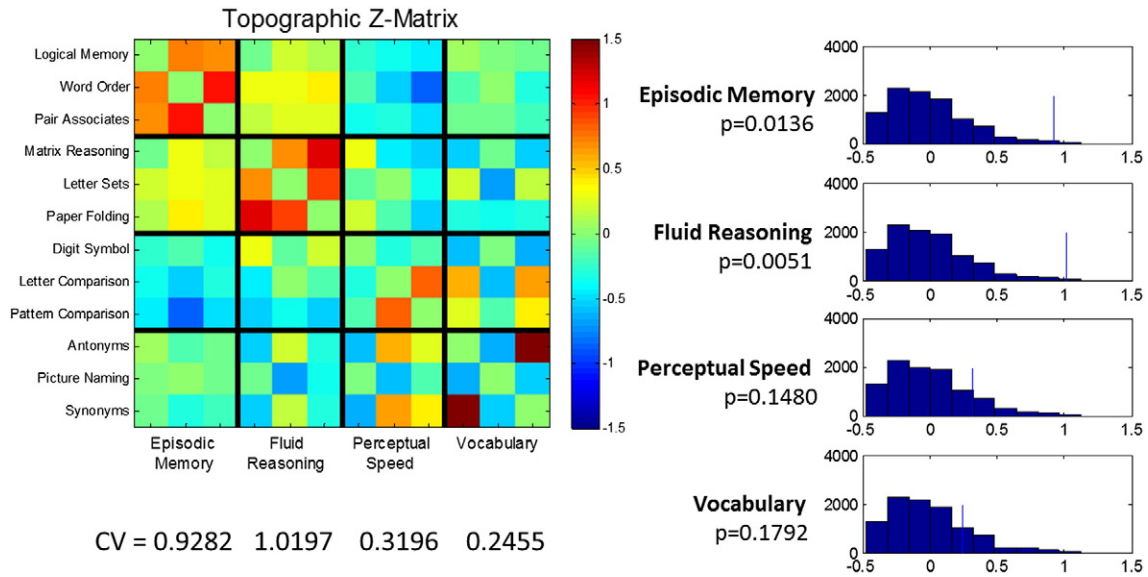
#### Discussion

This report introduces the Reference Ability Neural Network (RANN) study, an ambitious project designed to establish the neural substrates of four cognitive RA domains. Within the cognitive psychology literature it has previously been demonstrated that age-related variance in

**Table 2**  
Age-specific findings for global and within-domain construct validity (CV), for behavior and topography of activation. Age groups were divided at median age. Also provided is a test of the difference in magnitude of the CV between young and old CV ( $\Delta CV$ ). Significant results at  $p < 0.05$  are in bold.

	Memory	Fluid	Speed	Vocabulary	Global
<i>Behavioral construct validity</i>					
Old CV	0.25, $p = 0.08$	<b>0.44, <math>p = 0.02</math></b>	<b>0.74, <math>p &lt; 0.01</math></b>	<b>0.58, <math>p &lt; 0.01</math></b>	<b>0.50, <math>p &lt; 0.01</math></b>
Young CV	0.19, $p = 0.21$	<b>0.64, <math>p &lt; 0.01</math></b>	<b>0.99, <math>p &lt; 0.01</math></b>	<b>0.44, <math>p = 0.02</math></b>	<b>0.56, <math>p &lt; 0.01</math></b>
Young-Old $\Delta CV$	-0.06, $p = 0.70$	0.20, $p = 0.06$	0.25, $p = 0.10$	-0.14, $p = 0.83$	0.21, $p = 0.41$
<i>Topographic construct validity</i>					
Old CV	<b>0.82, <math>p = 0.01</math></b>	<b>0.99, <math>p &lt; 0.01</math></b>	0.32, $p = 0.12$	0.17, $p = 0.23$	<b>0.57, <math>p &lt; 0.01</math></b>
Young CV	<b>0.63, <math>p = 0.03</math></b>	<b>0.76, <math>p = 0.02</math></b>	0.39, $p = 0.06$	0.27, $p = 0.17$	<b>0.51, <math>p &lt; 0.01</math></b>
Young-Old $\Delta CV$	-0.19, $p = 0.82$	-0.23, $p = 0.80$	0.08, $p = 0.38$	0.10, $p = 0.29$	-0.06, $p = 0.73$



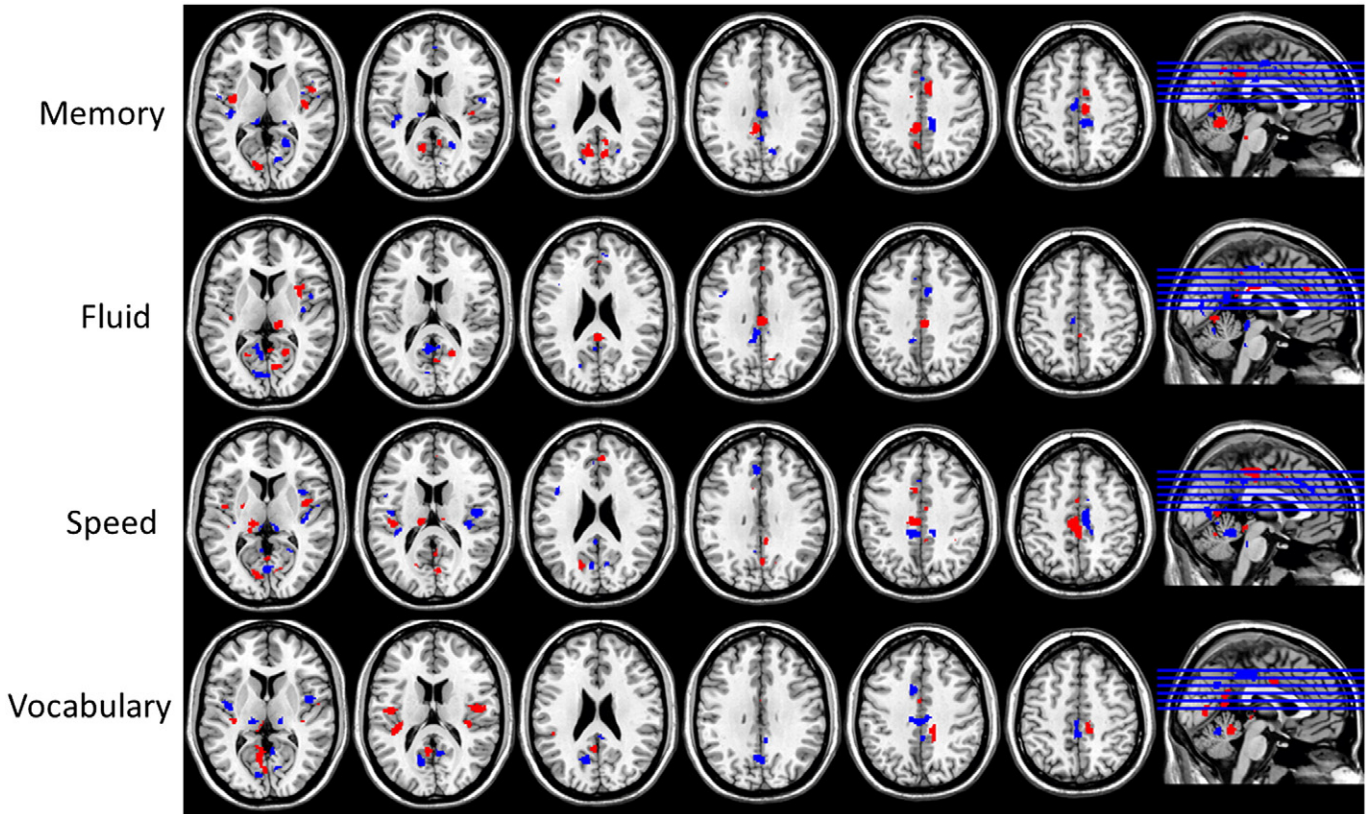


**Fig. 4.** Investigation of topographic construct validity of mean univariate task-activation patterns, similar to display of behavioral construct validity in Fig. 3. For the mean activation patterns, the SPEED and VOCAB domains yielded poor convergent and discriminant validity, resulting in non-significant p-levels.

the majority of cognitive tasks can be accounted for using the latent variables representing four RAs: episodic memory, fluid reasoning, perceptual speed, and vocabulary. Performance on tasks within each RA is more highly correlated than is performance on tasks across RAs. We reason that this pattern of convergent and discriminant validity must have a neural basis, and hypothesize that we can identify unique brain networks associated with each RA. The current study was designed to detect these RANNS. In addition, the RANN study aims to delineate whether and how the RA neural networks change in expression or

composition with age and in turn whether these network changes account for age-related changes in their corresponding cognitive domains.

We chose to focus on these specific four reference abilities because studies by Salthouse and others, conducted on several thousand individuals, have demonstrated that the twelve selected tasks represent four latent constructs (RAs) whose structure remains stable across aging with excellent convergent and discriminant validity. This provides us the ideal conditions to identify the neural networks underlying these RAs. It should be noted that similar structure of latent cognitive abilities



**Fig. 5.** Topographic activation patterns derived from the linear-indicator regression. Shown are the robustness Z-maps obtained from a bootstrap procedure with 500 iterations. Shown in blue are areas with loadings  $Z < -2$ , indicating negative task-activation, while areas with loadings  $Z > 2$ , indicating positive task-activation, are shown in red.

**Table 3**  
Brain regions for the Perceptual Speed pattern.

X	Y	Z	Brain structure	BA	Z-value	CS
<i>Positive</i>						
–6	–24	46	Cingulate gyrus	31	4.83	328
14	–50	–22	Culmen of cerebellum	*	6.22	290
12	–74	–4	Lingual gyrus	18	5.20	162
–12	–22	12	Thalamus	*	5.34	145
8	–26	–4	Thalamus	*	4.12	132
–10	–80	0	Lingual gyrus	18	5.03	123
44	0	10	Insula	13	3.01	94
–42	–24	14	Transverse temporal gyrus	41	3.38	89
–28	–58	–12	Fusiform gyrus	19	3.50	62
–18	–70	24	Precuneus	31	3.08	57
–8	10	36	Cingulate gyrus	24	2.59	53
<i>Negative</i>						
–10	–54	–18	Culmen of cerebellum	*	–5.82	360
2	–72	4	Cuneus	30	–4.41	256
44	–16	10	Insula	13	–3.64	217
8	–20	48	Cingulate gyrus	31	–4.74	192
–4	30	36	Medial frontal gyrus	9	–3.93	118
–2	–36	38	Cingulate gyrus	31	–2.82	96
–6	–30	6	Thalamus	*	–3.13	83
18	–64	–18	Declive of Cerebellum	*	–3.84	53
–18	–46	–4	Parahippocampal Gyrus	30	–2.67	51

Note. X, Y, Z are the MNI coordinates. BA = Brodmann Area, CS = cluster size, \* = BA not available.

has been found with variables from the Woodcock Johnson test (Salthouse, 1998), WAIS-IV and WMS-IV batteries (Salthouse, 2009), and other assorted variables (Salthouse and Davis, 2006), indicating that these latent variables effectively describe cognitive aging. Further, although the modeled RAs are based on performance of twelve specific tasks, these latent constructs have been found to account for the age-related performance differences on numerous other cognitive tasks (Salthouse, 2009; Salthouse and Davis, 2006; Salthouse et al., 2008) including tasks often assumed to reflect executive functioning such

**Table 4**  
Brain regions for the Episodic Memory pattern.

X	Y	Z	Brain structure	BA	Z-value	CS
<i>Positive</i>						
–20	–48	–24	Culmen of cerebellum	*	5.99	572
–14	–68	–6	Lingual gyrus	18	2.97	–
–6	–42	38	dACC	31	4.41	205
–12	–62	20	Precuneus	31	4.07	186
6	4	42	dACC	32	3.50	124
8	–20	48	Paracentral lobule	31	4.15	99
40	–14	12	Insula	13	3.60	98
–8	–30	–4	Thalamus	*	4.17	95
20	–40	–4	Lingual gyrus	30	3.12	92
16	–50	–6	Culmen	*	2.90	92
–36	–8	6	Clastrum	*	3.58	74
46	2	6	Insula	13	4.21	68
8	–52	20	Posterior cingulate	30	2.64	65
<i>Negative</i>						
14	–52	–22	Culmen of cerebellum	*	–4.74	379
–38	–24	10	Insula	13	–3.03	186
8	–32	46	Paracentral lobule	31	–3.27	153
22	–56	16	Posterior cingulate	30	–3.06	142
16	–66	32	Precuneus	7	–2.81	–
2	–66	–4	Culmen of cerebellum	*	–3.09	101
0	–24	30	Cingulate gyrus	23	–3.11	88
12	–74	6	Cuneus	23	–3.56	68
–10	–40	–10	Culmen of cerebellum	*	–3.11	65
–4	–18	46	Cingulate gyrus	24	–2.88	64
4	–26	2	Thalamus	*	–2.84	57
40	14	2	Insula	13	–3.17	51

Note. X, Y, Z are the MNI coordinates. BA = Brodmann Area, CS = cluster size, \* = BA not available.

**Table 5**  
Brain regions for the Fluid Reasoning pattern.

X	Y	Z	Brain structure	BA	Z-value	CS
<i>Positive</i>						
2	–66	–4	Culmen of Vermis	*	3.22	269
12	–72	10	Cuneus	30	3.10	–
4	–54	6	Posterior cingulate	30	2.79	–
0	–24	28	Cingulate gyrus	23	4.61	226
–12	–58	–18	Declive of cerebellum	*	4.49	137
14	–30	4	Thalamus	*	3.50	106
26	–48	–14	Parahippocampal gyrus	37	3.39	90
20	–58	14	Posterior cingulate	30	3.08	86
36	12	8	Insula	13	2.93	76
0	30	30	Cingulate gyrus	32	2.84	55
2	–40	24	Posterior cingulate	23	3.03	51
<i>Negative</i>						
–4	–56	14	Posterior cingulate	23	–3.61	276
–30	–50	–22	Culmen of cerebellum	*	–3.31	181
–6	–80	14	Cuneus	18	–2.98	124
–8	–44	36	Cingulate gyrus	31	–3.32	122
–6	–26	54	Medial frontal gyrus	6	–4.04	116
–6	–30	–6	Thalamus	*	–3.69	109
6	–4	36	Cingulate gyrus	24	–2.62	99
16	–48	–4	Lingual gyrus	19	–3.13	89
24	–58	–22	Declive of cerebellum	*	–3.01	60

Note. X, Y, Z are the MNI coordinates. BA = Brodmann Area, CS = cluster size, \* = BA not available.

as the Wisconsin Card Sort Test, Tower of Hanoi, figural, letter, and category fluency, Stroop, flanker, stop-signal, anti-cue, zoo map, mazes, and both complex span and updating working memory tasks. In subsequent studies we can assess the degree to which other types of tasks, including executive function tasks, are represented by the 4 RAs. Alternately, we could attempt to develop RANNs for other potential RAs.

In the current report, we describe the RANN study design and procedures, and present some preliminary data supportive of the idea that

**Table 6**  
Brain regions for the Vocabulary pattern.

X	Y	Z	Brain structure	BA	Z-value	CS
<i>Positive</i>						
–4	–56	14	Posterior cingulate	23	4.12	465
–16	–48	4	Parahippocampal gyrus	30	3.61	–
0	–50	–16	Culmen of cerebellum	*	3.56	–
–44	–40	20	Insula	13	4.49	165
14	–32	48	Paracentral lobule	5	4.28	135
–42	14	22	Middle frontal gyrus	46	5.31	116
–52	–12	16	Postcentral gyrus	43	4.97	–
–48	6	16	Broca's area	44	4.19	–
54	–8	16	Postcentral gyrus	43	4.23	102
42	–14	20	Insula	13	2.73	–
18	–64	–18	Declive	*	4.17	85
–8	–32	8	Thalamus	*	3.14	57
<i>Negative</i>						
0	–36	46	Precuneus	7	–3.55	252
–2	–26	44	Paracentral lobule	31	–3.22	–
14	–72	–6	Lingual gyrus	18	–4.60	249
–14	–64	20	Precuneus	31	–3.74	143
–14	–50	–22	Culmen of cerebellum	*	–3.66	130
–12	–76	2	Lingual gyrus	18	–4.12	115
8	–46	2	Culmen of cerebellum	*	–2.98	114
4	–56	10	Posterior cingulate	29	–2.81	114
12	–24	8	Thalamus	*	–3.53	89
–42	–6	8	Insula	13	–3.64	86
4	–62	–16	Culmen of cerebellum	*	–3.50	84
–12	–24	10	Thalamus	*	–3.57	82
0	–64	34	Precuneus	7	–3.37	76
36	–8	–10	Clastrum	*	–3.15	52

Note. X, Y, Z are the MNI coordinates. BA = Brodmann Area, CS = cluster size, \* = BA not available.

**Table 7**

Left-out data prediction success of the reference-domain label derived from linear indicator regression. The first 2 columns give the gold-standard identities of reference label and task. The last 4 columns give the fraction of predictions for the 4 reference labels. Every row has been normalized such that it sums to unity. Highlighted in bold face are correct predictions. For example activation on the logical memory task was correctly identified as belonging to the episodic memory domain 70% of the time, and was misclassified as belonging to fluid reasoning 20% of the time.

Actual domain and task		Predicted domain			
Domain	Task	Memory	Fluid	Speed	Vocabulary
Memory	Logical memory	<b>75%</b>	16%	3%	5%
Memory	Word order	<b>87%</b>	10%	1%	3%
Memory	Paired associate	<b>86%</b>	7%	2%	4%
Fluid	Matrix reasoning	5%	<b>84%</b>	9%	1%
Fluid	Letter sets	9%	<b>87%</b>	2%	2%
Fluid	Paper folding	6%	<b>84%</b>	9%	1%
Speed	Digit symbol	4%	4%	<b>87%</b>	5%
Speed	Letter comparison	2%	5%	<b>83%</b>	10%
Speed	Pattern comparison	2%	7%	<b>83%</b>	8%
Vocabulary	Antonyms	6%	6%	14%	<b>74%</b>
Vocabulary	Picture naming	3%	4%	7%	<b>85%</b>
Vocabulary	Synonyms	0%	5%	9%	<b>86%</b>

RANNs can be identified. Our intention was to begin by investigating whether expected behavioral and brain activation relationships could emerge naturally from the data without the use of brain-behavioral constraints. Thus we first examined the statistical correlation of behavioral performance within RA domains compared to that across domains. Overall, we found that the behavioral variables within each RA domain showed a statistically larger correlation within than between RA domains. While this clustering of the behavioral data within domains was expected based on extensive previous studies, all of the tasks had been modified to suit the demands of presentation in the scanner and to maximize power of the analysis of the BOLD data. This included inserting large gaps of time between items and blocks. Further, some tasks were more radically altered. Most notably, the form of all of the episodic memory tasks differed from their standard paper and pencil presentations. For example, for Word Order we replaced simple verbal recall with multiple-choice items querying the order of words in the lists. When we examined CV for each individual domain, comparing the 3 within-domain correlations to the 27 across-domain values, we found significant CV for fluid reasoning, perceptual speed and vocabulary, while CV for episodic memory reached only borderline significance. This may be due to stronger than expected correlation between performance in some memory and fluid reasoning tasks. No within domain-correlation was significant only in one age group. Global CV was significant in both young and old, without showing a significant age-related difference. The same findings occurred for Fluid Reasoning, Perceptual Speed, and Vocabulary. We will continue to monitor these associations as we collect more data. We will also consider alternate scoring schemes for the episodic memory tasks; at this point we simply compute the proportion correct. Finally, as more data is collected, we will use more advanced structural equation approaches to model these data.

We next attempted to see if good CV could also be demonstrated in activation maps using the most straightforward image-analytic and statistical approaches. The idea was to see how well CV emerged with the minimum amount of statistical constraints. That is, our initial approaches did not force CV, and minimized manipulation of the imaging data. For our initial examination of this issue we simply derived an activation map for each task using GLM with a block design after removing task-invariant activation from the total data array. The topography of the activation maps of tasks within-domain were more strongly related to each other than those of tasks across-domain. However when we examined the domain-specific CV of each RA by comparing the 3 within-domain correlations to the 27 across-domain values, only episodic memory and fluid reasoning showed significant validity.

There appeared to be large cross-domain relationships between the topography of activation for the perceptual speed and vocabulary tasks. This suggests that simple analytic approaches may hint at segregation of activation within each of the RAs, more advanced multivariate approaches are required to these derive the RANNs. After dividing the subjects into two groups based on median age, we found good global topographic CV in each group, and the domain-specific CVs did not differ by age. This suggests that we should be able to identify RANNs that are stable across age.

It should be noted that our CV measure for the behavioral measures reflects the degree to which different people exhibit more similar levels of performance for tasks within the same domain than for tasks between different domains, whereas for the task activation it reflects the degree to which different voxels exhibit more similar average levels of activation (collapsed across people) for tasks within the same domain than for tasks between different domains. This is potentially important because not only are different types of measures being evaluated (performance vs. average activation), but the comparisons are across different types of entities (people vs. voxels). We did do an analysis assessing CV for activation which yielded similar results, but we felt that it was too complex to present here.

Because the basic GLM analysis did not yield good domain-specific CV for each RA, we moved to a more directed approach at deriving a unique RANN for each RA. We chose linear-indicator regression analysis for this purpose because it simply tries to derive a linear combination of PCs that best associated with each RA, and does not require used of behavioral data. The regression analyses were successful in identifying a unique pattern associated with each of the four RAs. Each contained areas that were very reliably associated with its RA.

Significant regions in the four RANNs showed good correspondence with the respective processes involved in performing each RA. Broca's area (Brodmann area 44) was part of the vocabulary pattern, as expected for a skill heavily reliant on language (Bak et al., 2001). The same can also be said with bilateral insular activation (Oh et al., 2014). Oh and colleagues conducted a meta-analysis of language-related fMRI studies showing that insular activation, predominantly in the left hemisphere, was found for expressive language. The largest cluster of activation in the episodic memory pattern resided in the cerebellum. Involvement of the cerebellum in higher cognitive function has been well documented in both human (Stoodley et al., 2012; Strick et al., 2009) and animal studies (Schmahmann, 1996). In one study (Stoodley et al., 2012), cerebellar activation was found for an n-back task, a task that relies heavily on memory processes. The perceptual speed pattern is dominated by activation in the cerebellum and bilateral thalamus and visual cortices. Note that visuomotor processes were necessary for the speeded tasks. The thalamus (Saalmann and Kastner, 2011) and visual cortices (Wurtz and Kandel, 2000) participate in various levels of visual processes that are essential to translating visual stimuli into appropriate motor response. For the fluid reasoning pattern, activation was found in the dorsal anterior and posterior cingulate cortex. The anterior cingulate has been associated with cognitive control towards achieving a goal (Shenhav et al., 2013) and the dorsal anterior cingulate was found to specifically monitor changes in the environment to allow efficient adaptation to task at hand (Sheth et al., 2012). The function of the posterior cingulate cortex has been proposed to be control of arousal, internal or external direction of attention, and attentional focus (Leach and Sharp, 2014).

It is important to note that the RANN patterns were derived for maximal discriminability among the RAs, which is different from the usual functional analysis that derives a set of brain regions with greater activity than at rest. Thus, brain regions with activities that are distinct to the RA are more likely to be shown in the ability patterns, while brain regions common to more than one of the abilities are less likely to be significant. For example, it is expected that the medial temporal lobe participates in memory tasks, and activation in this area was significant in univariate t-tests in the GLM contrasts for each memory task



(not reported here). However, activation in this area may not have discriminated memory from the other each RAs well, perhaps due to activation of this region in some other tasks. Other analytic approaches may allow for RANNs that fully incorporate areas important for mediating their RA while still maintaining discriminant validity.

We next tested these “RANN candidates” using a leave one out procedure, by re-deriving them in 90% of the data and assessing for each scan in the remaining 10% of data the RANN that it loaded on to. This procedure was repeated multiple times and the frequencies with which each scan was successfully identified or summarized. Overall, the derived RANNs were quite successful in identifying the RA membership of individual scans. The misidentifications were also informative. The vocabulary tasks were most often mislabeled as perceptual speed tasks. Perhaps some subjects felt the need to respond quickly to the multiple-choice tasks. Notably, this is compatible with the observed correlations between performance on vocabulary and speed tasks in the behavioral data. Episodic memory tasks were most often mislabeled as fluid reasoning tasks. One could speculate that the recognition items had some strategic demands similar to those in the fluid reasoning tasks. More generally these misidentifications can represent the shared cognitive demands across RAs that is removed in structural equation modeling analyses by including a common latent variable. Future analyses could take two directions. First, we could initially remove common variance across task-related activation as we did in the GLM analyses. Alternately, we could adjust the RANN indicators based on shared or unique variance.

Previous reports have investigated cognitive aging fMRI-tasks and sought to provide unifying explanations with domain-specific networks that otherwise show invariance with respect to content generalized beyond particular tasks (Burianova et al., 2010). One recent study (Hampshire et al., 2012) administered a set of 12 distinct cognitive tasks in a small fMRI study ( $n = 16$ ). A factor model of the mean images from the 12 tasks yielded two functional networks roughly corresponding to working memory and fluid reasoning components. A similar factor structure was noted in behavioral data from the same tasks administered online to a large sample of subjects. This approach differs from our attempt to identify a RANN underlying a predetermined set of latent variables. Several studies have focused on tasks within the domain of memory. One study (Nyberg et al., 2002) used a partial least squares (PLS) technique with 3 memory fMRI tasks administered to 13 young adults to extract spatial patterns of activity associated with latent variables thought to reflect distinct memory systems including working memory, episodic memory, and autobiographical memory. Another study (Grady et al., 2006) similarly used a voxel-wise contrast of covariance between fMRI signal from memory tasks and age in a sample of 35 healthy adults spanning from 20 to 87 years of age and found that they could derive a common network of activation that covaries with age. The same group (Burianova and Grady, 2007) again conducted a study of the common fMRI networks underlying autobiographical, episodic, and semantic memory, as well as the distinct neural networks associated with each of these functions using PLS on a sample of 12 (21–37 years old). They (St-Laurent et al., 2011) later showed that while the common activation pattern expressed with these three components of memory persists across age, the unique spatial covariance patterns associated with each express themselves less with age. A larger study (Salami et al., 2012) also investigated the common and distinct spatial patterns of brain activity associated with the encoding and retrieval phases of a face-name paired associates task using PLS; 292 healthy adults ranging from 25 to 80 participated, and two distinct networks corresponding to encoding functions and more general cognitive control functions were extracted and described. All of these past studies demonstrate the recognition of a need to more clearly summarize which aspects of task-related activity are common, which are unique, and how these shared and unique networks change with age. To the best of our knowledge, our study is the first that will directly attempt to identify multiple RANNs in one study using several representative

tasks for each RA, and further, the first to demonstrate good CV for the topography of the task-related activation across these tasks.

We consider these results are preliminary and subject to change. While the initial GLM analyses were instructive, our preferred analytic strategy is multivariate. In addition the multivariate linear-indicator regression used in this study, we intend to perform multivariate analyses that use both tenets of good CV and robust brain-behavioral correlation as simultaneous constraints to help derive the RANNs as multivariate activation patterns. We will continue to work with the linear indicator regression analysis. Another promising strategy to employ for this objective is PLS (Lobaugh et al., 2001; McIntosh, 1999; McIntosh et al., 1996). For the 3 tasks within each RA domain, behavioral seeds can be used to compute three correlational images that can then be submitted to PC analysis. The resulting first PC would capture the major source of variance across these three correlational images. Since it will be dominated in large part by the mean correlational image, it can be used as a first approximation of the RANN for the RA domain under consideration. Pattern scores can be computed for each constituent task in any reference domain by forming the inner product between the RANN for that particular reference domain with any subject's activation map for the task in question. This procedure will generate subject pattern scores for 4 RANNs in 12 tasks that serve as candidates for the RANNs. CV of within-domain convergence and between-domain divergence can be investigated for the pattern scores in a similar manner to the procedure outlined in the present report. Eventually we would like to derive four patterns, one for each RA, that meet the criteria shown in Fig. 1. Ideally the patterns would show appropriate convergent and add discriminant validity. In addition, it would be ideal if subject pattern scores (i.e. expression of each pattern) was associated with performance of the related RA. While RANN derivation could be designed to force this relationship, our hope is that it will emerge naturally as we better delineate the individual RANNs.

When sufficient data is available, we hope to derive these multivariate RANNs in a data-pool of at least 100 younger participants (age <40 years), with a subsequent prospective application to data from each remaining decade, i.e., 40–50, ..., 70–80 in order to systematically evaluate how the composition and expression of these RANNs may change with aging. CV and brain-behavioral correlations can likewise be ascertained. This approach will allow us to make inferences about the nature of the neural substrates of age-related cognitive changes. For example, we hypothesize that the same covariance patterns (i.e., RANNs) exist in all age groups, while the level of expression of the RANNs (the pattern score) can vary with age group. This would parallel the observation that the RAs and their underlying neural networks remain invariant across aging while the level of performance declines over age for most of these abilities, and could reflect changes in the efficiency or capacity of the RANNs. Alternatively, we may find that the RANNs in younger adults do not forward apply to the older groups' imaging data, or that the forward applied RANN does not predict RA scores well. In that case we will repeat the process of deriving the RANNs from the older group's data. There are several possible outcomes to this process. We may find that the young RANNs are still present, but additional age-specific neural networks are necessary to predict RA scores. We have defined this outcome as *compensation* in previous published reports (Steffener et al., 2009; Stern, 2009). We may find that we are able to extract fewer than four RANNs (after testing for convergent and discriminant validity) across the four RAs. This is suggestive of *dedifferentiation*, or loss of functional specificity of neural networks. Finally, with aging we may find a unique RANN for each reference ability, indicative of functional reorganization without dedifferentiation.

We also intend to integrate data from other imaging modalities and investigate how the framework outlined in Fig. 1 might change, not only with age, but also with physiological substrates of age, obtained from imaging modalities such as gray-matter volume and cortical thickness, perfusion, white-matter integrity and resting-BOLD functional connectivity.



Another advantage to identifying RANNs using multivariate approaches is that these multivariate patterns could be readily forward applied to other functional imaging data sets. After assuring conventions regarding standardized sampling space and masking, a contrast of interest for a task would be calculated using standard GLM procedures, resulting in one parametric map per subject. The inner product between each RANN covariance pattern and the contrast map yields the subject's expression score for that RANN. These calculations can be informative in several ways: at the group level, investigation of the mean subject RANN expression can clarify the differential contribution of the 4 RANNs to task-specific networks. This would provide a standard metric for classifying any task in terms of the 4 RANNs. In addition, the investigator could explore whether subject expression of the RANN is correlated with indices of task performance. This would help determine the degree to which the 4 latent RAs contribute to task performance.

In summary, we have initiated an ambitious study that aims to identify four unique neural networks, each underlying a key RA that characterizes cognitive aging. Our hope is that by focusing on these RANNs, as opposed to task-related activation associated with individual tasks, we will be able to advance our knowledge regarding the key brain changes that underlie cognitive aging. More generally, the networks that we will develop will represent key aspects of cognitive function throughout the lifespan. The preliminary analyses presented here strongly suggest that we will be able to identify and fully characterize these networks.

### Acknowledgments

This research was supported by a grant from the National Institute on Aging (AG038465, PI Dr. Stern). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The sponsors had no role in the study design, data collection, analysis or interpretation, writing of the report, or decision to submit the article for publication.

### References

- Bak, T.H., O'Donovan, D.G., Xuereb, J.H., Boniface, S., Hodges, J.R., 2001. Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease-dementia-aphasia syndrome. *Brain* 124, 103–120.
- Blessed, G., Tomlinson, B.E., Roth, M., 1968. The association between quantitative measures of senile change in the cerebral grey matter of elderly subjects. *Br. J. Psychol.* 114, 797–811.
- Burianova, H., Grady, C.L., 2007. Common and unique neural activations in autobiographical, episodic, and semantic retrieval. *J. Cogn. Neurosci.* 19, 1520–1534.
- Burianova, H., McIntosh, A.R., Grady, C.L., 2010. A common functional brain network for autobiographical, episodic, and semantic memory retrieval. *NeuroImage* 49, 865–874.
- Burnham, K.P., Burnham, K.P., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Springer, New York.
- Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Ekstrom, R.B., French, J.W., Harman, H.H., Dermen, D., 1976. *Manual for Kit of Factor-referenced Cognitive Tests*. Princeton.
- Grady, C.L., Springer, M.V., Hongwanishkul, D., McIntosh, A.R., Winocur, G., 2006. Age-related changes in brain activity across the adult lifespan. *J. Cogn. Neurosci.* 18, 227–241.
- Hampshire, A., Highfield, R.R., Parkin, B.L., Owen, A.M., 2012. Fractionating human intelligence. *Neuron* 76, 1225–1237.
- Hastie, T., Tibshirani, R., Friedman, J.H., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York.
- Leech, R., Sharp, D.J., 2014. The role of the posterior cingulate cortex in cognition and disease. *Brain* 137, 12–32.
- Lobaugh, N.J., West, R., McIntosh, A.R., 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology* 38, 517–530.
- Mattis, S., 1988. *Dementia Rating Scale (DRS)*. Psychological Assessment Resources, Odessa, FL.
- McIntosh, A.R., 1999. Mapping cognition to the brain through neural interactions. *Memory* 7, 523–548.
- McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage* 3, 143–157.
- Nyberg, L., Forkstam, C., Petersson, K.M., Cabeza, R., Ingvar, M., 2002. Brain imaging of human memory systems: between-systems similarities and within-system differences. *Brain Res. Cogn. Brain Res.* 13, 281–292.
- Oh, A., Duerden, E.G., Pang, E.W., 2014. The role of the insula in speech and language processing. *Brain Lang.* 135C, 96–103.
- Raven, J.C., 1962. *Advanced Progressive Matrices, Set II*. H.K. Lewis, London, UK.
- Saalmann, Y.B., Kastner, S., 2011. Cognitive and perceptual functions of the visual thalamus. *Neuron* 71, 209–223.
- Salami, A., Eriksson, J., Nyberg, L., 2012. Opposing effects of aging on large-scale brain systems for memory encoding and cognitive control. *J. Neurosci.* 32, 10749–10757.
- Salthouse, T.A., 1998. Independence of age-related influences on cognitive abilities across the life span. *Dev. Psychol.* 34, 851–864.
- Salthouse, T.A., 2005. Relations between cognitive abilities and measures of executive functioning. *Neuropsychology* 19, 532–545.
- Salthouse, T.A., 2009. Decomposing age correlations on neuropsychological and cognitive variables. *J. Int. Neuropsychol. Soc.* 15, 650–661.
- Salthouse, T.A., Babcock, R.L., 1991. Decomposing adult age differences in working memory. *Dev. Psychol.* 27, 763–776.
- Salthouse, T.A., Davis, H.P., 2006. Organization of cognitive abilities and neuropsychological variables across the lifespan. *Dev. Rev.* 26, 31–54.
- Salthouse, T.A., 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., Kersten, A.W., 1993. Decomposing adult age differences in symbol arithmetic. *Mem. Cogn.* 21, 699–710.
- Salthouse, T.A., Pink, J.E., Tucker-Drob, E.M., 2008. Contextual analysis of fluid intelligence. *Intelligence* 36, 464–486.
- Schmahmann, J.D., 1996. From movement to thought: anatomic substrates of the cerebellar contribution to cognitive processing. *Hum. Brain Mapp.* 4, 174–198.
- Shenhav, A., Botvinick, M.M., Cohen, J.D., 2013. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 79, 217–240.
- Sheth, S.A., Mian, M.K., Patel, S.R., Asaad, W.F., Williams, Z.M., Dougherty, D.D., Bush, G., Eskandar, E.N., 2012. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature* 488, 218–221.
- Steffener, J., Brickman, A.M., Rakitin, B.C., Gazes, Y., Stern, Y., 2009. The impact of structure on age-related changes in working memory functional activity. *Brain Imaging Behav.* 3, 142–153.
- Stern, Y., 2009. Cognitive reserve. *Neuropsychologia* 47, 2015–2028.
- St-Laurent, M., Abdi, H., Burianova, H., Grady, C.L., 2011. Influence of aging on the neural correlates of autobiographical, episodic, and semantic memory retrieval. *J. Cogn. Neurosci.* 23, 4150–4163.
- Stoodley, C.J., Valera, E.M., Schmahmann, J.D., 2012. Functional topography of the cerebellum for motor and cognitive tasks: an fMRI study. *NeuroImage* 59, 1560–1570.
- Strick, P.L., Dum, R.P., Fiez, J.A., 2009. Cerebellum and nonmotor function. *Annu. Rev. Neurosci.* 32, 413–434.
- Woodcock, R.W., Johnson, M.B., Mather, N., 1989. *Woodcock-Johnson Psycho-Educational Battery—Revised*. DLM Teaching Resources.
- Wurtz, R.H., Kandel, E.R., 2000. The central visual pathways. In: Kandel, E.R., Schwartz, J.H. (Eds.), *Principles of Neural Science*, 4th ed. McGraw-Hill, New York, NY, pp. 523–547.