Assessing and mapping language, attention and executive multidimensional deficits in stroke aphasia

Rahel Schumacher[1, 2]*, Ajay D. Halai[2]*, & Matthew A. Lambon Ralph[2]*

[1] Neuroscience and Aphasia Research Unit, School of Biological Sciences, University of Manchester, Manchester, United Kingdom

[2] MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom

* Corresponding authors

Postal address:

MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, CB2 7EF, England

Email addresses:

rahel.schumacher@manchester.ac.uk,

ajay.halai@mrc-cbu.cam.ac.uk,

matt.lambon-ralph@mrc-cbu.cam.ac.uk

Running title: Nonverbal cognition in stroke aphasia

Abstract

There is growing awareness that aphasia following a stroke can include deficits in other cognitive functions and that these are predictive of certain aspects of language function, recovery and rehabilitation. However, data on attentional and executive (dys)functions in individuals with stroke aphasia are still scarce and the relationship to underlying lesions is rarely explored. Accordingly in this investigation, an extensive selection of standardized nonverbal neuropsychological tests was administered to 38 individuals with chronic post-stroke aphasia, in addition to detailed language testing and magnetic resonance imaging. To establish the core components underlying the variable patients' performance, behavioural data were explored with rotated principal component analyses, first separately for the nonverbal and language tests, then in a combined analysis including all tests. Three orthogonal components for the nonverbal tests were extracted, which were interpreted as shift-update, inhibit-generate and speed. Three components were also extracted for the language tests, representing phonology, semantics and speech quanta. Individual continuous scores on each component were then included in a voxel-based correlational methodology analysis, yielding significant clusters for all components. The shift-update component was associated with a posterior left temporo-occipital and bilateral medial parietal cluster, the inhibit-generate component was mainly associated with left frontal and bilateral medial frontal regions, and the speed component with several small right-sided fronto-parieto-occipital clusters. Two complementary multivariate brain-behaviour mapping methods were also used, which showed converging results. Together the results suggest that a range of brain regions are involved in attention and executive functioning, and that these non-language domains play a role in the abilities of patients with chronic aphasia. In conclusion, our findings confirm and extend our understanding of the multidimensionality of stroke aphasia, emphasize the importance of assessing nonverbal cognition in this patient group and provide directions for future research and clinical practice. We also briefly compare and discuss univariate and multivariate methods for brain-behaviour mapping.

Keywords

aphasia, executive functions, attention, principal components, univariate and multivariate brain-behaviour mapping

List of Abbreviations

FCP = fuzzy clustering fixed prototypes

FWEc = family-wise error corrected

GPR = gaussian process regression

KRR = kernel ridge regression

MKR = multi-kernel regression

MSE = mean square error

PCA = Principal Component Analysis

PRoNTo = pattern recognition of neuroimaging toolbox

RVR = relevance vector regression

SVR-LSM = support-vector regression lesion symptom mapping

TAP = Test of Attentional Performance

VBCM = voxel-based correlational methodology

VLSM = voxel-based lesion-symptom mapping

Introduction

There is a growing understanding that a left hemispheric stroke leading to impairments in language processing – aphasia – often also affects other cognitive functions, such as attention or executive functions (Glosser and Goodglass, 1990; Helm-Estabrooks, 2002; Jefferies and Lambon Ralph, 2006; Murray, 2012; Villard and Kiran, 2017) and it has been shown that impairments in these cognitive functions play an important role in aphasia recovery and rehabilitation (Fillingham *et al.*, 2005; van de Sandt-Koenderman *et al.*, 2008; Lambon Ralph *et al.*, 2010; Brownsett *et al.*, 2014; El Hachioui *et al.*, 2014; Geranmayeh *et al.*, 2017; Simic *et al.*, 2017). The occurrence and patterns of nonverbal cognitive dysfunctions in patients with aphasia, the relationship between nonverbal and language impairments, and their structural correlates have been examined separately in some studies. To date, however, no investigation has undertaken a detailed behavioural assessment of both verbal and nonverbal performance or combined this with structural imaging data.

A handful of previous behavioural studies have examined nonverbal cognition in patients with aphasia but did so either with a narrow focus, for instance investigating the impact of domain-general executive dysfunctions on semantic cognition (Thompson *et al.*, 2018), or on a rather general level with findings based on composite scores (Helm-Estabrooks, 2002), a few standardized tests per domain (Kauhanen *et al.*, 2000; Fucetola *et al.*, 2009; El Hachioui *et al.*, 2014; Lee and Pyun, 2014; Marinelli *et al.*, 2017; Wall *et al.*, 2017) or experimental tasks (Villard and Kiran, 2015; Kuzmina and Weekes, 2017). This limited test selection stands in contrast to research efforts with healthy participants or other patient populations, which have explored the nature of multiple components within attention and executive function (Mirsky *et al.*, 1991; Miyake *et al.*, 2000; Friedman and Miyake, 2017). One study including patients with aphasia used a broad range of attention assessments and indeed found that aspects of attention differed with respect to their predictive power regarding language function (Murray, 2012).

Another limitation of existing studies is that patient performance is often reported on a group level only (Glosser and Goodglass, 1990; Kauhanen *et al.*, 2000; El Hachioui *et al.*, 2014; Lee and Pyun, 2014; Naranjo *et al.*, 2018) and information about the prevalence of impaired performance based on normative data is seldom available or incomplete. This information is, however, of clinical significance and relevant when performance in different aspects of cognitive functioning is to be compared.

Underlying patterns in impaired and preserved abilities of heterogeneous patient populations can be extracted using data reduction techniques, such as principal component analysis (Kummerer *et al.*, 2013; Butler *et al.*, 2014; Mirman *et al.*, 2015; Halai *et al.*, 2017; Lacey *et al.*, 2017). Applied to large, detailed datasets containing language measures and a handful of executive function assessments, a previous study of chronic post-stroke aphasia found three principal components (phonology, semantics, executive function) underlying participants' performance (Butler *et al.*, 2014), which was supplemented by a fourth speech quanta component (the quantity of speech produced in connected-speech tasks) in a subsequent study (Halai *et al.*, 2017). One major advantage of data-driven approaches is that they can accommodate for the fact that multiple processes underlie performance in any given test (e.g., naming requires preserved visual perception, semantics, phonology and motor articulation) and no test is a pure measure of single cognitive/language processes. Indeed, sensibility regarding the linguistic demands of any test is particularly high within the field of aphasia. These concerns are usually expressed in the sense that impaired language functions may interfere with testing of other cognitive domains (Keil and Kaszniak, 2002), and more rarely the other way around (Heuer *et al.*, 2017). Data-driven approaches offer a formal method to establish the mutual influences of language and nonverbal ability on test performance.

Based on studies with healthy controls and various neurological populations, a bilateral fronto-cingulo-parietal network is known to be involved in attention and executive function processes (Miller and Cohen, 2001; Duncan, 2010; Niendam *et al.*, 2012; Petersen and Posner, 2012; Fedorenko *et al.*, 2013; Power and Petersen, 2013) but little is known about the structural correlates of attentional and executive dysfunctions in patients with aphasia. Recent research combining data-driven decomposition of behavioural assessment with neuroimaging data, has revealed the structural correlates of behavioural performance in patients with aphasia (Kummerer *et al.*, 2013; Butler *et al.*, 2014; Mirman *et al.*, 2015; Halai *et al.*, 2017; Lacey *et al.*, 2017). Whilst extracting clear brain-behaviour relationships for various aspects of language, these studies struggled to find significant associations of tissue integrity with scores on executive function (though see Lacey *et al.*, 2017), either because non-language assessment was not included (e.g. Kummerer *et al.*, 2013; Mirman *et al.*, 2015) or assessment coverage was too limited (Butler *et al.*, 2014; Halai *et al.*, 2017).

In addition to the form and analysis of patients' behavioural assessment, the approach to mapping brain-behaviour relationships could also be critical. Univariate approaches, such as voxel-based lesion-symptom mapping (VLSM) (Bates *et al.*, 2003) and voxel based

correlational methodology (VBCM) (Tyler *et al.*, 2005), are relatively easy to run and interpret. Recent debate has noted the potential shortcomings of univariate approaches (Karnath *et al.*, 2018) including the inability to detect conditional voxel combinations (DeMarco and Turkeltaub, 2018) and mis-localisation (Mah *et al.*, 2014), which might be addressed by multivariate analyses (though see Sperber *et al.*, 2019). The power of multivariate analyses, however, bring new interpretation challenges which are straightforward in univariate approaches: because all weights in multivariate models are conditional on each other, the interpretation or post hoc thresholding of individual weights becomes non-trivial (see Haufe *et al.*, 2014). Accordingly, making inferences about local brain-behaviour relationships based on multivariate models is, at best, complicated. One transparent way forward is for studies to begin to present both univariate and multivariate results. Therefore, in the current study we show the results for four different methodological approaches, which allows us to demonstrate some commonalities and differences.

To extend our understanding of stroke aphasia to potentially critical aspects of non-verbal cognitive function and their structural correlates, we administered a comprehensive battery of non-verbal tests of attention and executive function to a large and diverse group of individuals with chronic post-stroke aphasia. The key aims of the study were: (1) to assess the prevalence of attention and executive dysfunction in patients with post-stroke aphasia; (2) to explore the underlying relationships between the tests of attention and executive function, as well as the link to the patients' language profiles; and (3) to map the structural correlates for these underlying attention, executive and language features by means of four different methodological approaches.

Materials and methods

Participants

Thirty-eight participants were recruited for the present study (11 female, 27 male; mean age 64 ± 11.9 years, range 45-88 years; see Supplementary Table 1 for more details). All participants had a single left hemispheric stroke (ischaemic or haemorrhagic) at least one year before assessment and imaging (see Fig. 1 for lesion overlap map) and had no additional significant neurological conditions and no contraindications for MRI. They were pre-morbidly right-handed native English speakers with normal or corrected-to-normal vision. All had been diagnosed with aphasia but no restrictions were applied regarding the type of aphasia or the severity. Five patients are identical to patients whose data were reported in Halai *et al.* (2017) and Butler *et al.* (2014). Informed consent was obtained from all participants prior to participation, in line with the Declaration of Helsinki and as approved by the local NHS ethics committee. MRI data from a healthy age and education matched control group (10 female, 12 male) was used as a reference to identify lesion/abnormal tissue for each patient (Seghier *et al.*, 2008).

---------- Fig. 1 about here ----------

Neuropsychological assessments

In addition to comprehensive language testing, described in more detail in Butler *et al.* (2014) and Halai *et al.* (2017), a broad range of standardized neuropsychological tests of attention and executive functions were administered. This included the subtests Alertness, GoNoGo, Divided Attention, and Distractibility from the Test of Attentional Performance (TAP - Mobility version 1.3.1; Zimmermann and Fimm, 1995; www.psytest.net), a computerized test battery measuring reaction times and error rates in tests with varying attentional demands; the subtests Design Fluency and Trail Making (parts 2-4) from the Delis-Kaplan Executive Function System (D-KEFS; Delis *et al.*, 2001), the former assessing nonverbal idea generation by requiring participants to draw as many different figures as possible (connecting dots with lines), and the latter assessing visuospatial attention, processing speed and flexibility by requiring participants to connect numbers (part 2), letters (part 3) or alternatingly both (part 4) in ascending order; a computerized version of the Tower of London (TOL-F by Schuhfried; Kaller *et al.*, 2011), a visuospatial planning task; the Kramer test (Balzer *et al.*, 2011), a categorization task requiring

participants to find ways of sorting eight cards into two groups; the Raven's Coloured Progressive Matrices (Raven, 1962), assessing reasoning abilities; and the Brixton test (Burgess and Shallice, 1997), assessing visuospatial rule detection. Test scores were compared to published norms; age- and/or education-corrected norms were considered if available. For the Raven Matrices, the norms for part B were taken from Smits *et al.* (1997). Following Brooks *et al.* (2011), performance was considered as at least mildly-to-moderately impaired if it was more than 1.5 standard deviations below the mean (i.e., a T-score below 35, a percentile rank below 6 or a scaled score of 5 or lower).

Data analysis

For a descriptive comparison of the impairments per patient and measure, and to account for missing data, percentages of impaired scores were calculated based on 16 measures from the 10 nonverbal tests and 14 measures from 12 language tests. The percentage of impaired scores per patient was taken as an indicator of the severity of their impairment and subsequently used in correlation analyses. Based on the raw test scores, three principal component analyses (correlation-based) were performed (using IBM SPSS 22.0) to elucidate the data's underlying structure. The first PCA comprised just the nonverbal tests of attention and executive function. In the second PCA, only the language measures were included, which also provided a replication of previous results (Butler *et al.*, 2014; Halai *et al.*, 2017). Lastly, the third PCA comprised the combination of all measures included in the two other PCAs. To facilitate interpretation, it was ensured that a higher score would indicate better performance for all measures. To this end, reaction time measures were inverted, and accuracy rates were computed. Due to missing values and to include the same sample in all analyses, data of 32/38 patients were entered in the PCAs. TAP Distractibility and the letter and switching versions of the Trail Making test were not included in order to not further decrease the sample size. Importantly, analyses including these measures showed that they were highly correlated with measures of the GoNoGo test or the number version of the Trail Making test, respectively. To reduce the number of variables entered in the analysis, some comparable language measures were combined (Boston naming and Cambridge naming, immediate and delayed repetition of words and non-words, spoken and written word-picture matching, word and non-word minimal pairs). All components with eigenvalues ≥ 1 were extracted and then varimax rotated, yielding orthogonal and interpretable components. Two control analyses were performed to assess the stability and predictability of the PCA results. First, means and 95% confidence intervals for the component loadings were computed by leaving one case out each time. Second, the

similarity between the observed data and those predicted was determined using a leave one case out method (by projecting the left-out case into the component space using the coefficient matrix). Correlations were computed to explore the relationship between component scores and the severity of the impairment in the neuropsychological tests as well as with patient characteristics such as lesion volume, age, and years of education.

Neuroimaging data acquisition and analysis

High resolution structural T1-weighted Magnetic Resonance Imaging (MRI) scans were acquired on a 3.0 Tesla Philips Achieva scanner (Philips Healthcare, Best, The Netherlands) using an 8-element SENSE head coil. A T1-weighted inversion recovery sequence with 3D acquisition was employed, with the following parameters: TR (repetition time) = 9.0 ms, TE (echo time) = 3.93 ms, flip angle = 8°, 150 contiguous slices, slice thickness = 1 mm, acquired voxel size 1.0 × 1.0× 1.0 mm, matrix size 256 × 256, field of view = 256 mm, TI (inversion time) = 1150 ms, SENSE acceleration factor 2.5, total scan acquisition time = 575 s.

Structural MRI scans were pre-processed with Statistical Parametric Mapping software (SPM8: Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/). The images were normalised into standard Montreal Neurological Institute (MNI) space using a modified unified segmentation-normalisation procedure optimised for focal lesioned brains (Seghier *et al.*, 2008). Data from all participants with stroke aphasia and all healthy controls were entered into the segmentation- normalisation. Images were then smoothed with an 8 mm full width-half-maximum (FWHM) Gaussian kernel and used in the lesion analyses described below. An age and education matched healthy control group was used to determine the extent of abnormality per voxel. This was achieved using a fuzzy clustering fixed prototypes (FCP) approach, which measures the similarity between a voxel in the patient data with the mean of the same voxel in the control data (note: this method does not discriminate what caused the abnormality, but simply reflects how deviant the signal in the patient scan is from a healthy group). One can apply a threshold to the FCP to determine membership to abnormal/normal voxel. The default parameters were used apart from the lesion definition 'U-threshold', which was set to 0.5 to create a binary lesion image. We modified the U-threshold from 0.3 to 0.5 after comparing the results obtained from a sample of patients to what would be nominated as lesioned tissue by an expert neurologist. The images generated for each patient were visually inspected and manually corrected if necessary and were then used to create the lesion overlap map in Fig. 1.

The smoothed FCP images (% abnormality) were used to determine the brain regions where abnormality correlated with PCA component scores using a voxel-based correlational methodology (VBCM) (Tyler *et al.*, 2005), a variant of voxel-lesion symptom mapping (Bates *et al.*, 2003), in which both the behaviour and signal intensity measures are treated as continuous variables (conducted in SPM12). For the structural correlate analysis, we assume a negative correlation between abnormality and behavioural component score (i.e. greater abnormality leads to poorer performance). The participants' component scores from the combined PCA, were entered simultaneously into a VBCM analysis. The resulting clusters thus account for the unique variance of a component. In additional analyses, lesion volume (calculated from the lesion identified by the automated lesion identification method (Seghier *et al.*, 2008)), age, education, and time post stroke were entered as covariates. Unless noted otherwise, we applied the threshold at voxel-level $p < 0.001$ and family-wise error corrected (FWEc) cluster-level $p < 0.05$.

To supplement the univariate analysis, we conducted multivariate analyses in two ways. First, we used the support-vector regression lesion symptom mapping (SVR-LSM) toolbox recently updated by DeMarco and Turkeltaub (2018), which was based on Zhang *et al.* (2014). In this framework, we loaded the lesion binary images as the features and created a separate model for each component score. The following settings were used: MATLAB SVM implementation, hyper-parameter optimisation (Bayes Optimisation with default settings) and lesion threshold = 3 (~10% of sample). The resulting beta weights were evaluated by permutation testing (N=10,000, voxel-wise $p < 0.005$ and cluster-wise $p < 0.05$), but note that the model performance (predicted vs. observed scores) is not evaluated in this approach. We ran two models per component, with and without correction for lesion volume ('regress on both'). Second, we used the pattern recognition of neuroimaging toolbox (PRoNTo V2.1) (http://www.mlnl.cs.ucl.ac.uk/pronto/) (Schrouff *et al.*, 2013) as an alternative method because (1) it formally evaluates model predictions and (2) it does not truncate beta weights post-hoc. For this toolkit, we entered the FCP % abnormality images as a continuous measure and followed the pipeline through in two pathways: 1) using the whole brain as input (similar to the VBCM) and 2) restricted to lesion territory (N>3) (similar to VLSM/SVR-LSM). Given the simplicity of the toolkit, we ran models using four regression machine implementations: (1) kernel ridge regression (KRR; Hastie *et al.*, 2009), (2) relevance vector regression (RVR; Tipping, 2001), (3) gaussian processes regression (GPR; Rasmussen and Williams, 2006) and (4) multi-kernel regression (MKR; Bach *et al.*, 2004; Rakotomamonjy *et al.*, 2008). PRoNTo

relies on kernel methods in order to overcome the high dimensionality problem in neuroimaging (using NxN pair-wise similarity matrix) and features were mean centred. The default parameters were used for all machines and where necessary hyper-parameter optimisation was achieved using nested leave-one-out cross validation (default grid search). A leave-one-out cross-validation scheme was used to determine model performance. For model inference, we report p-values for correlation and mean square error (MSE) following a permutation test of the observed scores (N = 1000) with a $p < 0.05$ alpha threshold. As with the SVR-LSM, we ran each component model with and without lesion volume as a covariate.

The anatomical labels for the clusters were determined using the Harvard-Oxford atlas for grey matter and on the John Hopkins white matter atlas for white matter tracts. Furthermore, comparisons to existing findings were made by either overlapping the respective maps, if available, or by checking (in MRICron) whether published peak coordinates overlapped with the clusters from the VBCM.

Data availability

Behavioural data are available in the Supplementary material. Further data are potentially available by request to the last author.

Results

Neuropsychological profiles

The first aim of this study was to assess the prevalence of impairments in attention and executive functions in patients with post-stroke aphasia. Patients' performance was thus compared to available norm data in order to identify the number of impaired scores per patient and test. All participants scored below normal range in at least one measure of the ten tests of attention and executive function, but no participant was impaired in all of these tests (mean percentage of impaired scores per patient 36.7 ± 20.8 %, range 6.3-90.9 %). Fifteen patients were impaired in at least half of the administered nonverbal tests. In comparison to the nonverbal test performance, all participants scored below normal range in at least three measures of the twelve language tests, 30 patients were impaired in at least half of the administered language tests, and five participants were impaired in all of these tests (mean percentage of impaired scores per patient 65.0 ± 22.4%, range 21.4-100%). Details on impaired performance in the nonverbal and language tests are depicted in Fig. 2, while Fig. 3 shows the patients' overall impairment in the nonverbal versus language tests (as percentage of impaired scores in the respective tests). Individual patients' scores are available in Supplementary Table 2 and Supplementary Table 3 while Supplementary Fig. 2 gives details about impaired performance on the different principal components.

The Alertness test and the Distractibility without distractor condition were the only two nonverbal tests where the percentage of impaired scores was around or below 5% of the sample. These tests measure more basic attention functions and it has previously been reported that these aspects of attention are more commonly impaired in right-hemispheric stroke patients (Sturm *et al.*, 1997). The tests with the highest percentages of impaired scores were the Trail Making Test (numbers impaired in 25 patients (65.8%), letters in 32 patients (88.9%), and switching in 29 patients (85.3%)), the Design Fluency test (25 patients, 67.6%) and the Kramer test (21 patients, 58.3%). We split the sample into two groups of 'cognitive' severity based on a median split of overall impairment in the nonverbal tests (see Supplementary Table 2 for details). Comparison of the two groups revealed that only the more cognitively-severe patients had impaired scores in the Tower of London and TAP Divided Attention tests. As such, the test of divided attention might be especially clinically useful as a predictor of impaired cognition in aphasic populations. In contrast, both groups showed a similar and high degree of impairment in two other tests, the Kramer and the letter version of the Trail Making test. The

high percentage of impaired performance in the Trail Making Test is particularly important considering the widespread use of this test with aphasic patients. Thus, impaired performance in the switching condition of the trail making test need not necessarily stem from difficulties in switching but from reduced automaticity of accessing the letters in order (and, to a lesser extent, numbers), which is a prerequisite for task completion.

---------- Figures 2 & 3 about here ----------

Separate and combined principal component analyses of nonverbal and language tests

The second aim was to explore the underlying relationships between the tests of attention and executive function, as well as linking these to the patients' language profiles. We computed separate PCAs for the nonverbal and verbal tests, as well as a combined PCA including all tests. The PCA including only the nonverbal tests of attention and executive functions yielded three orthogonal components accounting for 68.5 % of the variance (KMO = 0.704). Based on the tests loading highest on each component (see Fig. 4A), the first component (accounting for 28.1 % of the variance) was interpreted as 'shift-update' as the tests loading highest are relatively demanding with respect to flexible (visuo-spatial) processing and working memory. Interestingly, the first component contains tests that are traditionally regarded as tests of executive function (Tower of London, Brixton) as well as tests that are more associated with attention (Divided attention and Trails numbers), which underlines the link between the two domains that is also reflected in the term 'executive attention' (Kane and Engle, 2002; Petersen and Posner, 2012). The second component (23.2 %) was interpreted as 'inhibit-generate' as it included tests like the Kramer sorting test (requiring idea generation as well as inhibition of salient aspects of the stimuli) as well as simple response inhibition tasks like the GoNoGo test. The third component (17.2 %) was interpreted as 'speed' as it contained the reaction time measures of both basic attention tasks.

The separate analysis of the language tests yielded three orthogonal components accounting for 78.3 % of the variance (KMO = 0.718). The components can be interpreted as 'phonology' (accounting for 31.5 % of the variance), 'semantics' (24.2 %), and 'speech quanta' (22.6 %), directly replicating previous research (Butler *et al.*, 2014; Halai *et al.*, 2017). The fact that the patient sample of this study largely consists of patients not included

in previous reports shows the stability of these results. Moreover, other groups report similar patterns (Mirman *et al.*, 2015; Lacey *et al.*, 2017).

The third PCA - combining the nonverbal and language tests - yielded six orthogonal components accounting for 78.6 % of the variance (KMO = 0.661). Fig. 4A shows that the components from the two separate analyses remained relatively stable (also evidenced by high correlations between the separate and combined component scores, see Table 1 and Supplementary Fig. 2). Their order and percentage of explained variance was as follows: phonology (21.6 %), shift-update (13.4 %), inhibit-generate (12.2 %), speech quanta (11.7 %), semantics (11.5 %), speed (8.2 %). Notably, apart from the phonology component which explained the highest amount of variance, the other language and nonverbal components are weighted similarly in terms of explained variance.

The stability analyses for all three PCAs revealed that all test loadings had very tight 95% confidence intervals. The most unstable tests were Design Fluency in the nonverbal PCA (mean loading = 0.58 ± 0.02), Camel and Cactus in the verbal PCA (0.86 ± 0.08), and Kramer in the combined PCA (0.75 ± 0.05). We also found generally high correlations between the predicted left-out cases and observed scores for the nonverbal (r = 0.83), verbal (r = 0.88) and combined (r = 0.88) PCAs.

Whilst the combined PCA preserves the nature of the six principal behavioural components, it is notable that many individual language tasks load across verbal and nonverbal components – reflecting the fact that many language activities and the tasks used to assess them require generalised attention and executive skills (e.g., comparing verbal stimuli, deciding between responses, etc.). This is true for both semantic tests (aligning with the fact that semantic cognition requires both access to semantic representation but also executively-related processes (see Jefferies and Lambon Ralph, 2006; Thompson *et al.*, 2018) and for phonological tests with demands on working memory (sentence comprehension) or abstract reasoning and problem-solving (minimal pairs).

----------Fig. 4 about here ----------

Relationship between impairment, component scores, and patient characteristics

Previous research documents both the presence (Fucetola *et al.*, 2009; Baldo *et al.*, 2015) and absence (Helm-Estabrooks, 2002) of a significant correlation between nonverbal and verbal impairment. We found a moderate but significant relationship between simple indices of nonverbal and language impairment (in terms of percentage of impaired nonverbal/language test scores per patient), as shown in Fig. 3 and Table 1. This finding seems to relate primarily to the nonverbal shift-update component that correlates with both indices of severity. Beyond this, there is considerable variation, which results from the fact that even when combined into one PCA there are statistically-orthogonal components for the language and nonverbal test scores; they would collapse into a shared PCA component if performance in nonverbal and language tests was a reflection of simple severity alone.

Regarding patient characteristics, also shown in Table 1, nonverbal as well as verbal severity correlated significantly with lesion volume, but neither correlated with age, education or time post stroke. More specifically, lesion volume correlated with the separate nonverbal shift-update component and with the semantic and speech quanta components of both PCAs. Age correlated with the nonverbal components apart from speed, and with the semantic component from the separate verbal PCA. Education only correlated significantly with the inhibit-generate component from the separate nonverbal PCA, and time post stroke correlated moderately with the shift-update components.

Notably, the first nonverbal and language components, shift-update and phonology, were still significantly correlated with the severity of the nonverbal and language impairment, respectively, when age, education, time post stroke and lesion volume were accounted for by means of partial correlation (separate shift-update component and nonverbal impairment $r = -0.629$; separate / combined phonology component and language impairment $r = -0.814$ / $r = -0.851$; all $p < 0.0004$).

----------Table 1 about here ----------

Structural correlates

The third aim was to map the structural correlates for the underlying attention, executive and language features. We simultaneously entered all component scores obtained in the combined PCA and performed a VBCM with tissue abnormality, which yielded significant clusters for all components (though shift-update and speech quanta were present at a lower voxel-level threshold of 0.01, FWEc at cluster-level $p < 0.05$). The clusters are depicted in Fig. 4B and Fig. 5, and details are listed in Table 2.

From the nonverbal components, shift-update was uniquely correlated with left lateral temporo-occipital regions (encompassing parts of the medial and inferior temporal gyrus, fusiform cortex as well as the lateral occipital cortex and extending to parahippocampal regions and brain stem), in addition to bilateral mainly parietal midline regions (postcentral gyrus, precuneous, superior parietal lobule). The inhibit-generate component was uniquely correlated with left lateral (middle and inferior frontal gyrus) and subcortical frontal regions (anterior thalamic radiation) as well as medial frontal regions bilaterally (subcallosal cortex, (para)cingulate gyrus, supplementary motor cortex), in addition to several smaller clusters in occipital and parietal regions. The speed component was also associated with several small, mainly right-sided parieto-occipital and frontal clusters.

The clusters associated with the three language components resembled the clusters reported in previous studies by our group (Butler *et al.*, 2014; Halai *et al.*, 2017). The phonology cluster was uniquely correlated with left temporo-parietal regions encompassing parts of the inferior, middle, and superior temporal gyri as well as supramarginal and angular gyrus. The semantics component was associated with a cluster of left cortical (anterior temporal lobe, extending inferiorly into occipital lobe) and subcortical (thalamus) regions. The speech quanta cluster was in the dorsal fronto-parietal cortex and included parts of the pre- and postcentral gyrus. When lesion volume was included as a covariate, inhibit-generate, speed, and phonology remained significant. Semantics was only significant at a less strict threshold; this applied as well to the shift-update component and is shown in Supplementary Fig. 3. The effects of including other patient characteristics such as age, education, and time post stroke in the VBCM are also shown and discussed in the Supplementary material.

---------- Table 2 about here----------

The multivariate analyses yielded similar results, as shown in Fig. 5. The SVR-LSM produced significant clusters for inhibit-generate, phonology, semantics and speech quanta. The evaluation of the best model within PRoNTo revealed significant brain-behaviour relationships for inhibit-generate (KRR model cross-validation r = 0.357, MSE = 0.854, p = 0.022), phonology (MKR model cross-validation r = 0.379, MSE = 1.008, p = 0.042), and semantics (KRR model cross-validation r = 0.750, MSE = 0.431, p < 0. 001) when using the whole brain. The results were the same when using the restricted lesion territory: inhibit-generate (KRR model cross-validation r = 0.400, MSE = 0.816, p = 0.019), phonology (GPR model cross-validation r = 0.359, MSE = 0.860, p = 0.013), and semantics (KRR model cross-validation r = 0.712, MSE = 0.478, p < 0.001). When lesion volume was added as a covariate, the SVR-LSM produced significant clusters for inhibit-generate and phonology only, while the PRoNTo toolkit found significant models for inhibit-generate and semantics (for both whole brain and restricted lesion territory), as detailed in Supplementary materials.

As can be seen in Fig. 5, the VBCM and SVR-LSM results were strikingly similar. For inhibit-generate, VBCM yielded bigger and more distributed clusters but there was an overlap with the significant SVR-LSM result in left frontal subcortical regions. For phonology, the SVR-LSM and VBCM clusters were nearly identical, with the former extending slightly more into the superior parietal cortex, and the latter extending more anteriorly in the temporal lobe. Likewise, the VBCM and SVR-LSM results for the semantics component overlapped largely, with the former being slightly bigger and extending further posteriorly in the ventral temporal lobe. Finally, the main difference regarding the speech quanta results was that the SVR-LSM cluster extended slightly more dorsally and anteriorly. Furthermore, the unthresholded beta maps from PRoNTo showed some correspondence to both VBCM and SVR-LSM in terms of the negative beta weights. Apart from a small set of voxels in the medial temporal lobe that was part of the VBCM semantics cluster, all voxels identified in the VBCM and SVR-LSM analyses were within regions that were given a (strong) negative weight in the PRoNTo models. In contrast to SVR-LSM, the PRoNTo beta maps show the weights of the entire input space after confirming the model significantly maps to behaviour.

---------- Fig. 5 about here----------

Discussion

Even though there is growing awareness of the importance of attentional and executive (dys)functions in aphasia, to date the occurrence and patterns of such impairments, the relationship between nonverbal and language functions, as well as their structural correlates have not been studied in detail in the same sample of patients. This study extended our understanding of the multidimensionality of chronic post-stroke aphasia and found that: (1) a considerable number of patients showed impaired performance in tests of attention and executive function; (2) the variance underlying nonverbal and language test performance was best captured by three orthogonal components each; (3) both univariate and multivariate mapping approaches revealed brain-behaviour relationships in line with previous studies based on other methodologies and populations.

Given that our sample consisted of patients diagnosed with aphasia, unsurprisingly the incidence of language impairments was high and performance in language tests was overall worse than in nonverbal tests. However, patients' performance in tests of attention and executive function was also considerably impaired, as none of the patients performed within normal range in all tests and nearly 50% of the patients showed deficits in at least half of the administered tests. While language impairments might be the most salient consequences of a left hemispheric stroke, our more thorough and systematic investigation replicates earlier observations of co-occurring deficits in other cognitive domains (e.g. Helm-Estabrooks, 2002; Murray, 2012; Marinelli *et al.*, 2017; Ramsey *et al.*, 2017); a pattern which is important for clinical management and response to rehabilitation.

Our comprehensive battery of nonverbal tests allowed us to identify three separable components of attention and executive function (shift-update, inhibit-generate, and speed) which mirror explorations in healthy participants (e.g. Petersen and Posner, 2012; Friedman and Miyake, 2017). This contrasts with current studies in aphasia and clinical practice that either fail to assess nonverbal functions at all, or if they do then only a few (screening) measures are used. Whilst there are clear co-occurrences and simple raw correlations between measures, there is little evidence that everything collapses to one simple severity-based metric. This is in line with a recent study by Marinelli et al. (2017), reporting that only a quarter of their severely aphasic patients was also severely impaired in nonverbal cognition, as well as classical findings showing that language and non-language performance in aphasia have low correlations, and

that aphasia cannot be reduced to simple cognitive severity (Basso *et al.*, 1973; Helm-Estabrooks, 2002; Fucetola *et al.*, 2009).

It is important to note that performance on the various components is independent, suggesting that patients have variable combinations of verbal and nonverbal deficits. The common co-occurrence is relevant for three main reasons: (1) many language assessments also load on attention and executive functions; (2) some aspects of language function require interactions between components (e.g., controlled semantic processing: Jefferies and Lambon Ralph, 2006); (3) response to therapy and recovery has been shown to relate not only to language severity but also to more domain-general functions (Lambon Ralph *et al.*, 2010; Geranmayeh *et al.*, 2017; Conroy *et al.*, 2018). Our findings thus imply that the three identified nonverbal cognitive components need to be assessed separately in future studies and in clinical practice, as they might have different implications for function and recovery. Likewise, interventions should be considered in this patient population that (a) specifically aim at improving domain-general cognitive deficits (Geranmayeh *et al.*, 2017), (b) integrate therapy of attentional or executive dysfunctions into speech-language remediation (e.g. Mayer *et al.*, 2017), and (c) adopt a multidisciplinary team approach.

Using univariate and multivariate brain-behaviour mapping approaches we identified separable structural correlates for all three nonverbal components, in addition to replicating previous findings regarding the structural correlates of the three verbal components. The clusters of all three nonverbal components overlapped to some degree with the multi-demand network (Duncan, 2010; Fedorenko *et al.*, 2013). In addition, the shift-update cluster overlapped with the dorsal attention and control network, while the inhibit-generate cluster overlapped with the ventral attention and control network (Yeo *et al.*, 2011). More specifically, the correlates of shift-update fit well with task-based functional imaging studies that report activations in lateral temporo-occipital areas for demanding visuo-spatial tasks (Fedorenko *et al.*, 2013; Humphreys and Lambon Ralph, 2017) or when location and feature information must be combined (Simpson *et al.*, 2011); both processes are inherent to shift-update. The findings for the inhibit-generate component are also in line with previous research. Although more extensive, this network of areas overlaps with the regions found in a previous study of aphasia (Lacey *et al.*, 2017) and those identified in a meta-analysis of functional imaging studies on executive functions (Niendam *et al.*, 2012).

From a methodological point of view, it is important to note the complementary differences between the interpretation of univariate and multivariate analyses (see Hebart and Baker, 2018). In general, with univariate analyses, the beta values assigned to voxels are relatively transparent (i.e., their sign and strength indicates meaningful relationships with behaviour) and thus inferences about local function are easier to make (although inference using cluster-level thresholds can only show that there is signal somewhere in the cluster; Woo *et al.*, 2014). However, univariate methods are limited by practical (i.e. multiple comparison correction, interactions between multiple variables that are typically not orthogonal) and theoretical concerns (i.e. assumption of voxel independence, mis-localisation of effects; Mah *et al.*, 2014; DeMarco and Turkeltaub, 2018; Karnath *et al.*, 2018). In contrast, multivariate methods can be used for encoding or decoding (Naselaris *et al.*, 2011; Hebart and Baker, 2018) and have different goals (i.e., to predict data from experimental conditions or to map brain status to behavioural performance and make formal predictions, respectively). These models can have problems with interpretability as feature weights become non-transparent (Haufe *et al.*, 2014; Hebart and Baker, 2018), although encoding can assist with this challenge to some degree (such as partial least squares and canonical correlation analysis). By definition, in multivariate analyses all voxel/feature weights are non-independent and thus the importance of these weights is not easy to interpret. Furthermore, analysis steps that select a subsample of weights automatically mean that the overall multivariate model has been changed and one would need to test (1) whether the contribution of a voxel to the model is greater than chance or (2) whether the contribution of a voxel to the model is stable across different samples (e.g. via bootstrapping; Kuceyeski *et al.*, 2016). Given these differences between the methods, it is striking that the multivariate models (both SVR-LSM and PRoNTo) produced beta maps that strongly correspond to the VBCM results. We assume this follows the fact that stroke tends to generate binary tissue status (intact vs. infarcted) and this will dominate the predictions of behavioural variation in all models (and are the most likely features to be selected in any form of weight truncation such as that used in SVR-LSM). There are some potential avenues to help improve interpretations of both univariate and multivariate methods in the future. First, a recent study showed that it may be possible to compute a correction for the mis-localisation caused by anatomical bias (Sperber and Karnath, 2017). Secondly, Haufe *et al.* (2014) and Naselaris *et al.* (2011) propose ways in which a decoding model can be transformed into an encoding model, which potentially leads to interpretable weights. Thirdly, alternative sparse algorithms (such as LASSO, elastic net or recursive feature selection) have the benefit of introducing a penalty for complexity and therefore provide a solution with the smallest number of features

(though the challenge of interpreting the resultant weights still holds). Finally, we note that multivariate decoding methodologies typically require a large dataset, as data are partitioned into training/test sets for cross validation. This can be practically challenging, as not only do we require neuroimaging data but also a large neuropsychological test battery to determine the underlying principal components. In a recent simulation study (Sperber *et al.*, 2019), it was suggested that approximately 100 subjects are required to have stable/reproducible beta parameter mapping, whereas for prediction of clinical outcomes the number peaked at 40 and was relatively stable from this point up to 100 cases. In the current study we obtained 32 cases (similar to Lacey *et al.*, 2017) and so future work will require replication based on larger groups sizes.

Overall, the structural correlates align with areas of different cognitive functions in healthy participants. The variable combinations of verbal and nonverbal deficits observed across post-stroke aphasia (see above) presumably reflect differential encroachment of each person's lesion on the various regions implicated for each nonverbal and verbal component and/or their connections. This would imply that interventions should target different brain regions depending on which component needs to be ameliorated to improve performance. Options to be explored include neurostimulation - for instance by targeting medial frontal areas (Sliwinska *et al.*, 2017) or pharmacology (Berthier *et al.*, 2011). It also has implications for building accurate prediction models (Price *et al.*, 2010; Hope *et al.*, 2013; Yourganov *et al.*, 2015; Yourganov *et al.*, 2016; Pustina *et al.*, 2017; Hope *et al.*, 2018; Thye and Mirman, 2018). First, it may be that predictions of language performance might be improved if the predictors include nonverbal cognitive abilities alongside patient characteristics. Secondly, it may be possible to improve prediction models of both verbal and nonverbal abilities by using these updated PCA-derived structural correlates (cf. Halai *et al.*, 2018).

In conclusion, this study was able to demonstrate that functionally distinct aspects of attention and executive skills are commonly impaired in patients with post-stroke aphasia. The assessments successfully utilised here could be adopted in clinical assessment to guide management and choices over clinical pathways. Furthermore, future investigations can explore which specific aspects of attention and executive function are crucial for effective therapy and good rehabilitation outcomes, and how these features of nonverbal abilities can be supported or boosted through novel interventions.

Acknowledgements

We would like to thank the patients and their carers for contributing to this research project. Special thanks go also to Blanca De Dios Perez for collecting some of the data used in the analysis.

Funding

This research was supported by the Swiss National Science Foundation (Grant P2BEP3_168670 to RS), by the Rosetrees Trust (Grant A1699 to ADH & MALR), by the Medical Research Council (Grant MR/R023883/1 to MALR), and by the European Research Council (Grant GAP: 670428 to MALR).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is provided in a separate file.

References

Bach FR, Lanckriet GR, Jordan MI. Multiple kernel learning, conic duality, and the SMO algorithm. Proceedings of the twenty-first international conference on Machine learning; 2004: ACM; 2004. p. 6.

Baldo JV, Paulraj SR, Curran BC, Dronkers NF. Impaired reasoning and problem-solving in individuals with language impairment due to aphasia or language delay. Front Psychol 2015; 6: 1523.

Balzer C, Berger JM, Caprez G, Gonser A, Gutbrod K, Keller M. Materialien und Normwerte fuer die neuropsychologische Diagnostik (MNND). Rheinfelden, Switzerland: Verlag Normdaten; 2011.

Basso A, De Renzi E, Faglioni P, Scotti G, Spinnler H. Neuropsychological evidence for the existence of cerebral areas critical to the performance of intelligence tasks. Brain 1973; 96(4): 715-28.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, *et al.* Voxel-based lesion–symptom mapping. Nature Neuroscience 2003; 6: 448.

Berthier ML, Pulvermuller F, Davila G, Casares NG, Gutierrez A. Drug Therapy of Post-Stroke Aphasia: A Review of Current Evidence. Neuropsychology Review 2011; 21(3): 302-17.

Brooks BL, Sherman EMS, Iverson GL, Slick DJ, Strauss E. Psychometric Foundations for the Interpretation of Neuropsychological Test Results. In: Schoenberg MR, Scott JG, editors. Little Black Book of Neuropsychology: a Syndrome-Based Approach. New York: Springer; 2011. p. 893-922.

Brownsett SLE, Warren JE, Geranmayeh F, Woodhead Z, Leech R, Wise RJS. Cognitive control and its impact on recovery from aphasic stroke. Brain 2014; 137: 242-54.

Burgess P, Shallice T. The Hayling and Brixton tests. Bury St Edmunds, England: Thames Valley Test Co Ltd; 1997.

Butler RA, Lambon Ralph MA, Woollams AM. Capturing multidimensionality in stroke aphasia: mapping principal behavioural components to neural structures. Brain 2014; 137: 3248-66.

Conroy P, Drosopoulou CS, Humphreys GF, Halai AD, Lambon Ralph MA. Time for a quick word? The striking benefits of training speed and accuracy of word retrieval in post-stroke aphasia. Brain 2018; 141: 1815-27.

Delis DC, Kaplan E, Kramer JH. Delis-Kaplan Executive function system: examiners manual. Psychological Corporation; 2001.

DeMarco AT, Turkeltaub PE. A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping. Human Brain Mapping 2018; 39(11): 4169-82.

Duncan J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends in Cognitive Sciences 2010; 14(4): 172-9.

El Hachioui H, Visch-Brink EG, Lingsma HF, van de Sandt-Koenderman MWME, Dippel DWJ, Koudstaal PJ, *et al.* Nonlinguistic Cognitive Impairment in Poststroke Aphasia: A Prospective Study. Neurorehabilitation and Neural Repair 2014; 28(3): 273-81.

Fedorenko E, Duncan J, Kanwisher N. Broad domain generality in focal regions of frontal and parietal cortex. Proceedings of the National Academy of Sciences of the United States of America 2013; 110(41): 16616-21.

Fillingham JK, Sage K, Lambon Ralph MA. Treatment of anomia using errorless versus errorful learning: are frontal executive skills and feedback important? International Journal of Language & Communication Disorders 2005; 40(4): 505-23.

Friedman NP, Miyake A. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. Cortex 2017; 86: 186-204.

Fucetola R, Connor LT, Strube MJ, Corbetta M. Unravelling nonverbal cognitive performance in acquired aphasia. Aphasiology 2009; 23(12): 1418-26.

Geranmayeh F, Chau T, Wise RJS, Leech R, Hampshire A. Domain-general subregions of the medial prefrontal cortex contribute to recovery of language after stroke. Brain 2017; 140: 1947-58.

Glosser G, Goodglass H. Disorders in executive control functions among aphasic and other brain-damaged patients. Journal of Clinical and Experimental Neuropsychology 1990; 12(4): 485-501.

Halai AD, Woollams AM, Lambon Ralph MA. Using principal component analysis to capture individual differences within a unified neuropsychological model of chronic post-stroke aphasia: Revealing the unique neural correlates of speech fluency, phonology and semantics. Cortex 2017; 86: 275-89.

Halai AD, Woollams AM, Lambon Ralph MA. Predicting the pattern and severity of chronic post-stroke language deficits from functionally-partitioned structural lesions. Neuroimage-Clinical 2018; 19: 1-13.

Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: prediction, inference and data mining. New York: Springer; 2009.

Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage 2014; 87: 96-110.

Hebart MN, Baker CI. Deconstructing multivariate decoding for the study of brain function. Neuroimage 2018; 180: 4-18.

Helm-Estabrooks N. Cognition and aphasia: a discussion and a study. Journal of Communication Disorders 2002; 35(2): 171-86.

Heuer S, Ivanova MV, Hallowell B. More Than the Verbal Stimulus Matters: Visual Attention in Language Assessment for People With Aphasia Using Multiple-Choice Image Displays. Journal of Speech Language and Hearing Research 2017; 60(5): 1348-61.

Hope TMH, Leff AP, Price CJ. Predicting language outcomes after stroke: Is structural disconnection a useful predictor? Neuroimage-Clinical 2018; 19: 22-9.

Hope TMH, Seghier ML, Leff AP, Price CJ. Predicting outcome and recovery after stroke with lesions extracted from MRI images. Neuroimage-Clinical 2013; 2: 424-33.

Humphreys GF, Lambon Ralph MA. Mapping Domain-Selective and Counterpointed Domain-General Higher Cognitive Functions in the Lateral Parietal Cortex: Evidence from fMRI Comparisons of Difficulty-Varying Semantic Versus Visuo-Spatial Tasks, and Functional Connectivity Analyses. Cerebral Cortex 2017; 27(8): 4199-212.

Jefferies E, Lambon Ralph MA. Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. Brain 2006; 129: 2132-47.

Kaller CP, Unterrainer JM, Kaiser S, Weisbrod M, Aschenbrenner S. Tower of London - Freiburg Version. Moedling, Austria: Schuhfried; 2011.

Kane MJ, Engle RW. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. Psychonomic Bulletin & Review 2002; 9(4): 637-71.

Karnath HO, Sperber C, Rorden C. Mapping human brain lesions and their functional consequences. Neuroimage 2018; 165: 180-9.

Kauhanen ML, Korpelainen JT, Hiltunen P, Maatta R, Mononen H, Brusin E, *et al.* Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. Cerebrovascular Diseases 2000; 10(6): 455-61.

Keil K, Kaszniak AW. Examining executive function in individuals with brain injury: A review. Aphasiology 2002; 16(3): 305-35.

Kuceyeski A, Navi BB, Kamel H, Raj A, Relkin N, Toglia J, *et al.* Structural Connectome Disruption at Baseline Predicts 6-Months Post-Stroke Outcome. Human Brain Mapping 2016; 37(7): 2587-601.

Kummerer D, Hartwigsen G, Kellmeyer P, Glauche V, Mader I, Kloppel S, *et al.* Damage to ventral and dorsal language pathways in acute aphasia. Brain 2013; 136: 619-29.

Kuzmina E, Weekes BS. Role of cognitive control in language deficits in different types of aphasia. Aphasiology 2017; 31(7): 765-92.

Lacey EH, Skipper-Kallal LM, Xing SH, Fama ME, Turkeltaub PE. Mapping Common Aphasia Assessments to Underlying Cognitive Processes and Their Neural Substrates. Neurorehabilitation and Neural Repair 2017; 31(5): 442-50.

Lambon Ralph MA, Snell C, Fillingham JK, Conroy P, Sage K. Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors. Neuropsychological rehabilitation 2010; 20(2): 289-305.

Lee B, Pyun SB. Characteristics of Cognitive Impairment in Patients With Post-stroke Aphasia. Ann Rehabil Med 2014; 38(6): 759-65.

Mah Y-H, Husain M, Rees G, Nachev P. Human brain lesion-deficit inference remapped. Brain 2014; 137(9): 2522-31.

Marinelli CV, Spaccavento S, Craca A, Marangolo P, Angelelli P. Different Cognitive Profiles of Patients with Severe Aphasia. Behavioural Neurology 2017: 15.

Mayer JF, Mitchinson SI, Murray LL. Addressing concomitant executive dysfunction and aphasia: previous approaches and the new brain budget protocol. Aphasiology 2017; 31(7): 837-60.

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience 2001; 24: 167-202.

Mirman D, Chen Q, Zhang YS, Wang Z, Faseyitan OK, Coslett HB, *et al.* Neural organization of spoken language revealed by lesion-symptom mapping. Nature Communications 2015; 6: 9.

Mirsky AF, Anthony BJ, Duncan CC, Ahearn MB, Kellam SG. Analysis of the elements of attention: A neuropsychological approach. Neuropsychology review 1991; 2(2): 109-45.

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cogn Psychol 2000; 41(1): 49-100.

Murray LL. Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. American Journal of Speech-Language Pathology 2012; 21(2): S51-S64.

Naranjo NP, Grande DD, Alted CG. Individual variability in attention and language performance in aphasia: a study using Conner's Continuous Performance Test. Aphasiology 2018; 32(4): 436-58.

Naselaris T, Kay KN, Nishimoto S, Gallant JL. Encoding and decoding in fMRI. Neuroimage 2011; 56(2): 400-10.

Niendam TA, Laird AR, Ray KL, Dean YM, Glahn DC, Carter CS. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. Cognitive Affective & Behavioral Neuroscience 2012; 12(2): 241-68.

Petersen SE, Posner MI. The Attention System of the Human Brain: 20 Years After. In: Hyman SE, editor. Annual Review of Neuroscience, Vol 35. Palo Alto: Annual Reviews; 2012. p. 73-89.

Power JD, Petersen SE. Control-related systems in the human brain. Current Opinion in Neurobiology 2013; 23(2): 223-8.

Price CJ, Seghier ML, Leff AP. Predicting language outcome and recovery after stroke: the PLORAS system. Nature Reviews Neurology 2010; 6(4): 202-10.

Pustina D, Coslett HB, Ungar L, Faseyitan OK, Medaglia JD, Avants B, *et al.* Enhanced Estimations of Post-Stroke Aphasia Severity Using Stacked Multimodal Predictions. Human Brain Mapping 2017; 38(11): 5603-15.

Rakotomamonjy A, Bach F, Canu S, Grandvalet Y. Simplemkl. Journal of Machine Learning Research 2008; 9: 2491-521.

Ramsey LE, Siegel JS, Lang CE, Strube M, Shulman GL, Corbetta M. Behavioural clusters and predictors of performance during recovery from stroke. Nature Human Behaviour 2017; 1(3): 10.

Rasmussen CE, Williams CK. Gaussian processes for machine learning. Cambridge, MA: MIT Press; 2006.

Raven JC. Coloured Progressive Matrices, Sets A, AB, B. London, England: H. K. Lewis; 1962.

Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, *et al.* PRoNTo: Pattern Recognition for Neuroimaging Toolbox. Neuroinformatics 2013; 11(3): 319-37.

Seghier ML, Ramlackhansingh A, Crinion J, Leff AP, Price CJ. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. Neuroimage 2008; 41(4): 1253-66.

Simic T, Rochon E, Greco E, Martino R. Baseline executive control ability and its relationship to language therapy improvements in post-stroke aphasia: a systematic review. Neuropsychological Rehabilitation 2017: 1-45.

Simpson GV, Weber DL, Dale CL, Pantazis D, Bressler SL, Leahy RM, *et al.* Dynamic Activation of Frontal, Parietal, and Sensory Regions Underlying Anticipatory Visual Spatial Attention. Journal of Neuroscience 2011; 31(39): 13880-9.

Sliwinska MW, Violante IR, Wise RJS, Leech R, Devlin JT, Geranmayeh F, *et al.* Stimulating Multiple-Demand Cortex Enhances Vocabulary Learning. Journal of Neuroscience 2017; 37(32): 7606-18.

Smits CHM, Smit JH, vandenHeuvel N, Jonker C. Norms for an abbreviated Raven's Coloured Progressive Matrices in an older sample. Journal of Clinical Psychology 1997; 53(7): 687-97.

Sperber C, Karnath HO. Impact of Correction Factors in Human Brain Lesion-Behavior Inference. Human Brain Mapping 2017; 38(3): 1692-701.

Sperber C, Wiesen D, Karnath H-O. An empirical evaluation of multivariate lesion behaviour mapping using support vector regression. Human Brain Mapping 2019; 40(5): 1381-90.

Sturm W, Willmes K, Orgass B, Hartje W. Do specific attention deficits need specific training? Neuropsychological Rehabilitation 1997; 7(2): 81-103.

Thompson HE, Almaghyuli A, Noonan KA, Barak O, Lambon Ralph MA, Jefferies E. The contribution of executive control to semantic cognition: Convergent evidence from semantic aphasia and executive dysfunction. Journal of neuropsychology 2018; 12(2): 312-40.

Thye M, Mirman D. Relative contributions of lesion location and lesion size to predictions of varied language deficits in post-stroke aphasia. Neuroimage-Clinical 2018; 20: 1129-38.

Tipping ME. Spare bayesian learning and the relevance vector machine. Journal of Machine Learning Research 2001; 1: 211-44.

Tyler LK, Marslen-Wilson W, Stamatakis EA. Dissociating neuro-cognitive component processes: voxel-based correlational methodology. Neuropsychologia 2005; 43(5): 771-8.

van de Sandt-Koenderman WME, van Harskamp F, Duivenvoorden HJ, Remerie SC, van der Voort-Klees YA, Wielaert SM, *et al.* MAAS (Multi-axial Aphasia System): realistic goal setting in aphasia rehabilitation. International Journal of Rehabilitation Research 2008; 31(4): 314-20.

Villard S, Kiran S. Between-session intra-individual variability in sustained, selective, and integrational non-linguistic attention in aphasia. Neuropsychologia 2015; 66: 204-12.

Villard S, Kiran S. To what extent does attention underlie language in aphasia? Aphasiology 2017; 31(10): 1226-45.

Wall KJ, Cumming TB, Copland DA. Determining the association between language and cognitive Tests in Poststroke aphasia. Frontiers in Neurology 2017; 8: 9.

Woo CW, Krishnan A, Wager TD. Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. Neuroimage 2014; 91: 412-9.

Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, *et al*. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology 2011; 106(3): 1125-65.

Yourganov G, Fridriksson J, Rorden C, Gleichgerrcht E, Bonilha L. Multivariate Connectome-Based Symptom Mapping in Post-Stroke Patients: Networks Supporting Language and Speech. Journal of Neuroscience 2016; 36(25): 6668-79.

Yourganov G, Smith KG, Fridriksson J, Rorden C. Predicting aphasia type from brain damage measured with structural MRI. Cortex 2015; 73: 203-15.

Zhang Y, Kimberg DY, Coslett HB, Schwartz MF, Wang Z. Multivariate lesion-symptom mapping using support vector regression. Human Brain Mapping 2014; 35(12): 5861-76.

Zimmermann P, Fimm B. Test for attentional performance (TAP). Herzogenrath, Germany: PsyTest; 1995.

Figures

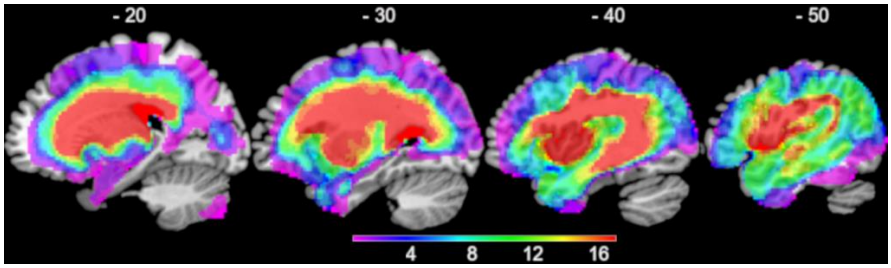


Figure 1. Overlap of the 38 patients' lesions. The image threshold corresponds to the maximum overlap (16 patients).

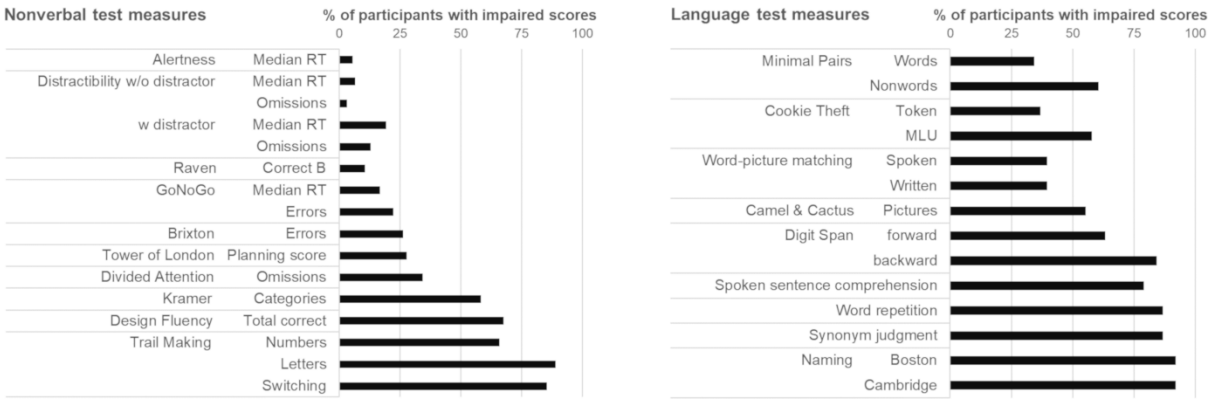


Figure 2. Percentage of participants with impaired performance on each measure of the nonverbal tests (left) and language tests (right).

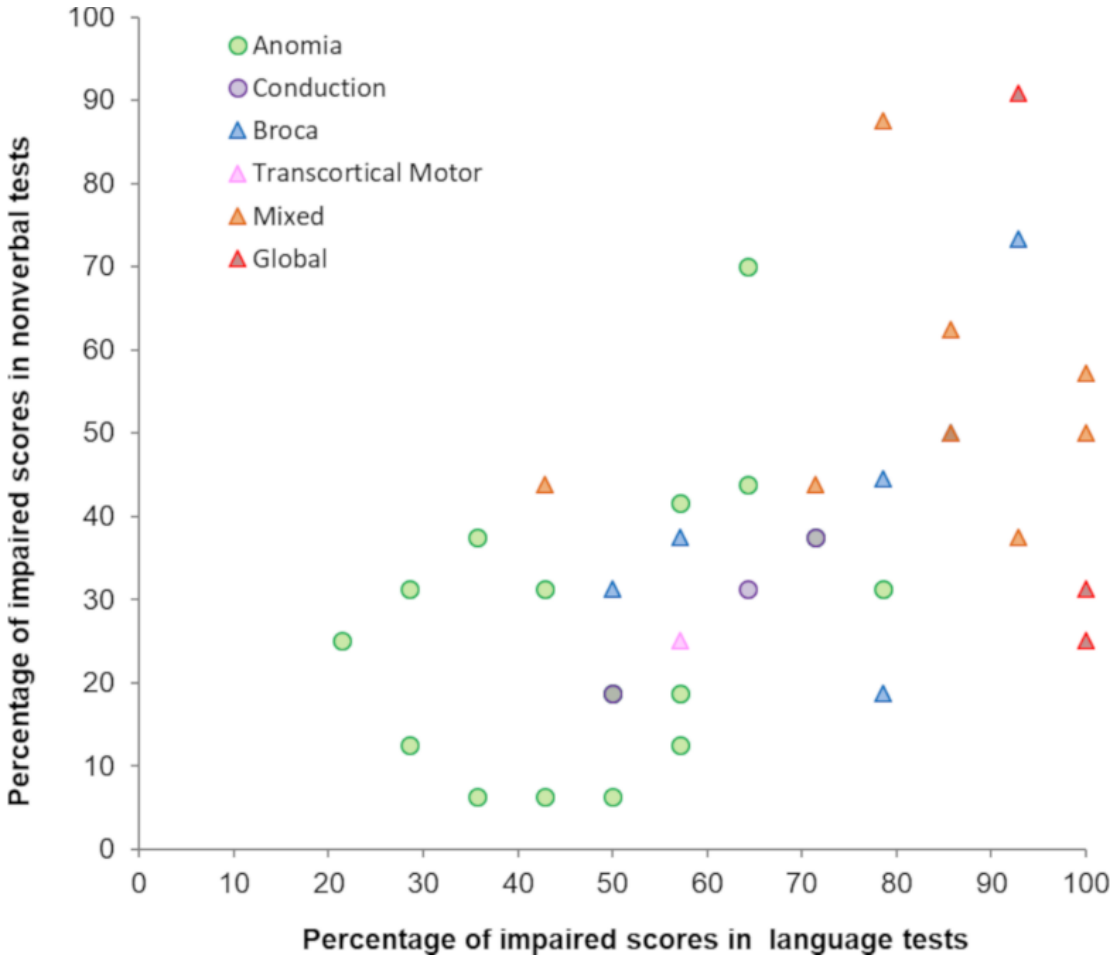


Figure 3. Patients' overall impairment in the nonverbal versus language tests. The percentages of impaired scores correlated significantly (r_s = 0.591, $p < 0.01$, n = 38, also if patient characteristics were accounted for by means of partial correlations). Symbols and colours denote an individual's aphasia type based on the BDAE (triangles for non-fluent, circles for fluent patients, for colours see figure legend). More saturated or differently coloured symbols denote two patients in the same spot.

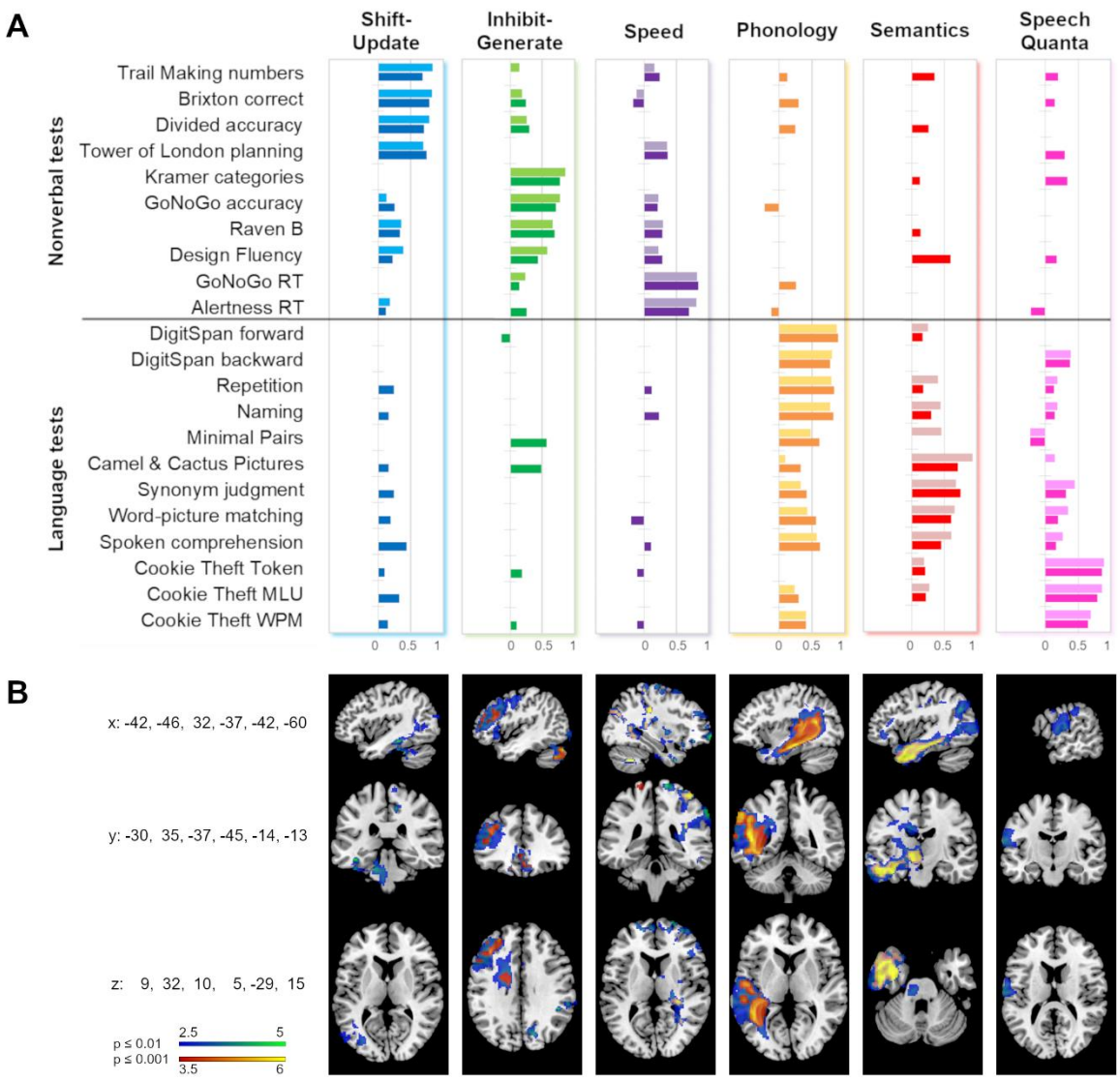


Figure 4. Component loadings and structural correlates associated with each component.

A) The darker colours (blue, green, purple, orange, red, pink) represent the loadings on the six components from the combined PCA. The lighter coloured bars represent the loadings on the three components in the separate nonverbal-only PCA (first three columns) and the language-only PCA (last three columns). Loadings < 0.1 are not depicted. MLU = mean length of utterance; WPM = words per minute

B) Structural correlates associated with each component from the combined PCA. Clusters shown in blue-green were obtained by applying a voxel-level threshold of $p \leq 0.01$, clusters in red-yellow correspond to a voxel-level threshold of $p \leq 0.001$. A family-wise error correction of $p \leq 0.05$ was applied to all clusters. The respective coordinates in MNI-space are indicated on the left side. Figures are in neurological convention (left is left).

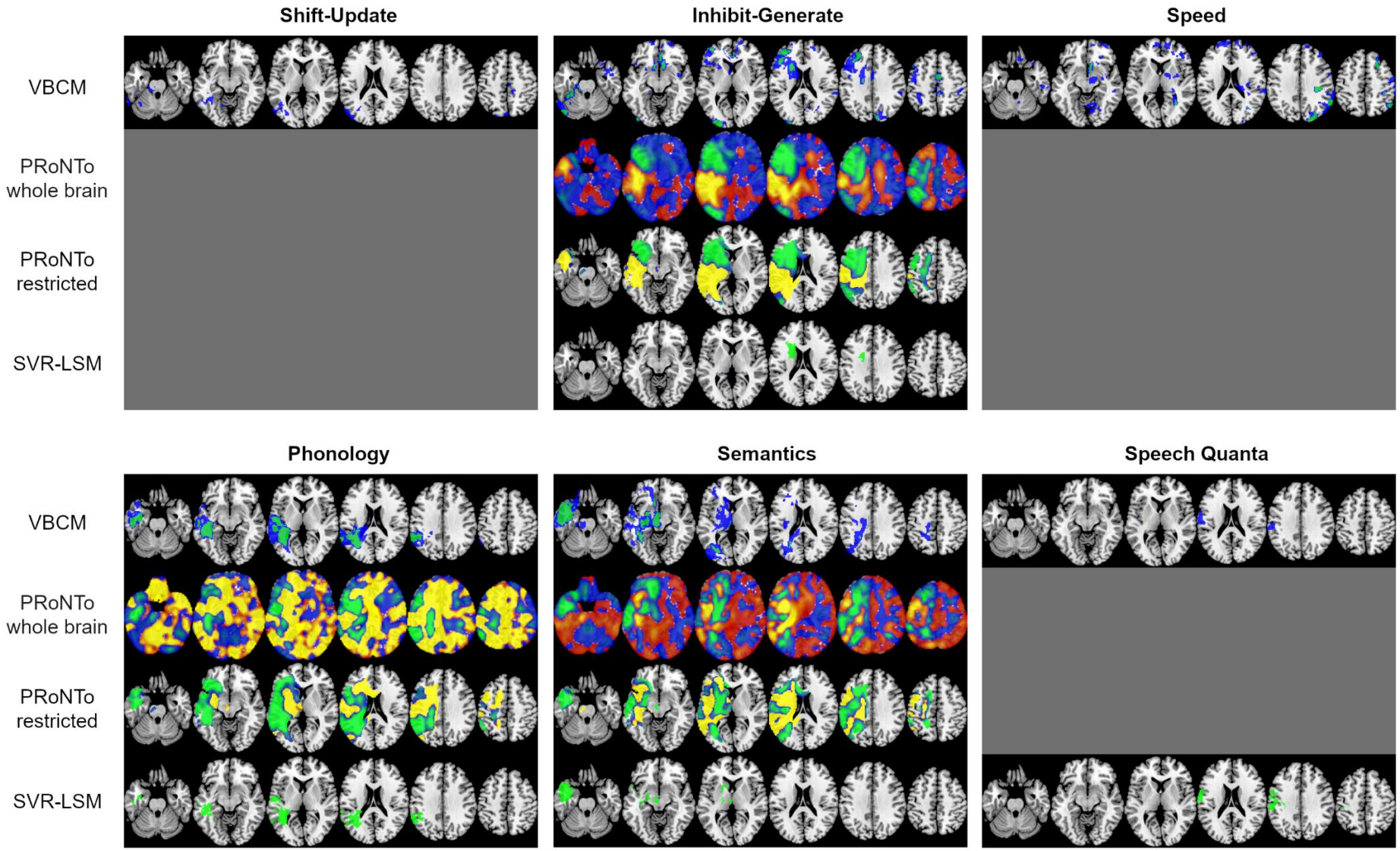

Shift-Update
Inhibit-Generate
Speed
VBCM
PRoNTo whole brain
PRoNTo restricted
SVR-LSM
Phonology
Semantics
Speech Quanta
VBCM
PRoNTo whole brain
PRoNTo restricted
SVR-LSM

Figure 5. Comparison of brain-behaviour mapping results based on the four different methodological approaches.

The significant VBCM clusters are shown in blue (voxel-level threshold 0.01) and green (voxel-level threshold 0.001), a family-wise error correction of $p \leq 0.05$ was applied to all clusters, and images are thresholded at the respective minimum t-value. The PRoNTo results depict the weights for the winning model if significant (see text), either including the whole brain space or restricting it to lesion territory ($N > 3$). They are thresholded from -0.005 to -0.0001 (green-blue) and 0.0001 to 0.005 (red-yellow). The negative weights are considered as more meaningful in this approach. The SVR-LSM images show voxels with significant beta weights after permutation testing (N=10,000, voxel-wise $p < 0.005$ and cluster-wise $p < 0.05$). MNI coordinates of slices, from left to right, are z = -25, -10, 5, 20, 35, 50 and they are in neurological convention (left is left). A grey surface indicates that no significant results were found for the respective component and methodological approach.

Tables

Table 1. Spearman correlations within and between severity of nonverbal and language impairment, component scores, and patient characteristics.

		Severity		Nonverbal PCA			Patient characteristics			
		nonverbal	verbal	S-U	I-G	Speed	Lesion	Age	Edu	Time post
Severity	verbal	.535*		-.521*	-.105	.150				
	nonverbal			**-.676***	-.283	-.110				
Verbal PCA	Phon	-.216	**-.719***	.261	-.294	.025	-.208	-.126	-.209	.131
	Sem	-.316	-.383*	.421*	.373*	.087	-.396*	-.433*	.288	-.115
	SQ	-.362*	-.427*	.443*	.097	-.283	-.504*	.068	.190	-.240
Combined PCA	Phon	-.164	**-.744***	.216	-.245	-.062	-.238	-.121	-.213	.126
	S-U	-.530*	-.259	.871*	-.063	-.018	-.308	-.445*	.143	-.393*
	I-G	-.235	-.173	-.109	.905*	-.184	.050	-.436*	.312	-.208
	SQ	-.325	-.349	.214	.178	-.195	-.376*	.053	.120	-.142
	Sem	-.139	-.118	.194	.201	-.010	-.370*	-.014	.247	.061
	Speed	-.172	.103	-.122	.037	.902*	.177	-.305	-.003	.294
Patient characteristics	Time post	.196	.151	-.381*	-.187	.240	.389*	.094	-.123	
	Edu	-.279	-.061	.254	.494*	-.174	-.132	-.321		
	Age	.323	.332	-.441*	-.455*	-.254	.251			
	Lesion	.353*	.555*	-.518*	-.146	.156				

Note: * p< 0.05 two-tailed; bold = significant after Bonferroni correction (p < 0.0004); n=32;
Phon = Phonology, S-U = Shift-Update, I-G = Inhibit-Generate, SQ = Speech Quanta, Sem = Semantics, Time post = Time post stroke, Edu = Education

Table 2. Clusters and peaks associated with the nonverbal and language components.

Component	Extent	Location	L/R	Z	x	y	z
Shift-Update	2032	Temporal fusiform cortex pos	L	4.29	-40	-32	-16
		Temporal fusiform cortex pos	L	3.69	-38	-34	-30
		Inferior longitudinal fas	L	3.58	-42	-36	-14
		Temporal fusiform cortex pos	L	3.32	-42	-30	-28
		Inferior temporal gyrus temocc	L	3.27	-60	-56	-22
		Occipital fusiform gyrus	L	3.22	-26	-64	-16
		Lateral occipital cortex sup	L	3.22	-56	-72	20
		Inferior temporal gyrus temocc	L	3.19	-56	-50	-22
	990	Left Precuneous cortex	L	4.04	-2	-62	66
		Postcentral gyrus	R	3.78	10	-36	72
		Precentral gyrus	R	3.76	10	-32	50
		Corticospinal tract	R	3.64	16	-34	54
		Superior parietal lobule	R	3.57	10	-48	72
Inhibit-Generate	1270	Frontal pole	L	5.00	-20	56	12
		Frontal pole	L	3.99	-28	50	16
		Middle frontal gyrus	L	3.94	-38	28	32
		Frontal pole	L	3.91	-28	42	36
		Frontal pole	L	3.63	-38	52	0
		Middle frontal gyrus	L	3.60	-44	24	24
		Inferior frontal gyrus p tri	L	3.50	-40	32	18
		Middle frontal gyrus	L	3.45	-52	18	30
	530	Subcallosal cortex	R	5.06	6	26	-14
		Accumbens	R	4.95	8	16	-6
		Cingulate gyrus ant	R	4.01	2	36	2
		Accumbens	L	3.88	-8	12	-8
		Subcallosal cortex	L	3.84	-12	28	-16
	447	Occipital pole	L	4.27	-24	-96	16
		Occipital pole	L	4.21	-22	-94	10
		Lateral occipital cortex inf	L	3.75	-42	-88	-10
	414	Supplementary motor cortex	L	3.55	-16	-10	34
		Anterior thalamic radiation	L	3.44	-20	20	18
		Superior longitudinal fas	L	3.34	-22	-4	30
	337	Supplementary motor cortex	R	4.50	6	-12	46
		Cingulate gyrus pos	R	4.42	4	-22	42
Speed	369	Lateral occipital cortex sup	R	4.51	26	-86	34
		Occipital pole	R	4.50	22	-90	32
	355	Angular gyrus	R	4.53	62	-54	38
		Lateral occipital cortex sup	R	4.12	54	-62	28
Phonology	5688	Inferior longitudinal fas	L	5.99	-42	-30	-16
		Inferior longitudinal fas	L	5.70	-42	-34	-14
		Temporal fusiform cortex pos	L	5.58	-40	-24	-18
		Inferior temporal gyrus post	L	5.11	-50	-18	-24
		Inferior temporal gyrus temocc	L	4.93	-48	-46	12
		Supramarginal gyrus pos	L	4.61	-60	-48	34
		Angular gyrus	L	4.56	-40	-54	14
		Middle temporal gyrus temocc	L	4.47	-42	-54	8
		Planum temporale	L	4.37	-36	-32	14

Semantics	4994	Temporal fusiform cortex pos	L	5.48	-40	-30	-16
		Inferior temporal gyrus post	L	5.15	-52	-16	-24
		Parahippocampal gyrus ant	L	5.12	-34	-6	-26
		Thalamus	L	5.12	-10	-22	-4
		Temporal Pole	L	5.04	-52	10	-36
		Hippocampus	L	5.02	-34	-10	-24
		Anterior thalamic radiation	L	4.90	-10	-18	-8
		Anterior thalamic radiation	L	4.82	-8	-18	-12
		Inferior longitudinal fas	L	4.71	-40	-36	-14
Speech Quanta	1010	Postcentral gyrus	L	3.24	-66	-16	16
		Postcentral gyrus	L	2.80	-56	-12	28
		Supramarginal gyrus ant	L	2.63	-62	-28	36
		Postcentral gyrus	L	2.51	-50	-24	38
		Postcentral gyrus	L	2.50	-44	-24	44
		Precentral gyrus	L	2.37	-60	0	38

Note: Only clusters with cluster-level FWEc $p \leq 0.001$ are shown in the table.
ant = anterior, fas = fasciculus, inf = inferior, p tri = pars triangularis, pos = posterior, temocc= temporo-occipital

SUPPLEMENTARY MATERIAL

Contents

Abbreviations

BDAE = Boston Diagnostic Aphasia Examination

CCT = Camel & Cactus Test

MLU = Mean length of utterance

NW = Nonword

RT = Reaction Time

TMT = Trail Making Test

TOL = Tower of London

W = Word

WPM = Words per minute

Supplementary Table 1. Participant background information.

Subj	Age	Gender	Education	Years post-stroke	Lesion Volume	BDAE classification*
1	55	m	17	9	11915	Broca
2	55	f	12	12	9767	Anomia
3	71	m	11	8	8788	Mixed Nonfluent
4	61	m	11	17	18392	Broca
5	72	m	12	10	41379	Global
6	47	m	11	6	8437	Anomia
7	76	m	11	4	22732	Mixed Nonfluent
8	50	f	11	8	6975	Anomia
9	79	f	11	7	13577	Anomia
10	63	f	19	6	9159	Anomia
11	80	m	13	5	34242	Mixed Nonfluent
12	71	m	11	4	3311	Anomia
13	71	m	13	6	13080	Broca
14	62	m	11	5	16433	Anomia
15	70	m	13	7	33239	Global
16	52	m	13	8	22948	Anomia
17	48	f	16	4	3897	Conduction
18	84	m	10	3	12131	Broca
19	46	f	13	5	18948	Anomia
20	75	f	11	6	23863	TMA
21	76	f	11	15	12057	Mixed Nonfluent
22	45	f	16	3	175	Anomia
23	66	m	11	4	33239	Mixed Nonfluent
24	69	m	11	5	31317	Mixed Nonfluent
25	81	m	11	6	33678	Mixed Nonfluent
26	47	m	11	3	10409	Anomia
27	59	f	11	24	12699	Anomia
28	68	m	11	2	4879	Conduction
29	53	m	11	7	37822	Global
30	88	m	9	2	8528	Anomia
31	75	m	11	11	36877	Broca
32	67	m	17	2	6557	Conduction
33	57	m	16	2	6974	Anomia
34	66	m	10	7	6607	Anomia
35	50	m	19	2	4538	Anomia
36	51	m	11	3	14681	Anomia
37	56	f	11	2	10081	Mixed Nonfluent
38	69	m	12	7	37907	Broca/Mixed Nonfluent

* This information is provided for completeness only. Issues with these diagnostic categories are discussed elsewhere (e.g. Dronkers and Larsen, 2001; Butler *et al.*, 2014; Kasselimis *et al.*, 2017).

Supplementary Table 2. Participants' scores in tests of executive function and attention.

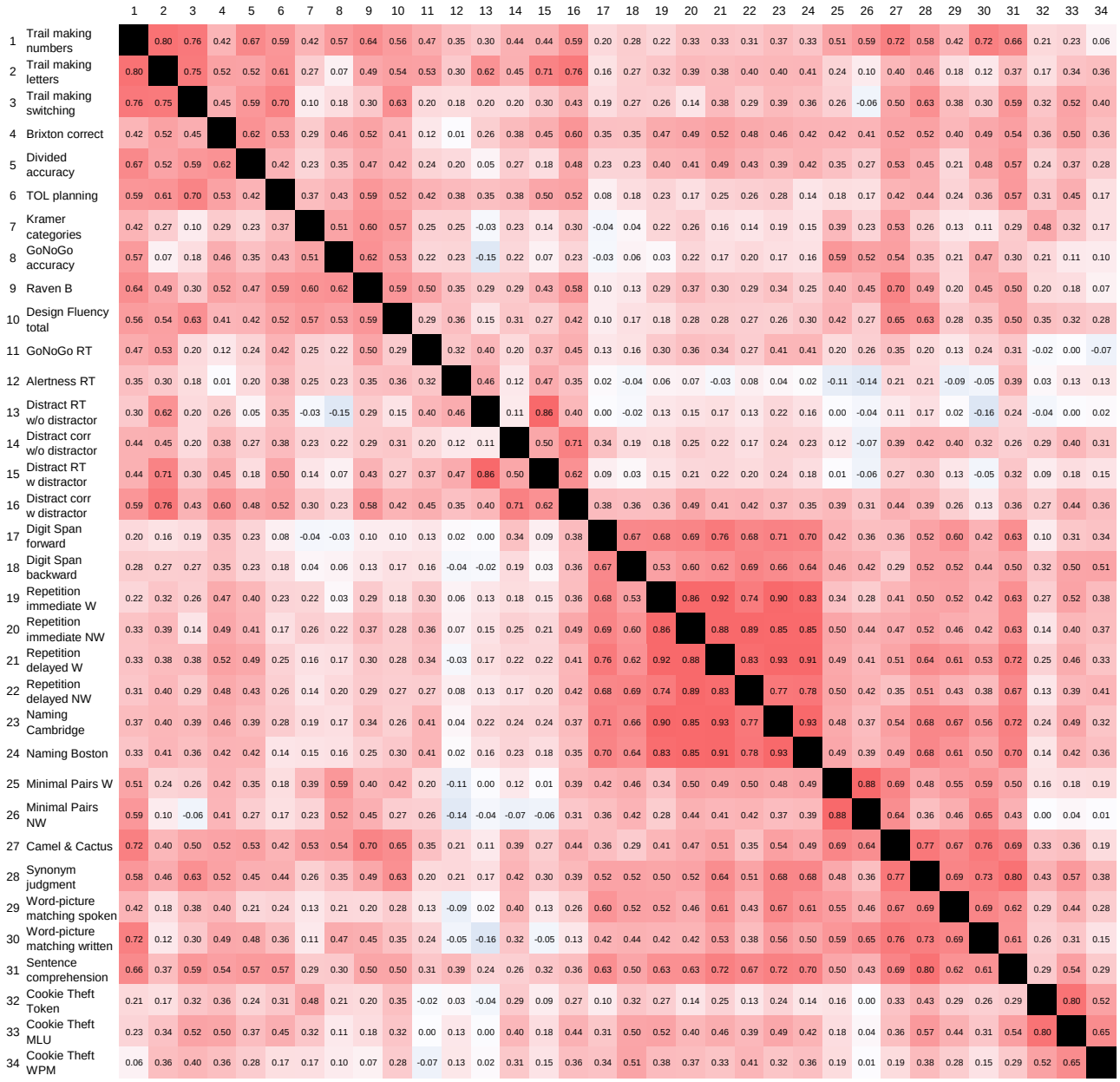
Subj	TMT			Brixton	Divided	TOL	Kramer	Raven B	Design Fluency	GoNoGo		Alertness	Distract w/o distractor		Distract w distractor	
	num	let	swi	err	omi	plan	cat	cor	cor	RT	err	RT	RT	omi	RT	omi
1	58	**82**	**161**	**27**	**6**	**6**	3	11	32	438	0	221	566	0	**624**	0
2	**74**	52	108	22	2	11	**2**	10	26	446	1	277	510	0	512	0
3	82	**121**	**277**	18	**14**	**6**	**0**	7	**14**	**569**	5	291	424	1	482	1
4	47	**163**	**309**	24	5	11	**2**	7	19	426	**9**	229	593	0	**639**	0
5	**a 500**	**a 500**	**a 500**	**34**	a	**0**	**0**	**1**	**4**	**737**	**11**	208	a	a	a	a
6	**62**	**95**	**182**	17	4	12	**2**	12	**17**	485	1	221	528	1	538	0
7	**540**	**a 240**	n/ad	23	n/ad	**0**	**0**	**3**	**3**	a	a	**1223**	n/ad	n/ad	n/ad	n/ad
8	**102**	**142**	**393**	**31**	**17**	8	4	10	**15**	407	**7**	211	482	0	517	1
9	**214**	n/ad	n/ad	**38**	**9**	**0**	**0**	6	**11**	525	**14**	234	n/ad	n/ad	n/ad	n/ad
10	44	**85**	118	19	2	13	5	12	24	452	1	256	506	0	520	0
11	**194**	**220**	**372**	22	**18**	6	**1**	9	**9**	557	3	324	586	0	**647**	1
12	51	76	**203**	19	1	11	**1**	5	**14**	478	3	207	502	0	524	0
13	59	**104**	**136**	**31**	n/av	15	n/av	10	n/av	n/av	n/av	n/av	n/av	n/av	n/av	n/av
14	**150**	**151**	**220**	12	4	14	3	11	24	527	0	194	513	0	463	0
15	**114**	**129**	**384**	28	2	10	**0**	10	**15**	484	4	214	426	0	442	0
16	52	**125**	**378**	20	2	16	**2**	9	**10**	462	**6**	286	519	0	513	0
17	**53**	**82**	**149**	11	**7**	13	3	11	22	**603**	3	228	475	1	499	**4**
18	**201**	**352**	n/av	21	n/av	n/av	**0**	10	**8**	539	3	259	n/av	n/av	n/av	n/av
19	**64**	**85**	122	13	2	16	**2**	10	24	437	2	214	462	0	437	0
20	64	**164**	**240**	27	5	15	2	8	**15**	477	**8**	256	497	0	527	1
21	**150**	**400**	**420**	**36**	**18**	**0**	2	**4**	**12**	**720**	3	320	**634**	0	650	**7**
22	**52**	**71**	**167**	16	1	12	3	11	**12**	534	2	285	558	0	600	0
23	**153**	**292**	**484**	19	6	**7**	3	10	**12**	**593**	2	274	**609**	1	**700**	3
24	**188**	**173**	**378**	**36**	**19**	9	3	7	**14**	498	3	214	456	1	484	2
25	**349**	n/ad	n/ad	**36**	n/av	n/av	n/av	6	**9**	360	**14**	256	n/av	n/av	n/av	n/av
26	**63**	**54**	**155**	13	2	15	**2**	11	**13**	403	4	233	497	0	493	0
27	**111**	**122**	**240**	20	4	**7**	**1**	9	**18**	380	1	241	529	0	555	1
28	55	**131**	**188**	14	3	12	**1**	7	20	542	3	203	542	0	568	0
29	**78**	**135**	**227**	26	5	7	3	9	**15**	437	1	225	511	1	555	1
30	91	107	**232**	16	**14**	6	**1**	5	**10**	596	1	**399**	a	a	a	a
31	**231**	**391**	**a 393**	**38**	a	**3**	**0**	**3**	**7**	532	5	272	594	**3**	**832**	**10**
32	**94**	**207**	**195**	21	6	14	3	9	24	**550**	0	266	584	0	**635**	0
33	34	**85**	**178**	16	0	12	3	8	29	505	1	244	494	0	505	0
34	54	67	155	16	6	14	**1**	10	19	371	4	231	492	0	507	1
35	32	**65**	127	14	0	14	3	11	26	421	2	219	474	0	487	0
36	**138**	**189**	**494**	17	3	**3**	3	10	**8**	373	0	247	484	0	479	0
37	**74**	**131**	**312**	**31**	**18**	10	**1**	10	**16**	374	1	254	518	1	556	**4**
38	**167**	**197**	**a 360**	28	**14**	14	**0**	**4**	**6**	529	**6**	224	526	0	541	3
n/imp	38/25	36/32	34/29	38/10	32/11	36/10	36/21	38/4	37/25	36/6	36/8	37/2	31/2	31/1	31/6	31/4
more imp	84.2	94.1	100	47.4	78.6	58.8	61.1	21.1	89.5	27.8	33.3	10.5	15.4	7.7	30.8	23.1
less imp	47.4	84.2	73.7	5.3	0	0	55.6	0	44.4	5.6	11.1	0	0	0	11.1	5.6

Notes: n/av = not available (organisational), n/ad = not administered (based on performance in other tests), a = abandoned (stopped after instruction or during test), imp = impaired, bold = score below cut-off thus considered as impaired¸ more/less impaired indicates the percentage of participants with impaired scores for the two subgroups of more/less impaired patients based on a median split of nonverbal severity

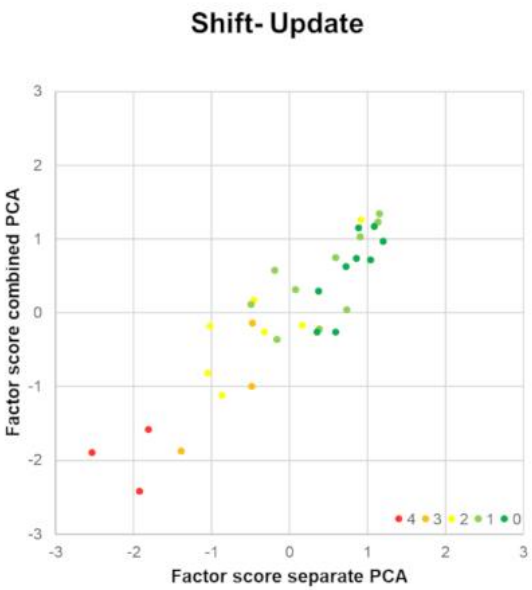
Supplementary Table 3. Participants' scores in language tests.

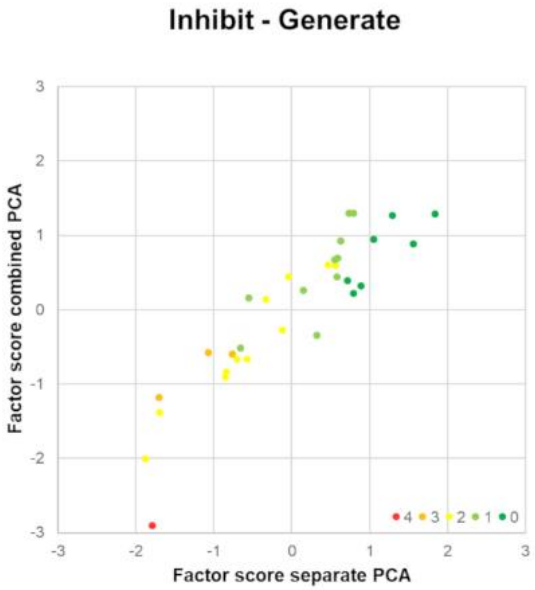
	Digit Span		Repetition immediate		Repetition delayed		Naming		Minimal Pairs		CCT	Synonym	Word-picture matching		Compre-hension	Cookie Theft		
Subj	forward	backward	W	NW	W	NW	Cambridge	Boston	W	NW	pictures	judgment	Spoken	Written	Sentence	Token	MLU	WPM
1	**3**	**0**	**59**	18	55	3	**48**	**43**	67	**58**	63	92	63	63	**18**	38	**6.86**	32.6
2	7	6	80	27	79	27	60	53	72	**66**	**51**	**90**	64	64	**27**	60	11.8	212
3	5	4	80	26	77	24	61	**40**	68	67	**47**	**80**	63	**60**	**25**	38	**8.17**	50.7
4	8	7	**62**	22	70	25	**56**	**47**	71	72	**48**	**86**	64	64	28	**30**	**7.4**	18
5	**2**	**0**	**30**	0	0	0	**0**	**0**	**38**	**16**	**17**	**47**	**37**	**20**	**4**	*32*	**8.2**	87.3
6	**4**	4	80	30	80	27	**57**	**33**	70	71	58	**87**	64	64	28	56	14.8	56
7	**3**	**2**	**40**	8	49	5	**30**	**23**	71	71	**38**	**55**	**62**	**60**	**4**	*33*	**3.18**	7.42
8	**3**	**2**	**75**	16	33	3	**56**	**30**	69	**65**	59	**75**	64	**62**	**20**	47	10	49.5
9	6	**3**	**46**	7	44	4	**34**	**30**	67	69	55	**79**	**62**	63	**22**	**25**	**4**	25.4
10	**4**	4	**72**	15	71	14	**57**	**26**	67	**54**	61	93	64	64	**26**	315	19.6	106
11	5	**3**	**35**	7	28	2	**25**	**6**	67	69	**48**	**63**	64	**62**	**11**	**11**	**1.4**	5.16
12	7	**3**	**77**	25	77	21	**54**	**37**	69	69	**54**	**76**	64	64	**25**	74	13	92.5
13	**2**	**2**	**41**	3	34	1	**35**	**14**	**62**	**64**	59	**79**	63	64	**18**	**31**	**6.83**	29.5
14	**3**	**2**	**76**	13	76	14	**46**	**34**	69	71	60	**86**	63	64	**25**	94	11.9	56.4
15	**0**	**0**	**1**	0	0	0	**0**	**0**	**56**	**59**	**43**	**72**	**50**	**60**	**15**	**0**	**0**	0
16	**3**	**2**	**69**	8	51	4	**49**	**25**	**63**	**58**	**51**	**72**	**61**	63	**24**	122	16.6	56.3
17	**3**	**0**	**31**	0	19	0	**22**	**9**	**58**	**57**	59	**86**	64	64	**23**	38	**3.15**	27.5
18	7	**2**	**52**	10	53	5	**42**	**23**	**59**	**62**	**52**	**82**	**62**	63	29	**18**	**8**	17.1
19	**4**	**2**	**74**	18	71	12	**54**	**36**	70	69	58	**84**	64	64	**27**	38	**6.83**	23.8
20	6	**0**	**68**	17	73	13	**39**	**23**	68	**58**	57	**80**	64	64	28	**25**	**4.71**	19.7
21	**2**	**0**	**0**	0	0	0	**3**	**3**	**61**	**53**	**50**	**71**	64	**61**	**18**	46	**5.22**	15.3
22	5	4	80	26	80	24	60	53	71	71	**54**	**83**	63	63	29	**22**	10.3	69.5
23	**3**	**0**	**71**	16	31	6	**9**	**2**	**62**	70	**42**	**45**	**40**	**59**	**15**	58	**5.17**	32.8
24	**2**	**0**	**4**	0	0	0	**1**	**0**	69	**66**	**38**	**47**	**50**	**36**	**16**	**1**	**0.17**	1
25	**2**	**0**	**50**	5	26	1	**16**	**10**	**31**	**34**	**34**	**44**	**55**	**39**	**14**	**12**	**4.75**	7.5
26	**4**	**2**	**70**	21	58	20	**56**	**44**	71	71	58	**84**	64	64	30	48	14.5	44.3
27	5	**0**	**76**	18	75	17	**54**	**46**	66	**59**	**53**	**85**	63	63	**26**	**23**	**4.67**	37.3
28	**2**	**2**	**11**	0	6	1	**4**	**3**	71	67	57	**67**	64	63	**19**	55	9.29	110
29	**0**	**0**	**0**	0	0	0	**0**	**0**	**62**	**61**	**40**	**50**	**38**	**42**	**11**	**0**	**0**	0
30	5	**3**	**61**	9	57	3	**35**	**33**	**55**	**57**	**48**	91	63	63	**24**	122	17	97.6
31	**2**	**2**	**52**	3	44	4	**38**	**31**	67	**66**	**42**	**67**	**59**	63	**20**	**19**	**3.29**	19.7
32	**2**	**2**	**17**	1	4	1	**2**	**1**	69	69	57	**86**	**62**	**62**	**20**	203	19	94.4
33	**4**	**3**	**73**	24	79	14	**56**	**47**	67	**62**	59	92	**61**	**62**	30	116	16.8	47.7
34	5	**3**	**57**	14	54	9	**50**	**33**	66	**63**	56	**84**	64	64	**27**	69	11.6	55.2
35	5	**3**	**68**	15	72	11	**50**	**46**	70	**63**	59	92	64	64	32	94	13.6	49.5
36	5	**3**	79	29	76	13	**58**	51	68	69	61	**77**	64	64	**20**	74	**6.69**	74
37	**0**	**0**	**0**	0	0	0	**0**	**0**	**51**	**54**	**46**	**67**	**34**	**58**	**14**	**0**	**0**	0
38	**2**	**2**	**40**	2	23	2	**31**	**14**	**40**	**33**	**26**	**67**	**54**	**59**	**14**	73	11.7	46.6
Total impaired	24	32	33	-	-	-	35	35	13	23	21	33	15	15	30	14	22	-

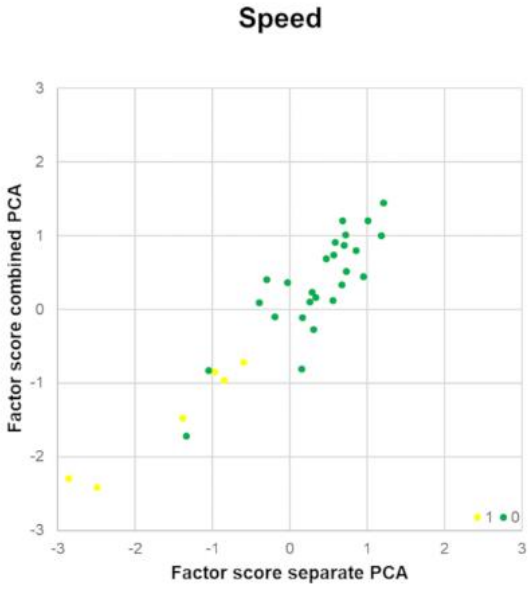
Notes: bold = impaired; - in the last row indicates that no norms were available

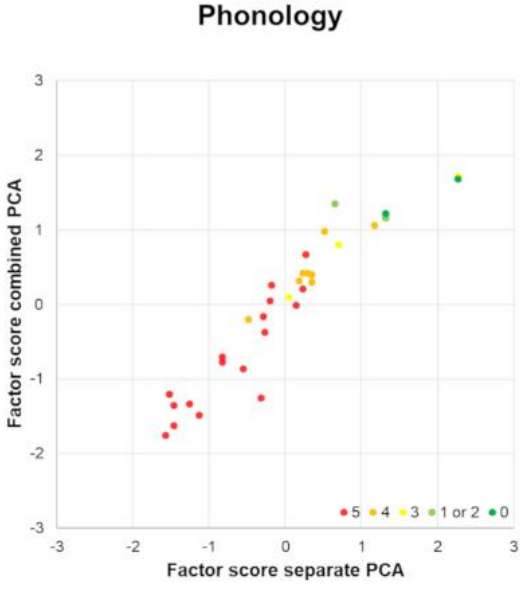


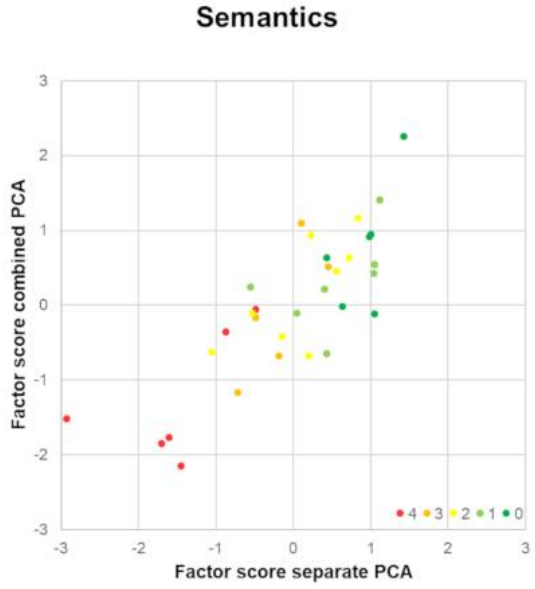
Supplementary Fig. 1. Pairwise Pearson correlations between all raw scores. Some scores have been transformed so that a higher score always corresponds to better performance (e.g. by taking the inverse of reaction time measures).

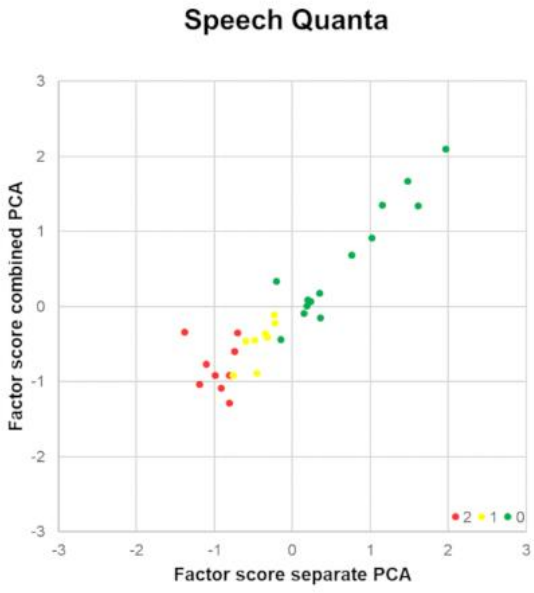

Shift- Update
Factor score combined PCA
Factor score separate PCA
4 3 2 1 0


Inhibit - Generate
Factor score combined PCA
Factor score separate PCA
4 3 2 1 0

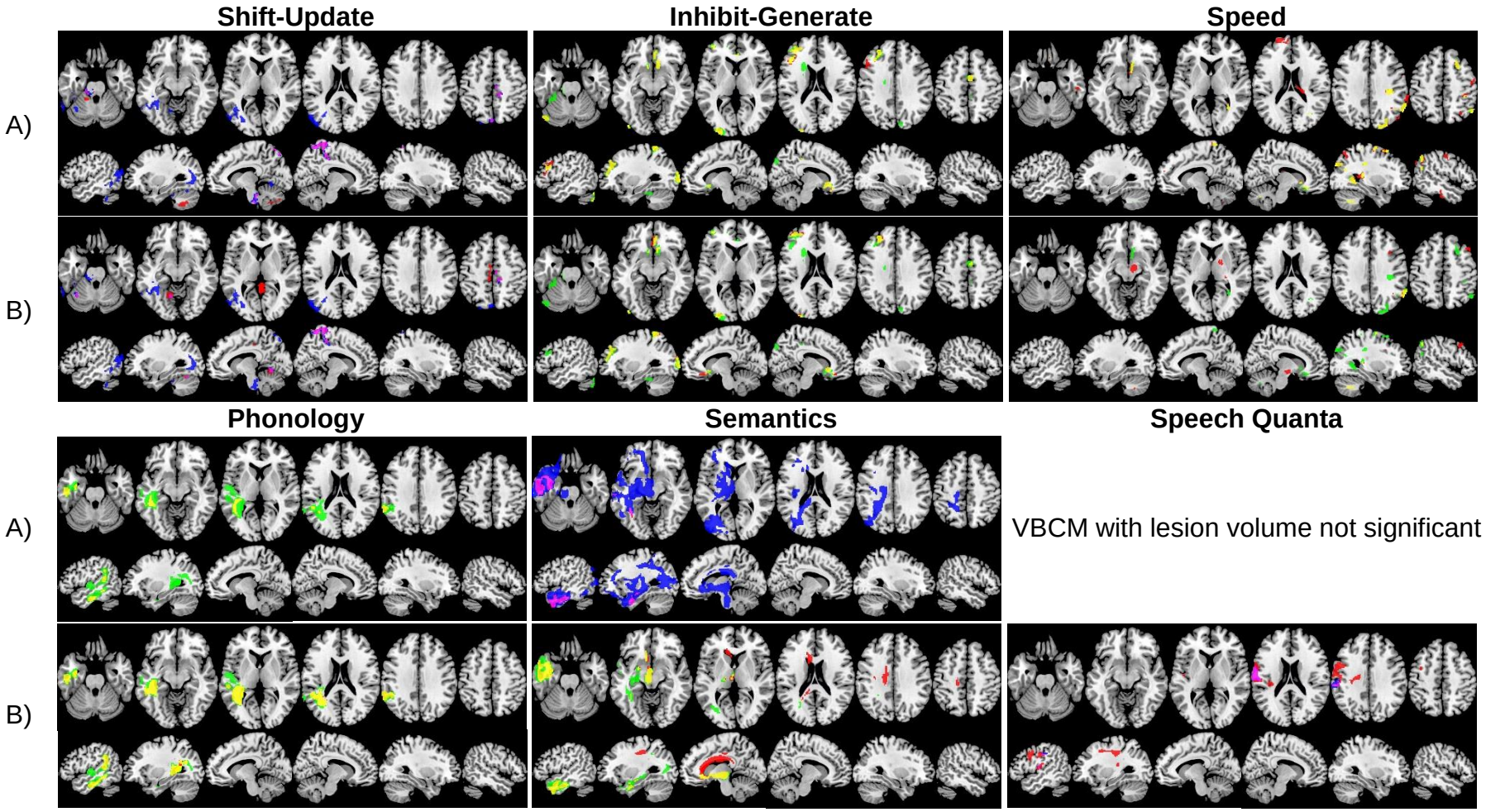

Speed
Factor score combined PCA
Factor score separate PCA
1 0


Phonology
Factor score combined PCA
Factor score separate PCA
5 4 3 1 or 2 0

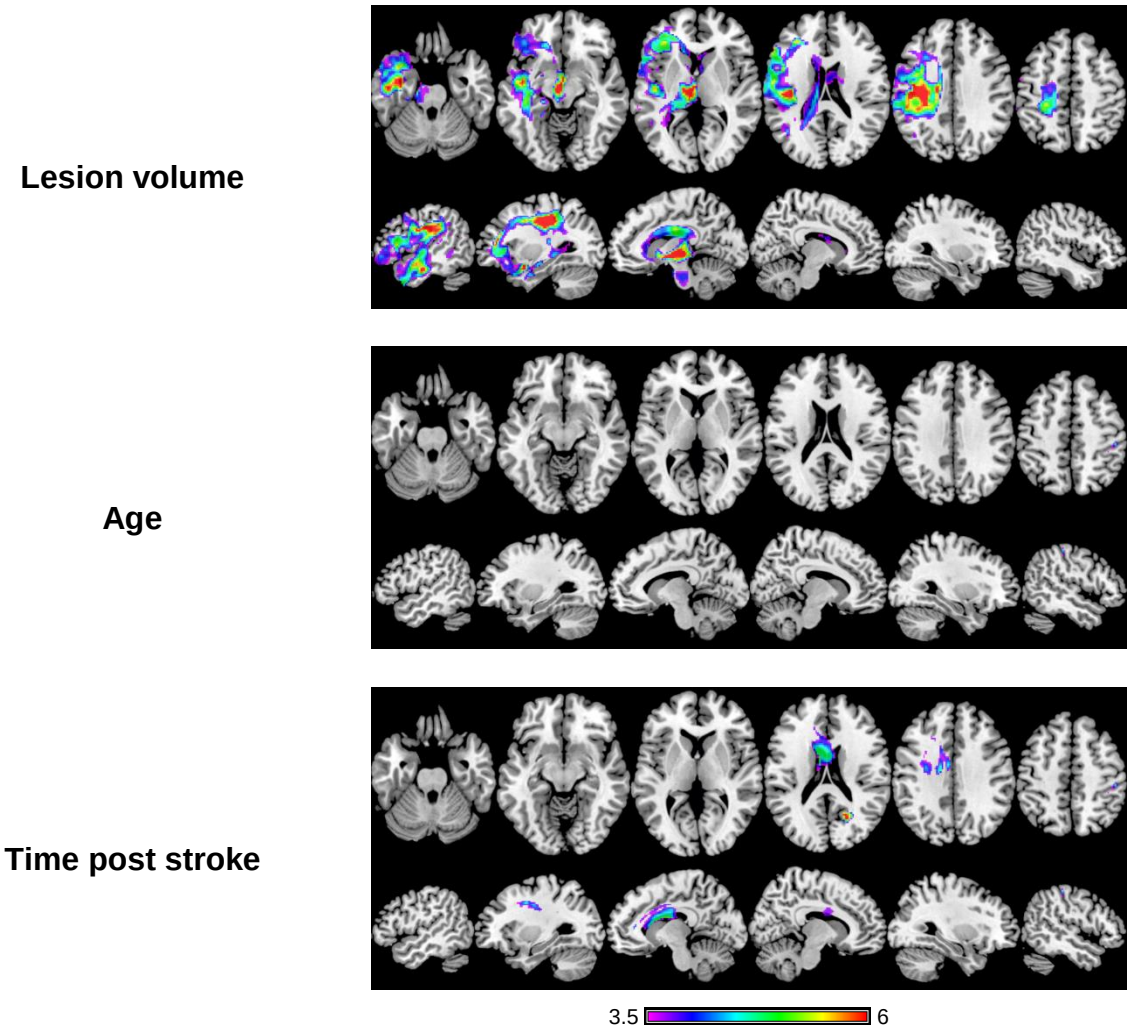

Semantics
Factor score combined PCA
Factor score separate PCA
4 3 2 1 0


Speech Quanta
Factor score combined PCA
Factor score separate PCA
2 1 0

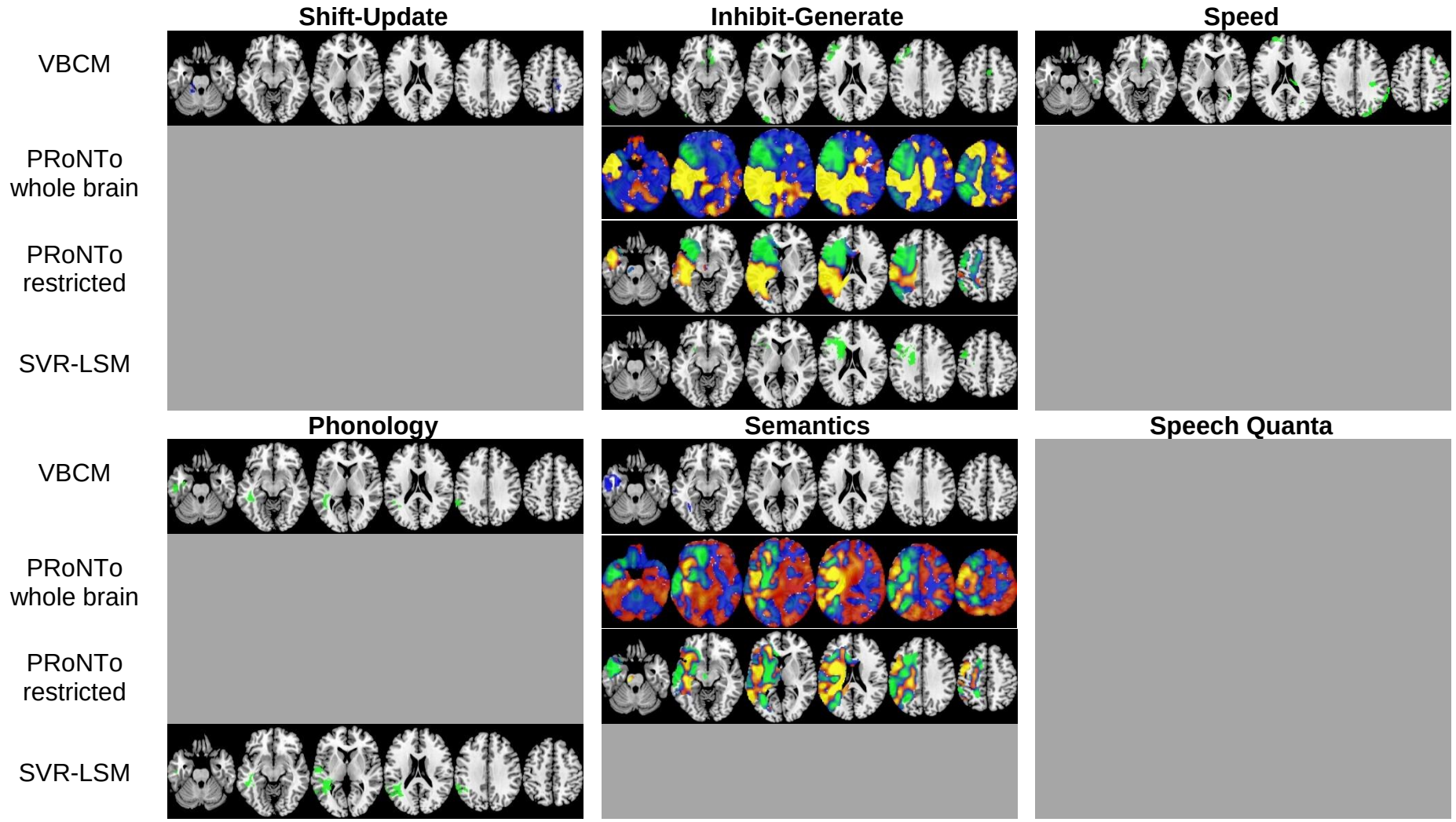
Supplementary Fig. 2. Relationship between and individual's separate and combined component scores and impaired test performance. The colour (legend in the bottom right corner of each graph) indicates the level of impairment (i.e. the individual's number of impaired test scores) in the tests contributing most to each component of the separate analyses (Shift-Update: Trails numbers, Brixton, Divided omissions, Tower of London; Inhibit-Generate: Kramer, Raven B, GoNoGo errors, Design Fluency; Speed: Alertness and GoNoGo median reaction time; Phonology: Boston and Cambridge Naming, Digit Spans, Repetition; Semantics: Synonym judgment, Camel & Cactus, spoken and written Word-Picture matching; Speech Quanta: Token and mean length of utterance of Cookie Theft; see also Fig. 4 in main text). A high correlation between the two component scores (see also Table 1 in main text) and a relatively good fit between component score and impairment can be observed (note that raw scores (sometimes combined) were included in the PCA while impairment in a test was based on norm data, which was corrected for age and education for some tests). Impairment in the Minimal Pairs and in spoken sentence comprehension is not considered as both tests loaded nearly identically on both language components in the separate analyses.


Shift-Update
Inhibit-Generate
Speed
A)
B)
Phonology
Semantics
Speech Quanta
A)
B)
VBCM with lesion volume not significant

Supplementary Fig. 3. Comparison of VBCM clusters with and without covariates. (A) shows the overlap of the VBCM clusters without covariates with the VBCM clusters including lesion as a covariate. (B) shows the comparison with the VBCM clusters including age, education, time post stroke as covariates; The significant VBCM clusters without covariates, presented in the main text, are shown in blue (voxel-level threshold 0.01) or green (voxel-level threshold 0.001), the clusters they are compared to (same voxel-level threshold) are shown in red, their overlap in magenta or yellow, respectively. The cluster-level threshold was set at $p < 0.05$, FWEc. MNI coordinates from left to right are z = -25, -10, 5, 20, 35, 50 for the axial slices and x = -50, -30, -10, 10, 30, 50 for the sagittal slices. Figures are in neurological convention (left is left) and thresholded at the respective minimum t-value.



Supplementary Fig. 4. Cluster associated with lesion volume, age, and time post stroke when only the patient characteristics were included in the VBCM analysis. There was no cluster associated with education at the applied threshold (voxel-level $p < 0.001$ and FWEc at cluster-level $p \leq 0.05$). The same slices as in Supplementary Fig. 3 are shown. Figures are in neurological convention (left is left).


Shift-Update
Inhibit-Generate
Speed
VBCM
PRoNTo whole brain
PRoNTo restricted
SVR-LSM
Phonology
Semantics
Speech Quanta
VBCM
PRoNTo whole brain
PRoNTo restricted
SVR-LSM

Supplementary Fig. 5. Comparison of the univariate and multivariate approaches when lesion volume is included as covariate. The VBCM clusters are shown in blue (voxel-level threshold 0.01) and green (voxel-level threshold 0.001), a family-wise error correction of $p \leq 0.05$ was applied to all clusters, and images are thresholded at the respective minimum t-value. The PRoNTo results depict the weights for the winning model if significant (see text), either including the whole brain space or restricting it to lesion territory ($N > 3$). They are thresholded from -0.005 to -0.0001 (green-blue) and 0.0001 to 0.005 (red-yellow). The negative weights are considered as more meaningful in this approach. The SVR-LSM images show voxels with significant beta weights after permutation testing (N=10,000, voxel-wise $p < 0.005$ and cluster-wise $p < 0.05$). MNI coordinates of slices, from left to right, are z = -25, -10, 5, 20, 35, 50 and they are in neurological convention (left is left). A grey surface indicates that no significant results were found for the respective component and methodological approach.

Effects of including patient characteristics as covariates

When lesion volume was included in the VBCM analysis, the extent of the clusters was generally reduced, as shown in Supplementary Fig. 3. Also, the speech quanta cluster was not significant anymore and the semantics cluster only appeared at a less strict threshold. Supplementary Fig. 4 shows that brain regions on the edge of the overall lesion extent were associated with lesion volume in the VBCM that included the patient characteristics only. Peak regions of this lesion volume cluster overlap with the semantics and speech quanta components, which explains why they emerged only at a less strict threshold or not anymore when lesion volume was taken into account. Similarly, when lesion volume was included in the multivariate approaches, as shown in Supplementary Fig. 5, only the models for inhibit-generate and for phonology remained significant in the SVR-LSM. Interestingly, the SVR-LSM result for inhibit-generate is more similar to the (uncorrected) VBCM cluster if lesion volume is added as a covariate. The PRoNTo approaches yielded significant results for inhibit-generate (KRR model whole brain: cross-validation $r = 0.338$, $MSE = 0.857$, $p = 0.028$; KRR model restricted lesion territory: $r = 0.384$, $MSE = 0.818$, $p = 0.028$) and semantics (KRR model whole brain: $r = 0.603$, $MSE = 0.402$, $p < 0.002$; KRR model restricted lesion territory: $r = 0.597$, $MSE = 0.419$, $p < 0.001$) only. Including lesion volume as a covariate has thus not the same effect in the different brain-behaviour mapping approaches, but generally leads to a reduction of significant models. It remains possible that including lesion volume in brain-behaviour mapping might lead to Type II error; hence it is unclear which strategy might be optimal.

The other patient characteristics (age, education, time post stroke) had a weaker effect on the brain-behaviour mapping. All nonverbal and verbal clusters still emerged at the same thresholds as when analysed by means of the VBCM without the covariates. However, their extent was somewhat smaller but could also include additional areas.

References

Butler RA, Lambon Ralph MA, Woollams AM. Capturing multidimensionality in stroke aphasia: mapping principal behavioural components to neural structures. Brain 2014; 137: 3248-66.

Dronkers NF, Larsen JL. Neuroanatomy of the classical syndromes of aphasia. In: Boller F, Grafman J, editors. Handbook of Neuropsychology. 2nd ed. New York: Elsevier Science; 2001. p. 19-30.

Kasselimis DS, Simos PG, Peppas C, Evdokimidis I, Potagas C. The unbridged gap between clinical diagnosis and contemporary research on aphasia: A short discussion on the validity and clinical utility of taxonomic categories. Brain and Language 2017; 164: 63-7.