

Vocabulary relearning in aphasia is supported by hippocampal memory and cortical language systems

Katherine R. Gore^{1,2}, Anna M. Woollams¹ and Matthew A. Lambon Ralph²

Author affiliations:

1 Division of Neuroscience and Experimental Psychology, School of Biological Sciences, University of Manchester, UK

2 MRC Cognition & Brain Sciences Unit, University of Cambridge, UK

Correspondence to:

Prof. Matthew A. Lambon Ralph
MRC Cognition and Brain Sciences Unit,
15 Chaucer Road,
Cambridge,
CB2 7EF
matt.lambon-ralph@mrc-cbu.cam.ac.uk

Running title: Aphasic neurotypical learning mechanisms

Keywords: aphasia; treatment; fMRI; word-learning;

Abbreviations: AAL = Automated Anatomical Labelling; AFNI = Analysis of Functional NeuroImages; ATL = anterior temporal lobe; BDAE = Boston Diagnostic Aphasia Examination; BOLD = blood oxygen level dependent; BNT = Boston Naming Test; CLS = Complementary Learning Systems; FDR = false discovery rate; FSL = FMRIB Software Library; IFG = inferior frontal gyrus; IPL = inferior parietal lobe; LINDA = Lesion Identification with Neighborhood Data Analysis; SLT = speech and language therapy; MNI = Montreal Neurological Institute; MTL = medial temporal lobe; pSTG = posterior superior temporal gyrus; RISP = repeated increasingly-speeded presentation; RT = reaction time; ROI = region of interest.

Open access: For the purpose of open access, the UKRI-funded authors have applied a Creative Commons Attribution (CC-BY) licence to any Author Accepted Manuscript version arising from this submission.

Abstract

Speech and language therapy can be an effective tool in improving language in post-stroke aphasia. Despite an increasing literature on the efficacy of language therapies, there is a dearth of evidence about the neurocognitive mechanisms that underpin language re-learning, including the mechanisms implicated in neurotypical learning. Neurotypical word acquisition fits within the idea of Complementary Learning Systems, whereby an episodic hippocampal system supports initial rapid and sparse learning, whilst longer-term consolidation and extraction of statistical regularities across items is underpinned by neocortical systems. Therapy may drive these neurotypical learning mechanisms, and efficacy outcome may depend on whether there is available spared tissue across these dual systems to support learning.

Here, for the first time, we utilised a reverse translation approach to explore these learning mechanisms in post-stroke aphasia, spanning a continuum of consolidation success. After three weeks of daily anomia treatment, 16 patients completed a functional magnetic resonance imaging protocol; a picture naming task which probed (i) premorbid vocabulary retained despite aphasia, (ii) newly re-learned treated items and (iii) untreated/unknown and therefore unconsolidated items. The treatment was successful, significantly improving patients' naming accuracy and reaction time post-treatment. Consistent with the Complementary Learning Systems hypothesis, patients' overall naming of treated items, like that of controls when learning new vocabulary, was associated with increased activation of both episodic and language regions. Patients with relatively preserved left hemisphere language regions, aligned with the control data in that hippocampal activity during naming of treated items was associated with lower accuracy and slower responses – demonstrating the shifting division of labour from hippocampally-dependent new learning towards cortical support for the efficiently-named consolidated items. In contrast, patients with greater damage to the left inferior frontal gyrus displayed the opposite pattern (greater hippocampal activity when naming treated items was associated with quicker responses), implying that their therapy-driven learning was still wholly hippocampally reliant.

Introduction

Around one third of people develop aphasia after cerebral stroke.¹ The most common symptom of post-stroke aphasia is word-finding difficulties (anomia) which often persist into the chronic phase of aphasia recovery. Anomia severely reduces patients' communication and, therefore, quality of life.² Thus, successful treatment of anomia is paramount. Although there is extensive evidence that speech and language therapy (SLT) can be effective³⁻⁵, the underlying neural correlates of therapy effectiveness and its variability remain unclear. There have been few, if any, studies that compare SLT-induced changes in neural activation with those associated with neurotypical mechanisms of vocabulary learning. In this reverse translation study, therefore, we utilised post-intervention fMRI to reveal the neural correlates of successful naming therapy, explored how these relate to the neural correlates of vocabulary learning in neurotypical participants, and investigated how lesions to core left hemisphere language regions change the neural division of labour that supports vocabulary relearning.

Previous studies have explored treatment-induced changes in brain activity⁶⁻⁸, and considered how SLT might induce neuroplasticity⁹, the proposed explanations of which have tended to mirror the hypotheses and debates about spontaneous recovery of language function in aphasia. Thus, some studies have proposed that recovery is dependent on the reconstitution of language systems in undamaged tissue around the lesion, resulting in left perilesional activity^{8,10-14}. Other studies have reported a shift of language function to the contralateral hemisphere, with increased activity in right language homologues.¹⁵⁻¹⁷ Accordingly, there is extensive debate about which patterns of reorganisation are advantageous to language outcome in stroke: the contribution of right hemisphere language homologues to speech production has been considered both beneficial^{15,16,18} or detrimental^{19,20} to recovery. Recent neurocomputational models suggest that these apparently contradictory patterns actually reflect the process of recovery after different levels of damage within an asymmetric yet inherently bilateral language production system.²¹ There is also some evidence that, beyond language-specific brain regions, there can be treatment-induced functional changes in domain-general brain networks²², including networks underpinning short-term memory.^{23,24}

An important next step in the field would be to secure a neurocognitive theory of SLT-induced language improvements, ideally underwritten by formal computational models. A first stride in this direction might be derived from the rich literature on learning and memory consolidation in healthy individuals. Explorations of the neural processes associated with

word training in healthy participants align with the Complementary Learning Systems theory (CLS).²⁵⁻²⁸ The CLS model proposes that sparse and highly plastic representations within the medial temporal lobes (MTL) and other parts of the episodic network support initial rapid knowledge acquisition. Slower learning across distributed representations within the neocortex allows for extraction of statistical structures and generalisation across episodes. Over time, knowledge consolidation is driven by a gradual MTL-to-neocortical shift in the division of labour across these complementary systems.²⁵ Previous fMRI investigations have shown that initial exposure to novel, abstract vocabulary items engages the hippocampus which gradually reduces over repeated presentations.²⁹⁻³² A more recent study in healthy participants, which adopted the typical multi-session cueing-based learning/therapy approach used for aphasic patients in SLT, investigated the neural division of labour for consolidated learning of previously-unknown native vocabulary items.²⁸ Consistent with the CLS theory, the newly-consolidated vocabulary was reliant upon both episodic regions (including MTL) and neocortical language regions; by contrast, pre-existing vocabulary only engaged these neocortical language regions. In line with the shifting MTL-to-neocortical division of labour, participants' language efficiency with these items was positively correlated with activation in the neocortical regions and negatively related to the hippocampal activation.

In this reverse translational study, we investigated the underlying neural correlates of successful speech and language therapy. By following the design of the recent exploration of native vocabulary learning in healthy participants (see above), we addressed two central questions: (1) does word re-learning in post-stroke aphasia fit within the CLS framework that supports word learning in healthy older adults?; and (2) does a certain level of critical damage to the core IFG speech production area cause deviations away from this typical learning framework? To explore these questions, we gave patients three weeks of speech and language therapy on items they could not consistently name in prior naming assessments. Directly after treatment, patients completed a picture naming task in the fMRI scanner in which we compared the activation associated with naming of treated items, untreated items they could name prior to scanning, and untreated/unknown items.

By mirroring the procedure of the healthy control study²⁸, we were able to compare the patients' fMRI results for name relearning against the combined language and episodic networks that are observed for vocabulary learning in healthy participants. In advance of the study, we anticipated three possible outcomes: (i) if enough of the patient's typical language systems were intact, then the naming of these relearned items should mirror the healthy control

pattern; (ii) if representations of the item's name exist but were weakened following stroke, re-learning could strengthen the representations and processing in the typical (retained) language system. This 'representational repair' process may be supported by 'new learning' via the episodic system, similar to when controls learn new vocabulary²⁸; (iii) if language areas were more compromised, then the remaining representations may be more severely damaged and thus (a) compensatory language-related systems may be engaged, or (b) if the neocortical representations are critically compromised then episodic systems would be dominant, somewhat akin to the pattern observed in healthy individuals during the initial phase of novel vocabulary learning²⁹⁻³² where there are yet to be any cortical representations.

Materials and methods

Participants

Sixteen participants were recruited from the North West of England (seven female, age range = 39-73, mean (M) = 57.5) with acquired language production impairment following a single left hemispheric stroke (either ischaemic or haemorrhagic) at least one year prior to the study (see Figure 1 for lesion overlap and Table 1 for demographic and selected neuropsychological data). Participant numbers were determined based on access. Inclusion criteria were: (i) anomia as determined by the Boston Naming Test (BNT; cut off 90%)³³; (ii) good spoken comprehension determined by the Comprehensive Aphasia Test (CAT)³⁴; (iii) good repetition of single words and non-words determined by the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA)³⁵ and (iv) absence of apraxia of speech. Participant's aphasia was classified using the Boston Diagnostic Aphasia Examination (BDAE).³⁶ All were native English speakers, with normal or normal-to-corrected vision and hearing, with no other significant neurological conditions or contraindications to MRI scanning. Participants gave informed consent prior to participation according to the Declaration of Helsinki, with approved ethics by the local Health Research Authority ethics committee. Structural MRI data from an age and education matched healthy control group ($n = 20$, 12 females, age range = 46-77, $M = 63.90$) were used for lesion identification.³⁷

Background neuropsychological assessments

All participants were tested on an extensive neuropsychological/aphasiological battery assessing language and cognitive abilities.³⁸ In addition to the BDAE³⁶, this battery consisted of (i) naming tests, including the Boston Naming Test (BNT³³) and the 64-item Cambridge naming test; (ii) the spoken sentence comprehension task from the CAT³⁴; (iii) semantic tasks from the 64-item Cambridge Semantic Battery³⁹ and the 96-trial synonym judgement test⁴⁰; (iv) speech production tests from the PALPA³⁵; (v) cognitive tests including the forward and backward digit span,⁴¹ Raven’s Coloured Progressive Matrices⁴² and the Brixton Spatial Rule Anticipation Task⁴³. Patient’s demographic details are displayed in Table 1, alongside results of the BNT to demonstrate that there was a word-finding deficit. An overlap map of participants’ left hemispheric lesions is shown in Fig. 3C.

Table 1. Patient’s demographic information ordered by lesion volume.

Patient ID	Age (years)	Gender	Education (years)	BDAE classification	Time post-stroke (years)	Lesion volume (voxels)	BNT
P1	56-60	F	16	Anomia	6	175	61.67
P2	71-75	M	11	Anomia	6	3136	76.67
P3	61-65	M	19	Anomia	4	4234	50.00
P4	51-55	F	11	Anomia	5	6630	78.33
P5	56-60	M	16	Anomia	5	6974	55.00
P6	46-50	M	11	Anomia	6	8163	43.33
P7	51-55	F	19	Anomia	6	8810	88.33
P8	66-70	F	12	Anomia	9	9383	75
P9	56-60	F	11	Anomia	3	10027	71.67
P10	61-65	M	17	Broca	10	11179	76.67
P11	56-60	F	11	Anomia	5	12767	50.00
P12	51-55	F	11	Anomia	6	13080	78.33
P13	51-55	M	11	Anomia	8	14681	85
P14	51-55	M	11	Broca	3	17465	35
P15	46-50	M	11	Broca	9	20681	41.67
P16	56-60	M	13	Anomia	6	22298	38.33

Note: BDAE = Boston Diagnostic Aphasia Examination; BNT = Boston Naming Test.

Pre-treatment naming assessments and stimuli

The design mirrored that used in Gore et al.²⁸ with healthy controls. Patients were assessed on all stimuli items three times prior to training. There were three sets of items: already known, to-be-treated and untreated. 200 items were taken from the International Picture Naming Project⁴⁴ and matched with discernible, coloured images with a white background to build the already known and to-be-treated item sets. Items were selected with high word frequency, short reaction times (<1000ms) and high accuracy (85-100%) in the normative IPNP data. Items that were correctly named by a patient on all occasions were added to that individual's already known set. Items that could not be named on one or more occasions were randomly divided between the to-be-treated or untreated sets. Final sets comprised 50 items each and were matched across sets for their psycholinguistic properties of word frequency, number of syllables and word length. Participants who could name over 80% of all items before training were provided with a more challenging set of to-be-treated items (3 participants in total). This set contained unfamiliar, rare items with low word frequency names, drawn from the British National Corpus.⁴⁵ Images from the already known, treated and untreated sets were phase scrambled for use as stimuli in the baseline task.

Anomia treatment

Patients received a laptop to practice naming the to-be-treated items for 45 minutes a day, four days a week, for three weeks. Patients completed, on average, a total of 7.3 hours of naming treatment. During the first two weeks, participants received cue-based training. In the third week, participants received speeded training.

During cue training, the number of items was incrementally increased, beginning with 10 items. When 70% accuracy was reached on these 10 items, another 10 items were added incrementally up to the total 50 items. Participants first viewed each item picture and item name in orthographic and audio form. The patients repeated the name out loud. The cue training was similar to standard error-reducing speech and language therapy^{46,47}, but with a choice of cue to provide a self-determined approach. An item picture was shown, and patients had an option to choose cues or to proceed with naming. For example, for a picture of a butterfly, the cue options were (i) written semantic description "A flying insect with large colourful wings"; (ii) an initial phonemic cue 'bu'; (iii) initial and second phonemic cue

‘butter’ and (iv) the whole word cue ‘butterfly’. All cues were delivered orthographically and audibly by the laptop. After each naming attempt, the whole correct word was given. Patients then answered a semantic question “Is it living?”.

The third week of training adopted the repeated increasingly-speeded presentation (RISP⁴⁸) method. To emphasise speed as well as naming accuracy, patients were instructed that they needed to name the item before the computer did. When participants reached a speeded naming success rate of 70%, the timing was incrementally decreased from 3s to 1.4s, to 1s. When participants beat the 1s target for 70% of items, the set size was increased by 10 items and the time-to-beat returned to 3s.

Procedure

Patients were asked to name the 200 IPNP pictures without cues before and after the 3-week anomia treatment. Within two days of finishing treatment, patients performed a picture naming task in the scanner. Patients also performed a semantic task, which is the focus of another study. The order of pictures and the order of task was counterbalanced within and across patients.

Neuroimaging acquisition

High-resolution structural T1-weighted MRI scans were acquired on a 3T Philips Achieva scanner using an eight-element SENSE head coil. The parameters were as follows: repetition time = 9.0ms, echo time = 3.93ms, acquired voxel size = 1.0 x 1.0 x 1.0 mm squared, matrix size = 256 x 256, slice thickness = 1mm, flip angle = 8 degrees, FOV = 256mm, inversion time = 1150ms, with 150 contiguous slices and a SENSE acceleration factor of 2.5.

A triple gradient echo EPI sequence was used for functional imaging to improve signal-to-noise ratio in the anterior temporal lobes (ATL).^{49,50} In single echo imaging, there is signal dropout and distortion in the ATL.^{51–53} The functional scans were also acquired with a 45-degree tilt from AC-PC to reduce ghosting artefacts in the temporal lobes. The functional sequences consisted of 31 slices covering the whole brain, TR = 2.5s, TE = 12, 30 and 48ms, flip angle = 85 degrees, resolution matrix = 80 x 80, FOV = 240 x 240mm and voxel size = 3.75 x 3.75 x 4mm.

Stimuli were presented during scanning using E-Prime 2.0, with block order pseudo-randomised and optimised for statistic power using OptSeq

(<http://surfer.nmr.mgh.harvard.edu/optseq/>). Spoken in-scanner responses were recorded using a fibre optic, noise-cancelling microphone for fMRI (FOMRI; Optoacoustics). Patients practised speaking with as little movement as possible prior to scanning to reduce motion artefacts.

There were two tasks during imaging acquisition, one of which is the focus of a separate study. For this study, a blocked-design picture naming task with four conditions was used; already known, treated, untreated and baseline. In the known condition, patients named pictures they could consistently name prior to treatment (e.g., pencil). In the treated condition, patients named newly treated items (e.g., penguin). If patients could not remember an item name, or the item was novel or phase scrambled they responded: “Don’t know”.

Each trial lasted 3700ms. A fixation cross was presented for 700ms, followed by the image to be named for 3000ms. Each 11.1s block consisted of three trials. Eight jittered length rest blocks were included per run, with an average length of 11.1s. There were two 7 minute and 4 second long blocks per specific task.

Neuroimaging preprocessing and analysis

T1-weighted structural images were pre-processed in FSL, version 6.0.0,⁵⁴ using the `fsl_anat` pipeline (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/fsl_anat), excluding FAST tissue segmentation. A supervised lesion segmentation algorithm (LINDA) was used for lesion identification.⁵⁵ T1-weighted images were normalised to Montreal Neuroimage Institute (MNI) space and LINDA transforms were applied. The functional data were despiked and slice time corrected per echo in the AFNI neuroimaging suite (v19.2.10; 3dDespike; 3dTshift).^{56,57} Motion correction parameters were calculated from the first slice of the first echo, then this transformation was applied to all three echoes. A subject-specific, LINDA-generated T1 mask was transformed to functional space and applied to each echo. Multi-echo data were optimally combined using the T2* combination method⁵⁸ in `tedana`.^{49,59} Optimally combined data were normalised to MNI space using LINDA transforms. Smoothing was performed on this data using FSL, using 8mm full-width half-maximum equivalent sigma. Statistical whole brain analyses were performed non-parametrically using `SnPM13` (<http://warwick.ac.uk/snpm>). Implicit masking defaults were set to ‘-Inf’ to remove the arbitrary analysis threshold and avoid lesion voxels being masked out for all participants. Region of interest analyses were performed using `SPM12`⁶⁰ and `MarsBaR`⁶¹. Two-step

multiple regression and simple slope analyses were performed in R version 4.0.5⁶² using the emmeans package.

Regions of interest were defined based upon the previous literature. Five *a priori* regions of interest were defined to include key semantic and episodic network areas. The left inferior frontal gyrus (IFG; MNI: -46 28 10), a key area of damage in post-stroke aphasia is considered critical in speech production.^{63–65} The right IFG homologue (MNI: 46 28 10) was included to investigate right-sided homologue compensation theories.^{66–68} A left ventral anterior temporal lobe (vATL; MNI: -36 -15 -30) was taken from a key semantic cognition reference.⁶⁹ For the episodic memory network, a left hippocampal ROI (MNI: -28 -14 -15) was defined based on previous literature of initial hippocampal activation during word learning.^{29,30} In addition, an ROI in the left inferior parietal lobe was included (IPL; MNI: -47 -64 34) due to consistent activation of this region in episodic processing.^{70,71} Pearson's *r* correlations were used to explore associations between ROI BOLD activity and behavioural performance measures (accuracy and RT). Critical tests of significant differences between correlations were adjusted for multiple comparisons using the Benjamini-Hochberg procedure, FDR = 0.05.⁷²

To explore the interaction between critical language region damage and performance/functional activity, a further left IFG ROI was created. This ROI included left IFG pars opercularis and left IFG pars triangularis, as designated by the Automated Anatomical Labelling atlas.⁷³

Behavioural performance analysis

Individual behavioural performance was probed for differences between the three pre-training test scores using a McNemar's test. Pre-treatment scores were averaged per participant as baseline performance was stable across the group and individuals. Repeated-measures *t*-tests were performed for each stimuli type, between the time points of pre- and post-treatment for accuracy and RT. Two repeated-measures analyses of variance (ANOVA) were conducted, one for accuracy performance and one for RT. These ANOVA analyses included within-subject main factors of Time (pre-treatment, post-treatment) and Treatment (treated, untreated and already known). Two-tailed *p* values and an alpha level of 0.05 were applied for all statistical tests. Due to the global pandemic, maintenance data was not available.

Results

Behavioural performance

There was no significant difference in the proportion of accurate responses between the three baseline pre-tests across the group ($p = .88$). Pre-treatment performance showed no significant change across time points. Thus, baseline pre-tests were averaged to provide a single representative sum of pre-treatment baseline performance. There was no significant difference in accuracy ($t = 1.67, p = .12$) or RT ($t = -1.76, p = .10$) for untreated items between time points (Fig. 2A). The patients demonstrated a substantial and significant improvement in naming accuracy ($t = 5.18, p < .0001$) and improved naming speeds ($t = -7.63, p < .0001$) when naming treated items after treatment (Fig. 2B). The patients also showed a small yet significant increase in naming accuracy for the known items ($t = 3.31, p = .004$), but the small change in naming speed failed to reach significance ($t = -2.08, p = .06$; Fig. 2C).

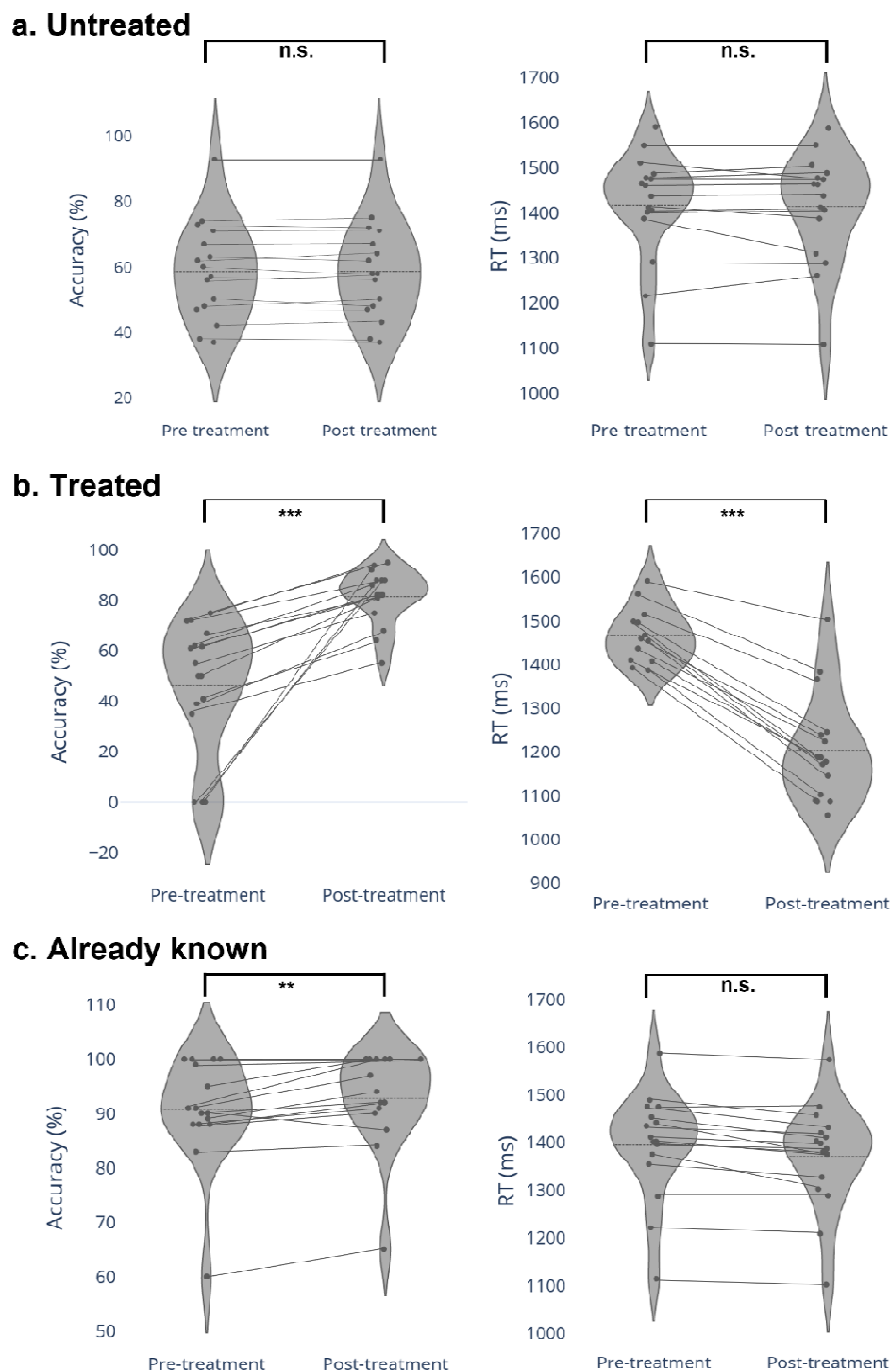


Figure 2. Accuracy and RT of naming a) untreated items, b) treated items and c) already known items. Dotted line represents the mean. Solid lines between points represent individual patient differences in time points. Group differences significant at $p < .0001$ and $p < .005$ are denoted with ‘***’ and ‘**’ respectively.

Accuracy

A repeated-measures 2 x 3 ANOVA revealed a significant main effect of Time ($F(1,15) = 90.35, p < .001$), with greater accuracy post-treatment. There was also a significant effect of Treatment ($F(2,30) = 118.31, p < .0001$) with greater accuracy when naming newly treated than untreated or known items. There was a significant Time x Treatment interaction ($F(2,30) = 35.48, p < .0001$). Naming of treated items was more accurate post-treatment (81%) than pre-treatment (46%). This unstandardised effect size (35%) was significantly greater than the mean difference (1%) between naming of untreated items pre-treatment (58%) and post-treatment (59%; $p < .0001$). The unstandardised effect size was also significantly greater than the mean difference (2%) between naming of already known items pre-treatment (90%) and post-treatment (92%; $p < .0001$).

RT

There were also significant main effects of both Time (quicker RTs post-treatment; $F(1,15) = 24.084, p < .001$) and Treatment (faster RTs when naming newly treated than untreated or known items; $F(2,30) = 27.307, p < .0001$). There was also a significant Time x Treatment interaction for RT ($F(2,30) = 17.507, p < .001$). Naming of newly treated items was faster (1203ms) than pre-treatment naming (1467ms). This unstandardised effect size (264ms) was significantly greater than the mean difference (4ms) between untreated items pre-treatment (1417ms) and post-treatment (1413ms; $p = .002$). Additionally, the unstandardised effect size was also significantly greater than the mean difference (23ms) between already known items pre-treatment (1394ms) and post-treatment (1371ms; $p = .002$).

Whole brain fMRI contrasts

The results of whole brain analyses are reported in Table 2. Multiple contrasts were performed; treated > already known, treated > untreated, already known > untreated and the inverse of these contrasts. There were significant clusters for treated > untreated and already known > untreated (Fig. 3). In the treated > untreated contrast, there was a significant cluster of activation over the right superior temporal gyrus (STG), right rolandic operculum and the right postcentral gyrus. There was also a further cluster in the dorsal occipital cortex spanning bilateral lingual gyri and the cerebellum. In the already known > untreated contrast, there was a similar pattern of activation. The right-sided cluster spanned the right anterior parietal and

superior temporal cortex areas. The posterior cluster centred on the bilateral lingual gyri, including posterior left parahippocampal cortex.

Table 2. Significant clusters of activity during naming of treated, already known and untreated items.

Contrast	Region of activation	Peak region	Cluster size	Peak MNI			Pseudo <i>t</i>
				x	y	z	
Treated > untreated	R STG, postcentral gyrus	R STG	3947	65	-14	11	5.26
		R postcentral		54	-6	13	5.05
		R STG		64	-6	4	3.86
	Occipital cortex, cerebellum	R cerebellum	24708	17	-59	-23	5.15
		L calcarine		2	-86	-5	4.84
		L cerebellum		-13	-62	-13	4.33
Known > untreated	R SMG, R STG, R Heschl gyrus	R SMG	2855	61	-35	29	4.84
		R STG		65	-16	12	4.35
		R STG		64	-7	7	3.76
	Bilateral lingual gyrus, left MTL	L fusiform	9039	-31	-39	-11	4.71
		R lingual		5	-79	-7	4.65
		L calcarine		2	-85	0	4.39

Note: Clusters significant at $p < .001$ voxel height and $p < .05$ FWE cluster correction. Up to 3 strongest peaks listed per cluster. Pseudo-*t* statistics were computed with smoothed variance in SnPM. R, right; STG, superior temporal gyrus; SMG, supramarginal gyrus; MTL, medial temporal lobe.

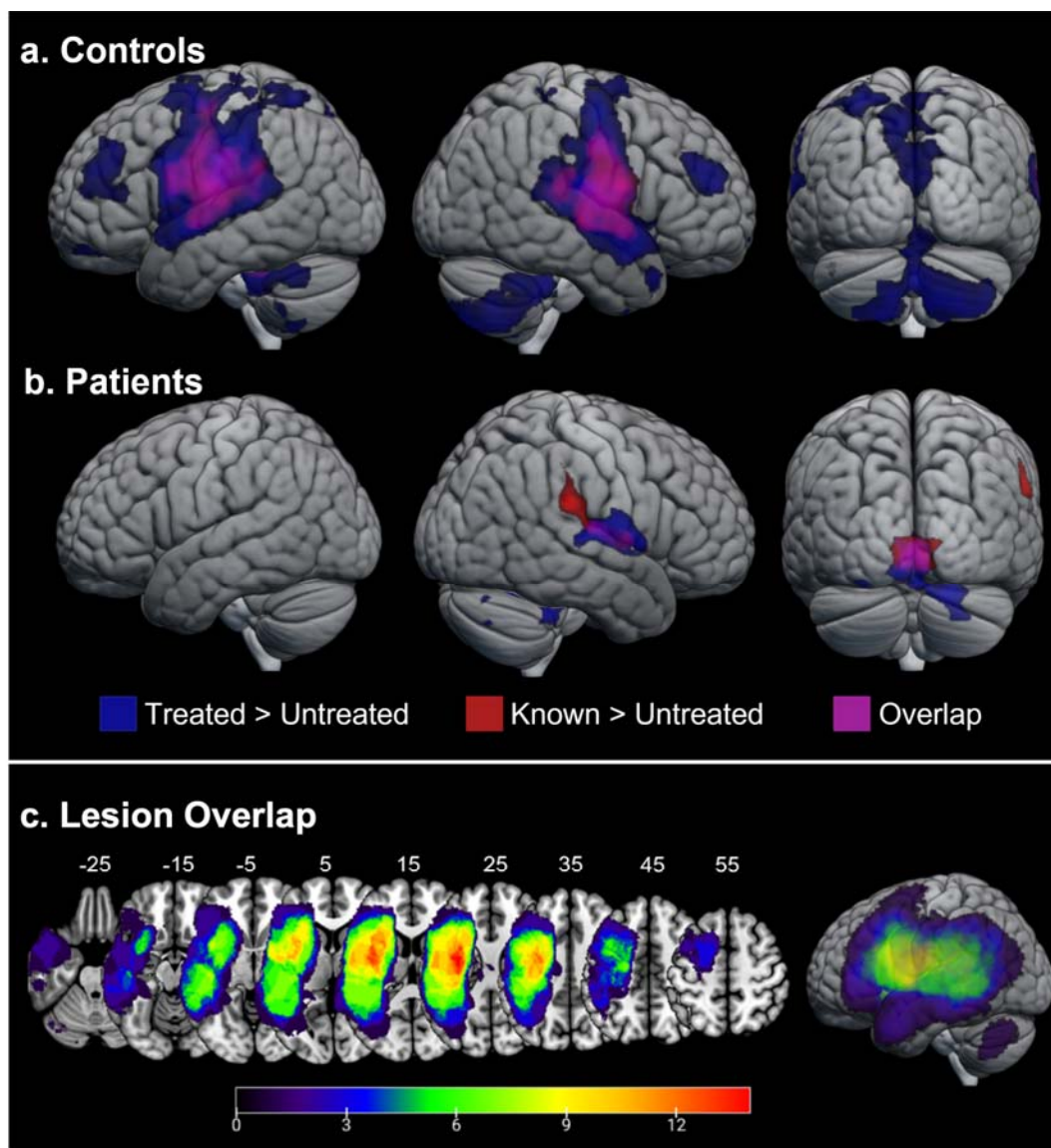


Figure 3. Whole brain clusters of increased activation during picture naming for (A) controls and (B) patients. Patient lesion overlap map on axial slices (C). Picture naming of treated > untreated (blue) and known > untreated (red) contrasts, with overlap (purple). Image thresholded at voxel level $p < .001$, cluster corrected FWE $p < .05$. R; right, P; posterior. (C) Colour bar indicates the number of participants included in the overlap. Numbering above slices indicates z MNI coordinate. In neurological convention, left is left.

Region of interest analyses

To explore the core hypothesis that neural processes underlying word re-learning in aphasia would be similar to those that support word learning in healthy controls (e.g., CLS theory) five *a priori* regions of interest (ROIs) were defined (Fig. 4A). These included language-semantic network areas (left IFG, right IFG and left ATL) and episodic network areas (left hippocampus and left IPL). Percentage damage to 8mm spherical ROIs is listed in Supplementary Table 1. An initial exploratory analysis was performed correlating BOLD activity of these ROIs with naming reaction times (RT) and accuracy. There were no significant correlations between behaviour and ROI activity in the already known > untreated contrast.

In Gore et al.²⁸ the underlying neural correlates of word learning fitted within the CLS model. To investigate whether speech and language therapy also drives these neurotypical learning mechanisms, we explored whether treated > untreated behavioural-activation correlations differed significantly from already known > untreated items. In healthy participants, left hippocampal activity was associated with worse performance, whereas left IFG and left ATL activity was associated with better performance.²⁸ For the patients in this study, there was a strong negative correlation between accuracy and left hippocampal activity in the treated > untreated condition ($r = -.517$, $p = .039$, Fig. 4B), which was significantly different to the weak *positive* correlation of accuracy and left hippocampal activity in the already known > untreated condition ($r = .363$, $p = .166$) using Fisher's r-to-z transformation ($z = -2.43$, $p = .008$, adjusted $p = .024$). There was a positive correlation between naming accuracy and left IFG activity in the treated > untreated condition ($r = .540$, $p = .043$, Fig. 4B), which was significantly different to the weak negative correlation of naming accuracy and left IFG activity in the already known > untreated condition ($r = -.025$, $p = .929$) using r-to-z transformation ($z = 2.19$, $p = .014$, adjusted $p = .021$). There were no further significant correlations between accuracy and ROI BOLD activity, including in the right IFG $r = .041$, $p = .881$.

Associations between region of interest BOLD activity and naming RTs were also explored. There was a strong positive correlation between longer RTs and left hippocampal activity in the treated > untreated condition ($r = .576$, $p = .020$, Fig. 4C), which was significantly different to the weak positive correlation in the already known > untreated condition ($r = -0.011$, $p = .971$; $z = -1.65$, $p = 0.04$, adjusted $p = .04$). There were no further significant

correlations between RT and ROI BOLD activity, including in the right IFG ($r = .229$, $p = .260$).

To examine the second study question, hierarchical multiple regression was used to investigate whether neural changes may deviate from those found in normal word learning paradigms dependent on levels of damage to critical regions. Given its importance in language processing and speech production, we focussed on the level of damage to the left IFG. All covariates were mean centred, and the interaction term was generated. In the first step, variables of left hippocampal activity and percentage left IFG damage were entered as predictors, accounting for a non-significant amount of variance in RTs ($R^2 = .340$, $F(2, 13) = 3.345$, $p = .067$). In the second stage, the interaction term between left hippocampal activity and left IFG damage was entered accounting for a significant proportion of variance ($R^2 = .765$, $F(1, 12) = 5.655$, $p = .012$, $b = -.508$, $t = -2.669$, $p = .02$). For visualisation, predicted simple slopes were generated to display the interaction at three levels of left IFG damage, mean and one standard deviation above/below the mean (18%, 58% and 98%; Fig. 4C). Greater hippocampal activity was associated with slower responses for those with less left IFG damage, and faster responses for those with more left IFG damage.

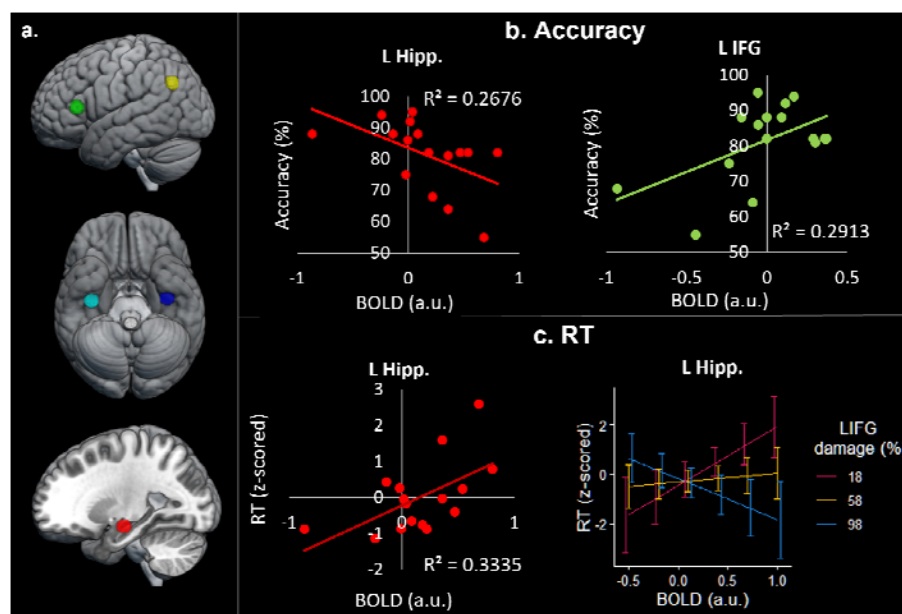


Figure 4. Correlations between regional activation and behavioural performance. (A) Regions of interest: green = left inferior frontal gyrus (IFG); yellow = left inferior parietal lobe; cyan = left anterior temporal lobe (ATL); blue = right ATL; red = left hippocampus (Hipp.). **(B)** Correlations between treated > untreated BOLD activity and accuracy. **(C)** Left:

correlation between treated > untreated left hippocampal BOLD activity and RT. Right: Predicted simple slopes of the association between treated > untreated left hippocampal BOLD activity and RTs at three levels of left IFG damage.

DISCUSSION

Whilst there is an increasing body of research on the positive efficacy of speech and language therapy for aphasia,^{3,4,74–76} there is still limited knowledge about the neurocognitive bases of both successful therapy and its variability across patients. In this reverse translation study, therefore, we examined a commonly adopted cueing-based naming treatment and used post-intervention fMRI to explore two core questions: (1) does word re-learning in post-stroke aphasia engage the Complementary Learning Systems (CLS)²⁷ that support word learning in healthy older adults?; and (2) does the level of damage to the core IFG speech production area cause deviations away from this typical learning framework?

The study yielded clear answers to both these questions. The overall result for the patients mirrored that observed in a previous post-learning fMRI study of healthy participants who had been asked to learn unfamiliar native vocabulary²⁸. Specifically, following a three-week combined cueing-based and speeded naming therapy, the patients' naming of the relearned items engaged a combination of the cortical language regions associated with production of established vocabulary plus the hippocampal episodic system. In line with the CLS framework, the gradual shift in the division of labour across these systems aligned with the consolidation status of the items (reflected in the speed and accuracy of naming). Thus, increased hippocampal activity during the naming of treated words was associated with worse performance (i.e., slower RTs and lower accuracy) whilst, in contrast, increased activity in a critical language production area, the left IFG, was associated with quicker RTs. Taken together, these counterpointed activation profiles are consistent with the CLS proposal that new learning, or in this case relearning, is initially dependent on the MTL episodic system and then gradually shifts towards long-term consolidation within the cortical systems. The results also imply that (a) the process of correcting or cleaning-up the partially-damaged cortical representations of impaired premorbid vocabulary requires a re-engagement of the MTL episodic system, and (b) long-term vocabulary reconsolidation ultimately involves re-engagement of the same cortical language production systems as those for preserved vocabulary.

In relation to the second core study question, we found that there was a crucial deviation away from this typical CLS pattern of vocabulary recovery which depended on the level of damage to the left IFG. Thus, when there was minimal IFG damage, the patients exhibited the counterpointed IFG-hippocampal correlations with naming efficiency described above and found in healthy participants when learning new vocabulary. Yet in the context of IFG damage, patients' naming of relearned items appears to remain reliant upon the hippocampal episodic system, such that better naming performance was then associated with increased hippocampal activation. There are at least two possible explanations for this altered pattern of results when there is more damage to the cortical language systems. The first is that there is insufficient cortical tissue left to support the longer-term consolidation of the relearned words and thus the hippocampal system has to retain sole "responsibility" for relearning. The second is that, with damaged cortical systems, more restoration work is required and this correction of the impoverished representations might also take much longer than before. Accordingly, the hippocampal contribution to relearning is both greater and may be required for much longer.

These results imply that it may be important to consider the level of damage to critical language regions during stratification of patients to therapy approach. Patients with more damage to critical language regions who are likely to be, therefore, more hippocampal dependent, may show worse poorer retention over time. Indeed, the previous investigation of new vocabulary learning in healthy participants found that higher levels of hippocampal engagement after learning was associated with poorer retention at six months.²⁸ To test this hypothesis in patients, further research is needed with post-treatment testing to determine maintenance of re-learned items. Furthermore, if hippocampally-dependent participants either have less efficient learning or are unable to shift the division of labour to the damaged cortical language systems, a longer-term treatment and maintenance strategy may be crucial to treatment outcomes.

Going beyond the focus on the language dominant area and hippocampal areas, the whole brain results demonstrated an overlapping cluster of increased right posterior superior temporal gyrus (pSTG) activity for both the contrasts of already known > untreated and treated > untreated. Over half of the patients had lesion overlap within the left pSTG (9; Fig. 3C). The right pSTG is consistently activated in controls and patients with aphasia during speech production tasks.⁷⁷ Indeed, there was right pSTG activation for controls performing the same task as that used in this study.²⁸ Although typically associated with receptive

auditory-language process rather than speech output, the posterior STG and neighbouring regions have been argued to be a critical part of the internal feedback systems that help to monitor and thus control the accuracy of speech production.⁷⁸

Finally, we should note that we observed very small yet reliable performance gains in the already known and untreated item sets (though considerably smaller than on the treated vocabulary). There are important considerations regarding the source of these gains in this study. Patients were tested on all items a total of five times during the course of the study. Previous studies have shown that such re-assessment alone may be sufficient to induce small yet reliable performance increases.⁷⁹ Other studies have suggested that generalisation to untreated items via cueing therapy may be specific to the minority of patients with relatively intact semantic processing and deficits only in the phonological encoding stages of speech production.⁸⁰ Consistent with this alternative hypothesis, the patients in this study were generally mildly aphasic and, except for one patient, all had only minor semantic deficits.

Conclusions

The results of this study demonstrated that word re-learning post-stroke engages the same CLS dual-system framework that is found for novel vocabulary learning in healthy participants, and shows a gradual shift in the division of representational labour from the hippocampal episodic system to the cortical systems required for speech production (e.g., IFG). However, this neurotypical pattern was moderated considerably by the degree of damage to key left hemispheric language regions. After critical levels of damage to the left IFG, better performance requires a continued and increasing engagement of the hippocampal system.

Funding

Medical Research Council National Productivity Investment Fund PhD studentship; European Research Council (GAP: 670428 - BRAIN2MIND_NEUROCOMP); Medical Research Council programme grant (MR/R023883/1) and intra-mural funding (MC_UU_00005/18).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary Table 1. Percentage damage to the regions of interest (ROI) per participant. All ROIs refer to an 8mm spherical region, except AAL L IFG which refers to the whole left IFG pars opercularis and pars triangularis.

ROI	Patient ID															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
L IFG	84	-	-	-	100	6.8	-	-	-	61	82	-	22	-	-	55
R IFG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L ATL	-	-	12	-	-	-	-	-	-	-	-	-	-	-	-	-
L Hipp	-	42	13	-	-	-	-	2	-	-	-	-	6	-	-	1
L IPL	87	-	-	-	-	-	79	-	-	-	-	62	34	-	-	-
AAL L IFG	95	89	10	37	94	38	28	89	-	88	99	97	95	-	4	47

Note: “-“ indicates no lesion overlap with the ROI. L, left; R, right; IFG, left inferior frontal gyrus; ATL, anterior temporal lobe; Hipp, hippocampus; IPL, inferior parietal lobe; AAL, Automated Anatomical Labelling.

References

1. Engelter ST, Gostynski M, Papa S, et al. Epidemiology of aphasia attributable to first ischemic stroke: Incidence, severity, fluency, etiology, and thrombolysis. *Stroke*. 2006;37(6):1379-1384. doi:10.1161/01.STR.0000221815.64093.8c
2. Code C. The quantity of life for people with chronic aphasia. *Neuropsychol Rehabil*. 2003;13(3):379-390. doi:10.1080/09602010244000255
3. Crinion JT, Leff AP. Recovery and treatment of aphasia after stroke: functional imaging studies. *Curr Opin Neurol*. 2007;20(6):667-673. doi:10.1097/WCO.0b013e3282f1c6fa
4. Fillingham JK, Sage K, Lambon Ralph MA. The treatment of anomia using errorless learning. *Neuropsychol Rehabil*. 2006;16(2):129-154. doi:10.1080/09602010443000254
5. Crinion JT, Leff AP. Using functional imaging to understand therapeutic effects in poststroke aphasia. *Curr Opin Neurol*. 2015;28(4):330-337. doi:10.1097/WCO.0000000000000217
6. Nardo D, Holland R, Leff AP, Price CJ, Crinion JT. Less is more: neural mechanisms underlying anomia treatment in chronic aphasic patients. 2017;(October). doi:10.1093/brain/awx234
7. Abel S, Weiller C, Huber W, Willmes K. Neural underpinnings for model-oriented therapy of aphasic word production. *Neuropsychologia*. 2014;57(1):154-165. doi:10.1016/j.neuropsychologia.2014.03.010
8. Abel S, Weiller C, Huber W, Willmes K, Specht K. Therapy-induced brain reorganization patterns in aphasia. *Brain*. 2015;138(4):1097-1112. doi:10.1093/brain/awv022
9. Berthier ML, Pulvermüller F. Neuroscience insights improve neurorehabilitation of poststroke aphasia. *Nat Rev Neurol*. 2011;7(2):86-97. doi:10.1038/nrneuro.2010.201
10. Meinzer M, Breitenstein C. Functional imaging studies of treatment-induced recovery in chronic aphasia. *Aphasiology*. 2008;22(12):1251-1268. doi:10.1080/02687030802367998
11. Saur D, Lange R, Baumgaertner A, et al. Dynamics of language reorganization after stroke. *Brain*. 2006;129(6):1371-1384. doi:10.1093/brain/awl090

12. Fridriksson J. Preservation and modulation of specific left hemisphere regions is vital for treated recovery from anomia in stroke. *J Neurosci*. 2010;30(35):11558-11564. doi:10.1523/JNEUROSCI.2227-10.2010
13. van Oers C, Vink M, van Zandvoort M, et al. Contribution of the left and right inferior frontal gyrus in recovery from aphasia. A functional MRI study in stroke patients with preserved hemodynamic responsiveness. *Neuroimage*. 2010;49(1):885-893. doi:10.1016/J.NEUROIMAGE.2009.08.057
14. Hillis AE, Kleinman JT, Newhart M, et al. Restoring Cerebral Blood Flow Reveals Neural Regions Critical for Naming. *J Neurosci*. 2006;26(31):8069-8073. doi:10.1523/JNEUROSCI.2088-06.2006
15. Blasi V, Young AC, Tansy AP, Petersen SE, Snyder AZ, Corbetta M. Word retrieval learning modulates right frontal cortex in patients with left frontal damage. *Neuron*. 2002;36(1):159-170. doi:10.1016/S0896-6273(02)00936-4
16. Leff AP, Crinion JT, Holland SK, Turkheimer FE, Howard D, Wise RJS. A physiological change in the homotopic cortex following left posterior temporal lobe infarction. *Ann Neurol*. 2002;51(5):553-558. doi:10.1002/ANA.10181
17. Turkeltaub PE, Coslett HB, Thomas AL, et al. The right hemisphere is not unitary in its role in aphasia recovery. *Cortex*. 2012;48(9):1179-1186. doi:10.1016/J.CORTEX.2011.06.010
18. Crinion J, Price CJ. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*. 2005;128(12):2858-2871. doi:10.1093/brain/awh659
19. Naeser MA, Martin PI, Baker EH, et al. Overt propositional speech in chronic nonfluent aphasia studied with the dynamic susceptibility contrast fMRI method. *Neuroimage*. 2004;22(1):29-41. doi:10.1016/j.neuroimage.2003.11.016
20. Winhuisen L, Thiel A, Schumacher B, et al. Role of the Contralateral Inferior Frontal Gyrus in Recovery of Language Function in Poststroke Aphasia: A Combined Repetitive Transcranial Magnetic Stimulation and Positron Emission Tomography Study. *Stroke*. 2005;36(8):1759-1763. doi:10.1161/01.STR.0000174487.81126.ef
21. Chang YN, Lambon Ralph MA. A unified neurocomputational bilateral model of spoken language production in healthy participants and recovery in poststroke aphasia. *Proc Natl Acad Sci U S A*. 2020;117(51):32779-32790. doi:10.1073/pnas.2010193117
22. Geranmayeh F, Brownsett SLE, Wise RJS. Task-induced brain activity in aphasic stroke patients: What is driving recovery? *Brain*. 2014;137(10):2632-2648.

- doi:10.1093/brain/awu163
23. Peñaloza C, Mirman D, Tuomiranta L, et al. Novel word acquisition in aphasia: Facing the word-referent ambiguity of natural language learning contexts. *Cortex*. 2016;79:14-31. doi:10.1016/j.cortex.2016.03.009
 24. Tuomiranta L, Rautakoski P, Rinne JO, Martin N, Laine M. Long-term maintenance of novel vocabulary in persons with chronic aphasia. *Aphasiology*. 2012;26(8):1053-1073. doi:10.1080/02687038.2012.693583
 25. McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychologi*. 1995;102(3):419-457.
 26. McClelland JL, McNaughton BL, Lampinen AK. Integration of new information in memory: new insights from a complementary learning systems perspective. *Philos Trans R Soc B Biol Sci*. 2020;375(1799):20190637. doi:10.1098/rstb.2019.0637
 27. Davis MH, Gaskell MG. A complementary systems account of word learning: neural and behavioural evidence. *Philos Trans R Soc B Biol Sci*. 2009;364:3773-3800. doi:10.1111/j.1467-9922.2010.00600.x
 28. Gore KR, Woollams AM, Bruehl S, Halai AD, Lambon Ralph MA. Direct Neural Evidence for the Contrastive Roles of the Complementary Learning Systems in Adult Acquisition of Native Vocabulary. *Cereb Cortex*. 2021;2021:1-14. doi:10.1093/CERCOR/BHAB422
 29. Breitenstein C, Jansen A, Deppe M, et al. Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage*. 2005;25(3):958-968. doi:10.1016/J.NEUROIMAGE.2004.12.019
 30. Davis MH, Di Betta AM, Macdonald MJE, Gaskell MG. Learning and consolidation of novel spoken words. *J Cogn Neurosci*. 2009;21(4):803-820. doi:10.1162/jocn.2009.21059
 31. Mestres-Missé A, Càmarà E, Rodríguez-Fornells A, Rotte M, Münte TF. Functional neuroanatomy of meaning acquisition from context. *J Cogn Neurosci*. 2008;20(12):2153-2166. doi:10.1162/jocn.2008.20150
 32. Takashima A, Bakker I, van Hell JG, Janzen G, McQueen JM. Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain Lang*. 2017;167:44-60. doi:10.1016/J.BANDL.2016.05.009
 33. Kaplan E, Goodglass H, Weintraub S. *Boston Naming Test*. Lea & Febiger; 1983.

34. Swinburn K, Porter G, Howard D. *Comprehensive Aphasia Test*. Psychology Press; 2005.
35. Kay J, Lesser R, Coltheart M. PALPA: Psycholinguistic assessments of language processing in aphasia. *PALPA*. Published online 1992:686.
36. Goodglass H, Kaplan E, Barresi B. *Boston Diagnostic Aphasia Examination*. Lippincott Williams & Wilkins; 2001.
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):1253-1266. doi:10.1016/j.neuroimage.2008.03.028
38. Butler RA, Lambon Ralph MA, Woollams AM. Capturing multidimensionality in stroke aphasia: mapping principal behavioural components to neural structures. *Brain*. 2014;137(Pt 12):3248-3266. doi:10.1093/brain/awu286
39. Bozeat S, Lambon Ralph MA, Patterson K, Garrard P, Hodges JR. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*. 2000;38(9):1207-1215. doi:10.1016/S0028-3932(00)00034-8
40. Jefferies E, Patterson K, Jones RW, Ralph MAL. Comprehension of concrete and abstract words in semantic dementia. *Neuropsychologia*. 2009;23(4):492. doi:10.1037/A0015452
41. Wechsler D. *Manual for the Wechsler Memory Scale-Revised*. Psychological Corporation; 1955.
42. Raven J. *Coloured Progressive Matrices. Sets A, Ab & B*. H. K. Lewis; 1962.
43. Burgess PW, Shallice T. *The Hayling and Brixton Tests: Thames Valley Test Company Bury St*. Thames Valley Test Company; 1997.
44. Szekeley A, D'Amico S, Devescovi A, et al. Timed action and object naming. *Cortex*. 2005;41(1):7-25. Accessed July 17, 2018. <http://www.ncbi.nlm.nih.gov/pubmed/15633703>
45. Consortium B. The British National Corpus. Published online 2007. <http://hdl.handle.net/20.500.12024/2554>
46. Abel S, Schultz A, Radermacher I, Willmes K, Huber W. Decreasing and increasing cues in naming therapy for aphasia. Published online 2005. doi:10.1080/02687030500268902
47. Conroy P, Sage K, Lambon Ralph MA. The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia. *Aphasiology*. 2009;23(6):707-730. doi:10.1080/02687030802165574

48. Conroy P, Sotiropoulou Drosopoulou C, 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(6):1815-1827. doi:10.1093/brain/awy087
49. Kundu P, Inati SJ, Evans JW, Luh WM, Bandettini PA. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage*. 2012;60(3):1759-1770. doi:10.1016/J.NEUROIMAGE.2011.12.028
50. Kundu P, Voon V, Balchandani P, Lombardo M V., Poser BA, Bandettini PA. Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals. *Neuroimage*. 2017;154:59-80. doi:10.1016/j.neuroimage.2017.03.033
51. Poser BA, Versluis MJ, Hoogduin JM, Norris DG. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magn Reson Med*. 2006;55(6):1227-1235. doi:10.1002/mrm.20900
52. Halai AD, Welbourne SR, Embleton K, Parkes LM. A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Hum Brain Mapp*. 2014;35(8):4118-4128. doi:10.1002/hbm.22463
53. Halai AD, Parkes LM, Welbourne SR. Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension. *Neuroimage*. 2015;122:214-221. doi:10.1016/j.neuroimage.2015.05.067
54. Woolrich MW, Jbabdi S, Patenaude B, et al. Bayesian analysis of neuroimaging data in FSL. *Neuroimage*. 2009;45(1 Suppl):S173-86. doi:10.1016/j.neuroimage.2008.10.055
55. Pustina D, Coslett HB, Turkeltaub PE, Tustison N, Schwartz MF, Avants B. Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis. *Hum Brain Mapp*. 2016;37(4):1405. doi:10.1002/HBM.23110
56. Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996;29(3):162-173. doi:10.1006/cbmr.1996.0014
57. Cox RW, Hyde JS. Software tools for analysis and visualization of fMRI data. *NMR Biomed*. 1997;10(4-5):171-178. doi:10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L
58. Posse S, Wiese S, Gembris D, et al. Enhancement of BOLD-contrast sensitivity by

- single-shot multi-echo functional MR imaging. *Magn Reson Med*. 1999;42(1):87-97. Accessed June 28, 2018. <http://www.ncbi.nlm.nih.gov/pubmed/10398954>
59. Community T tedana, Ahmed Z, Bandettini PA, et al. ME-ICA/tedana: 0.0.11. Published online September 30, 2021. doi:10.5281/ZENODO.5541689
60. Penny W, Friston K, Ashburner J, Kiebel S, Nichols T. Statistical Parametric Mapping: The Analysis of Functional Brain Images. *Stat Parametr Mapp Anal Funct Brain Images*. Published online 2007.
61. Brett M, Anton J-L, Valabregue R, Poline J-B. Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain. *Sendai, Japan Available on CD-ROM in NeuroImage*. 2002;16.
62. RStudioTeam. Integrated development for R. *RStudio*. 2020;42:14. <https://support.rstudio.com/hc/en-us/articles/206212048-Citing-RStudio%0Ahttps://rstudio.com>
63. Blank SC, Scott SK, Murphy K, Warburton E, Wise RJS. Speech production: Wernicke, Broca and beyond. *Brain*. 2002;125(8):1829-1838. doi:10.1093/BRAIN/AWF191
64. Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci*. 2007;8(5):393-402. doi:10.1038/nrn2113
65. Price CJ. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*. 2012;62(2):816-847. doi:10.1016/j.neuroimage.2012.04.062
66. Rosen HJ, Petersen SE, Linenweber MR, et al. Neural correlates of recovery from aphasia after damage to left inferior frontal cortex. *Neurology*. 2000;55(12):1883-1894. Accessed June 8, 2017. <http://www.ncbi.nlm.nih.gov/pubmed/11134389>
67. Heiss W-D, Thiel A. A proposed regional hierarchy in recovery of post-stroke aphasia. *Brain Lang*. 2006;98(1):118-123. doi:10.1016/j.bandl.2006.02.002
68. Naeser MA, Martin PI, Nicholas M, et al. Improved naming after TMS treatments in a chronic, global aphasia patient--case report. *Neurocase*. 2005;11(3):182-193. doi:10.1080/13554790590944663
69. Binney RJ, Embleton K V, Jefferies E, Parker GJM, Lambon Ralph MA. The Ventral and Inferolateral Aspects of the Anterior Temporal Lobe Are Crucial in Semantic Memory: Evidence from a Novel Direct Comparison of Distortion-Corrected fMRI, rTMS, and Semantic Dementia. *Cereb Cortex*. 2010;20:2728-2738.

- doi:10.1093/cercor/bhq019
70. Cabeza R, Ciaramelli E, Olson IR, Moscovitch M. The parietal cortex and episodic memory: An attentional account. *Nat Rev Neurosci.* 2008;9(8):613-625. doi:10.1038/nrn2459
 71. Humphreys GF, Jackson RL, Lambon Ralph MA. Overarching Principles and Dimensions of the Functional Organization in the Inferior Parietal Cortex. *Cereb Cortex.* 2020;30(11):5639-5653. doi:10.1093/CERCOR/BHAA133
 72. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B.* 1995;57(1):289-300. doi:10.1111/J.2517-6161.1995.TB02031.X
 73. Rolls ET, Huang CC, Lin CP, Feng J, Joliot M. Automated anatomical labelling atlas 3. *Neuroimage.* 2020;206:116189. doi:10.1016/J.NEUROIMAGE.2019.116189
 74. Breitenstein C, Grewe T, Flöel A, et al. Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *Lancet (London, England).* 2017;389(10078):1528-1538. doi:10.1016/S0140-6736(17)30067-3
 75. Conroy P, Sage K, Ralph ML, Lambon. Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalize to connected speech? *Int J Lang Commun Disord.* 2009;44(6):1036-1062. doi:10.1080/13682820802585975
 76. Crinion JT, Leff AP. Using functional imaging to understand therapeutic effects in poststroke aphasia. *Curr Opin Neurol.* 2015;28(4):330-337. doi:10.1097/WCO.0000000000000217
 77. Stefaniak JD, Alyahya RSW, Lambon Ralph MA. Language networks in aphasia and health: A 1000 participant activation likelihood estimation meta-analysis. *Neuroimage.* 2021;233:117960. doi:10.1016/J.NEUROIMAGE.2021.117960
 78. Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci.* 2009;12(6):718-724. doi:10.1038/nn.2331
 79. Howard D, Best W, Nickels L. Optimising the design of intervention studies: critiques and ways forward. *Aphasiology.* 2015;29(5):526-562. doi:10.1080/02687038.2014.985884
 80. Best W, Greenwood A, Grassly J, Herbert R, Hickin J, Howard D. Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex.* 2013;49(9):2345-2357. doi:10.1016/J.CORTEX.2013.01.005

Figure legends

Figure 1. Lesion overlap map of the 16 participants on axial slices. Colour bar indicates the number of participants included in the overlap. Numbering above slices indicates z MNI coordinate. In neurological convention, left is left.

Figure 2. Accuracy and RT of naming a) untreated items, b) treated items and c) already known items. Dotted line represents the mean. Solid lines between points represent individual patient differences in time points. Group differences significant at $p < .0001$ and $p < .005$ are denoted with ‘***’ and ‘**’ respectively.

Figure 3. Whole brain clusters of increased activation during picture naming for (A) controls and (B) patients. Picture naming of treated $>$ untreated (blue) and known $>$ untreated (red) contrasts, with overlap (purple). Image thresholded at voxel level $p < .001$, cluster corrected FWE $p < .05$. R; right, P; posterior.

Figure 4. Correlations between regional activation and behavioural performance. (A) Regions of interest: green = left inferior frontal gyrus (IFG); yellow = left inferior parietal lobe; cyan = left anterior temporal lobe (ATL); blue = right ATL; red = left hippocampus (Hipp.). **(B)** Correlations between treated $>$ untreated BOLD activity and accuracy. **(C)** Left: correlation between treated $>$ untreated left hippocampal BOLD activity and RT. Right: Predicted simple slopes of the association between treated $>$ untreated left hippocampal BOLD activity and RTs at three levels of left IFG damage.