

## ***Reorganization of the neurobiology of language after sentence overlearning***

Jeremy I Skipper<sup>1\*</sup>, Sarah Aliko<sup>1,2</sup>, Stephen Brown<sup>3</sup>, Yoon Ju Jo<sup>1</sup>, Serena Lo<sup>4</sup>, Emilia Molimpakis<sup>5</sup> and Daniel R Lametti<sup>1,6</sup>

<sup>1</sup>Experimental Psychology, University College London, UK

<sup>2</sup>London Interdisciplinary Biosciences Consortium, University College London, UK

<sup>3</sup>Natural Sciences, University College London, UK

<sup>4</sup>Speech and Language Sciences, University College London, UK

<sup>5</sup>Wellcome Centre for Human Neuroimaging, University College London, UK

<sup>6</sup>Department of Psychology, Acadia University, Nova Scotia, Canada

\*Corresponding author, [jeremy.skipper@ucl.ac.uk](mailto:jeremy.skipper@ucl.ac.uk)

### **Abstract**

There is a widespread assumption that there are a static set of ‘language regions’ in the brain. Yet, people still regularly produce familiar ‘formulaic’ expressions when those regions are severely damaged. This suggests that the neurobiology of language varies with the extent of word sequence learning and might not be fixed. We test the hypothesis that perceiving sentences is mostly supported by sensorimotor regions involved in speech production and not ‘language regions’ after overlearning. Twelve participants underwent two sessions of behavioural testing and functional magnetic resonance imaging (fMRI), separated by 15 days. During this period, they repeated two sentences 30 times each, twice a day. In both fMRI sessions, participants ‘passively’ listened to those two sentences and novel sentences. Lastly, they spoke novel sentences. Behavioural results confirm that participants overlearned sentences. Correspondingly, there was an increase or recruitment of sensorimotor regions involved in sentence production and a reduction in activity or inactivity for overlearned sentences in regions involved in listening to novel sentences. The global network organization of the brain changed by ~45%, mostly through lost connectivity. Thus, there was a profound reorganization of the neurobiology of speech perception after overlearning towards sensorimotor regions not considered in most contemporary models and away from the ‘language regions’ posited by those models. These same sensorimotor regions are generally preserved in aphasia and Alzheimer’s disease, perhaps explaining residual abilities with formulaic language. These and other results warrant reconsidering static neurobiological models of language.

### **Introduction**

It is widely assumed that the brain regions supporting speech production, perception and language comprehension are largely spatially fixed. Historical models consider ‘Broca’s area’ the locus of speech production and ‘Wernicke’s area’ the locus of comprehension<sup>1,2</sup>. Popular contemporary models include these left hemisphere inferior frontal gyrus and perisylvian regions and a small number of others, most notably premotor cortex and bilateral superior and middle temporal gyri<sup>3-11</sup>. Illustrating this fixity assumption, various aspects of the inferior frontal gyrus and superior temporal plane are regularly described as ‘the language network’, with the latter phrase appearing in more than 5500 articles on Google Scholar (assessed September, 2020)<sup>8</sup>.

Yet, there is a long history of evidence suggesting that language is more distributed throughout the brain than is belied by these models<sup>12-14</sup>. This is illustrated by test-retest reliability studies that use language stimuli or that explicitly pertain to the reliability of regions involved in language processing. These show that stable individual participant activity patterns and networks are more variable and distributed than the set of aforementioned ‘language regions’<sup>15-20</sup>. Though there are a number of reasons this might be, a significant amount of the variance can be accounted for by individual differences in task strategies and cognitive style, like the tendency to visualize words<sup>19,21</sup>.

A more specific example derives from the neurobiology of lexical processing. Words activate sensory and motor regions associated with their meaning, including ‘language regions’ and far beyond. The written word ‘telephone’ activates auditory cortex more than words that do not have auditory connotations<sup>22</sup>, ‘red’ activates the visual colour region V4<sup>23</sup>, ‘kick’ activates dorsal motor regions involved in moving the legs<sup>24</sup> and ‘garlic’ activates olfactory cortex<sup>25</sup>. These activity patterns occur early, within 50-150 ms of word onset, while they are still being read or heard<sup>22,26-28</sup>. This implies that the increased involvement of sensory and motor regions is not simply a post-perceptual process, somehow separable from true ‘language regions’.

These test-retest reliability and word processing examples suggest that ‘language regions’ or ‘the language network’ underrepresent the distributed nature of the neurobiology of language. The difference between the empirical data and popular theoretical models might be explained by the reliance on measures of central tendency. That is, the majority of experiments upon which these models are (presumably) based, average over individuals’ unique though reliable differences in language related activity patterns. This results in ‘language regions’, even if the peaks of any given participant's blobs are not in those regions<sup>17</sup>. Similarly, the vast majority of studies average over individual words and sentences from many unrelated semantic categories, averaging out their unique distributed patterns, leaving only ‘language regions’.

### ***Formulaic Language***

Individual differences and word processing are only two sources of spatial variability suggesting that the organization of language and the brain is more distributed than suggested by contemporary models. Another possible source of variability, related to both examples, is how well one has learned words and sequences of words. Neuroimaging studies that use multiword or sentence stimuli typically average over word sequences that are more or less formulaic. Formulaic language can be defined as:

*A sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.*<sup>29</sup>

Formulaic language is synonymous with or includes many other terms like ‘automatic speech’, ‘chunks’, ‘collocations’, ‘holistic’, ‘fixed’, ‘multiword expressions’ ‘noncompositional’, ‘preassembled’, ‘prefabricated’ and ‘ready-made’, to name ten<sup>29,30</sup>. Examples range from explicatives (‘shit’<sup>31</sup>), idioms (‘sort of’<sup>32</sup>), lists (‘one, two, three’), high frequency n-grams (‘the end of the’<sup>33</sup>), to individualised overlearned material (such as ‘I have of late, but wherefore I know not, lost all my mirth’<sup>34</sup>). Formulaic expressions are important for first and second language acquisition<sup>35-37</sup> and are ubiquitous in everyday usage, comprising

a third or more of language<sup>38</sup>. They are processed faster with fewer errors in both children and adults compared to novel words, attesting to their special status<sup>39-44</sup>.

Indeed, when they are not averaged together with less formulaic words, there is evidence that formulaic language processing might have a more distributed pattern than is suggested by popular neurobiological models of language. That is, since the 19th century, it has been known that the ability to produce formulaic speech is often preserved in aphasia, even with severe language impairment and damage to most and sometimes all of ‘the language network’ (as in global aphasia)<sup>45-54</sup>.

Where in the brain are formulaic expressions stored and processed when ‘language regions’ have been destroyed? One possibility is suggested by the relative behavioural and corresponding cortical preservation in aphasia and Alzheimer’s disease. Lesion locations strongly associated with language do not typically include sensorimotor regions around the central sulcus, i.e., primary motor and somatosensory regions<sup>55-62</sup>. Similar to aphasia, people with Alzheimer’s disease also produce more formulaic language than controls<sup>63-66</sup> and the amount of formulaic language produced predicts disease progression<sup>65</sup>. In addition to formulaic expression production, individuals with Alzheimer’s disease retain other motor functionality. This corresponds to the relative degradation of temporal cortices and sparing of regions and connectivity around the central sulcus<sup>67-73</sup>. In contrast, individuals with Parkinson’s disease produce less formulaic language<sup>63,66,74,75</sup>, have more motor problems and this corresponds to a breakdown of cortical/subcortical sensorimotor networks<sup>76-78</sup>.

### ***Speech Perception***

This work suggests that preserved formulaic language production in aphasia and Alzheimer’s disease is more reliant on sensorimotor regions and less on typical ‘language regions’ or ‘the language network’. By extension, this implies that, as learning increases and words become more formulaic, they become more reliant on these same sensorimotor regions in healthy individuals. In some models of speech perception, it has been suggested that regions of the brain supporting speech production, including sensorimotor regions, are also variously involved in speech perception<sup>79</sup>. As such, the perception and comprehension of formulaic expressions might also rely more on sensorimotor regions and less on ‘language regions’.

This suggestion is supported by neuroimaging studies of music and speech learning. These collectively show that perception after learning involves more engagement of the production systems used during learning. For example, in both monkeys and humans, learning to play sounds on a keyboard is subsequently associated with specific activation of sensorimotor regions involved in making finger and hand movements when listening to those sounds<sup>80-85</sup>. More generally, musical training improves sound and speech perception and these improvements are typically associated with sensorimotor regions<sup>85-90</sup>. Similarly, learning new speech sounds, novel words and artificial languages are all associated with more sensorimotor activity during perception following learning and consolidation<sup>79,91-99</sup>.

What mechanistic account makes sense of ‘hearing’ more with sensorimotor cortices and less with ‘language regions’? All listeners are faced with the problem of achieving perceptual constancy during speech perception. This is because there is variance in acoustic patterns both across and within talkers and no, as of yet, discoverable mapping between these patterns and speech categories (i.e., phonemes, syllables, etc.). We and others have argued that this

difficulty might be solved if, during perception, the brain makes use of the contextual information that accompanies speech and the capacity of motor systems to predict the sensory consequences of movements (an ability important for motor learning and adjusting movements in real time, known as ‘efference copy’)<sup>12,79,100,101</sup>.

For example, hearing ‘She was tired of her life and felt ready for a ...’ preactivates ‘change’. This is sequenced by motor regions involved in speech production as if it were to be spoken. Through efference copy from sensorimotor regions, the motor pattern for producing ‘ch’ activates acoustic patterns for ‘ch’ in auditory cortices. If there is an overlap with incoming acoustic information, interpretation is confirmed and further processing of ‘change’ is unnecessary, conserving metabolic resources (relative to a less predictive context). Indeed, there is a large reduction in activity in the entire superior temporal plane in a variety of more predictive contexts during speech perception and language comprehension<sup>101–105</sup>. By this account, the more overlearned and formulaic a sequence of words, the earlier the whole sequence becomes predictable. In the example, ‘change’ might be predictable at ‘ready’ before overlearning and ‘tired’ after. Correspondingly, perception will be more supported by sensorimotor regions and much less by ‘language regions’.

### ***Hypothesis***

Based on these speculations, we tested the hypothesis that, as sentences become formulaic through production-based overlearning, there will be a reorganization of the brain regions supporting the perception of those sentences. Behaviourally, we operationalised overlearning as a decrease in reaction times to predict the final word of two sentences, to process the individual words in those sentences and the ability to accurately remember the sentences following two weeks of home production based listening and repetition.

Neurobiologically, reorganization was predicted to correspond to a large increase in activity or new activations of sensorimotor regions. Here and throughout, we define ‘sensorimotor regions’ as pre and primary motor and primary and secondary somatosensory cortices in and adjacent to the central sulcus. These and other regions were expected to support the production of novel sentences. Concomitantly, we predicted a large decrease or inactivity in ‘language regions’ while listening to overlearned sentences. ‘Language regions’ were defined as activation when listening to novel sentences not previously heard or overlearned. Finally, we expected these changes to correspond to a global network reorganization of the brain.

Importantly, we used a natural or so called ‘passive’ design in which participants listened without making meta-linguistic judgements or corresponding button or vocal responses. This is because our hypothesis maintains that sensorimotor regions involved in production support perception after overlearning. If we had included motor responses, there would be no verifiable way to argue that resulting sensorimotor activity was not due to those movements. Figure 1 provides an overview of the study design used to test these predictions.

-----  
**Figure 1**  
-----

## Methods

### *Participants*

Twelve native British English speakers participated (6 females; 21-25 years old;  $M = 23.17$ ;  $SD = 1.41$ ). All but one was monolingual. Participants were right-handed as determined by the Edinburgh Handedness Inventory<sup>106</sup>. All had unimpaired hearing and (corrected) vision. None had any contraindication for MRI, history of psychiatric or neurological disorder or language-related learning disabilities. All participants gave informed consent and the study was approved by the University College London ethics committee.

### *Procedure*

The experiment lasted 17 days, including two testing days, each with three different behavioural tasks (completed multiple times) and fMRI to assess overlearning of two sentences (Figure 1). On the first day, participants performed a sentence completion and lexical decision task on a desktop computer in a noise attenuated testing room, using headphones. Both tasks included the words from the two sentences that the participants would overlearn over the subsequent 15 days.

Following these tasks, participants were escorted to the scanning suite. There they chose comfortable earbud sizes for noise-attenuating headphones. After being instructed, they were put in the head-coil with pillows under and on the sides of their head and (if desired) under the knees for comfort and to reduce movement during scanning. Once in place, participants chose an optimal stimulus volume by determining a level that was loud but comfortable. Once scanning began, participants' first fMRI task was to 'passively' listen to the two 'overlearned' sentences multiple times and previously unheard or 'novel' sentences. Participants' last fMRI task was to listen to and repeat some of the novel sentences that they had earlier heard in the scanner. Finally, we acquired high-resolution anatomical scans. After scanning, participants were returned to the testing room where they did the sentence completion and lexical decision tasks a second time and a sentence recall task to assess learning over the first day.

Over the next 15 days, participants trained at home by listening to and repeating the two overlearned sentences they heard during fMRI. They did this twice a day, sending us recordings of their productions when they were done. Participants returned on the final (17th) day and performed the sentence completion, lexical decision and sentence recall behavioural tasks, fMRI passive listening and speech production tasks and anatomical scans as on day one, in the same order. When these were complete, participants were given £7.5 per hour for behavioural testing and £10 per hour for scanning to compensate for their time and sent home.

### *Stimuli*

Participants were divided into three groups. Each group of four participants overlearned two different sentences. Three pairs of sentences were used to help assure that results are generalizable. Specifically, a male talker recorded the 498 sentences from the Supplemental Materials of Block and Baldwin<sup>107</sup>. Sentences were edited in Praat (<http://www.fon.hum.uva.nl/praat/>) to be 2.5 seconds, with the final word lasting 500 ms (see next 'Sentence Completion' section for a rationale). Four raters judged whether the sentences were appropriate for British English listeners (e.g., they do not discuss American football or 'pants') and sounded natural (i.e., they were not sped up or slowed down). The latter was done

using a Likert scale from one (not natural) to 10 (very natural). Inappropriate sentences, those with a naturalness rating less than six and sentences with proper nouns were discarded. From the remaining sentences, three sets of two sentences were created that were matched on number of words, cloze probability (i.e., the probability of the final word completing the preceding words)<sup>107</sup>, complexity and average word frequency as determined by Subtlex-UK<sup>108</sup> (Table 1). From the remaining sentences, 60 high cloze probability sentences were selected to be used in behavioural tasks and during scanning.

### ***Behavioural Tasks***

Three behavioural tasks were used to assess overlearning of the two sentences assigned to each participant. Sentence completion and lexical decision tasks were completed both before and after the two fMRI sessions. The sentence recall task was done only after each imaging session.

#### **Sentence Completion**

Participants listened to two sentences that were to be overlearned (sessions one and two) or had been overlearned (sessions three and four) and 30 novel sentences. In each case, the final word in the sentence was removed. Participants were asked to press a button as soon as they believed they knew the final word. The sentence stopped playing when the button was pressed and participants then typed in the final word as quickly as possible. Reaction times for the task were measured from the start of sentence playback until participants finished entering the predicted final word. On a per participant basis, reaction times greater than four standard deviations from the mean reaction were not included in the analysis.

#### **Lexical Decision**

Participants had to judge the lexical status (word or not word) of overlearned words, words similar to overlearned words, words dissimilar to overlearned words and nonwords. Thirty overlearned words were extracted from the overlearned sentences used in the study. Eighteen similar words were selected by searching through the remaining novel sentences and finding words that had a Levenshtein distance of two or less from the overlearned words extracted from the overlearned sentences. Levenshtein distance is the minimum number of single character alterations required to change one word into another<sup>109</sup>. Sixteen dissimilar words were also identified. These had a Levenshtein distance of six or more and also had a frequency that differed by  $\pm 10\%$  from the overlearned words. Sixty-four unintelligible but speech-like nonwords were created from the selected (i.e., overlearned, similar, dissimilar) words using a local time reversal script in Praat with 150 ms steps<sup>110</sup>.

Each participant heard only 21 words (nine overlearned words taken from the two sentences they repeated, six similar words and six dissimilar words) and 21 nonwords (reversed versions of the overlearned, similar and dissimilar words they heard). After hearing each word they indicated if they heard a word or a non-word with a button-press as quickly and accurately as possible. Reaction times were measured from the start of playback to the moment a response was indicated. On a per participant basis, reaction times greater than four standard deviations from the mean reaction were not included in the analysis. We then examined how reaction times changed between the two scanning sessions.

## **Sentence Recall**

After scanning sessions, participants were asked to type in up to 10 sentences that they remembered hearing in the scanner. The sentences they recalled (e.g., ‘He was a lousy cook and ordered out’) were matched (based on semantics and word use) to the sentences that were actually played in the scanner (e.g., ‘He eats out because he is a lousy cook’). The cosine similarity was then found between the two sentences to provide a measure of recall accuracy. The cosine similarity is the cosine of the angle between two normalized vectors. In this case, the two sentences were transformed into vectors of word counts and the cosine similarity was found for these word count vectors. Cosine similarity takes a value between zero and one, where a value of zero means that the sentences share no words and a value of one means that the sentences are identical in terms of words used.

## **Sentence Overlearning**

Commencing the day after the first fMRI session, participants overlearned two sentences through repetition at home, twice a day. Each day, participants listened to a prerecorded set of their two sentences, each repeated 30 times in a random order, lasting five minutes and 32 seconds. There was a three second gap between sentences during which the participant repeated the sentence out loud. The participants did this task again at least six hours later using another randomised prerecorded file of the same two sentences. All 30 prerecorded files had a different randomisation. To verify home learning took place, participants recorded themselves listening to and repeating their sentences using Audacity (<http://www.audacityteam.org>) and immediately shared the recording to a cloud storage folder. In total, participants listened and repeated their two sentences 1,800 times for two hours and 46 minutes. This was verified by checking recordings.

## ***fMRI***

### **Task**

A slow random event related design was used to compare how the response to overlearned sentences changed from fMRI session one to two. In each scanning session, audio stimuli were presented during six listening runs and a speech production run. Each listening run lasted six minutes and 53 seconds. Across these, participants heard two sentences repeated 60 times each and 60 novel sentences that were each only heard once, all in a randomised order. Stimuli were presented in a jittered manner such that, following each 2.5 second sentence, there was a minimum of 10 seconds of silence and a mean of 10.675 seconds ( $SD = 0.93$ ) and a maximum of 15 seconds of silence. The speech production run always followed the six listening runs and lasted eight minutes and 10 seconds. In this run, participants listened to 30 of the novel sentences they had just heard and were asked to repeat each sentence as soon as it finished. After each 2.5 second sentence and allowing for another 2.5 second period to produce the sentence, there was, again, a minimum of 10 seconds of silence with a mean of 10.69 seconds ( $SD = 1.34$ ) and a maximum of 15.625 seconds of silence. All listening and production runs included 10 seconds of silence at the start to allow magnetization to reach a steady state. Sessions one and two were the same, with the exception that in session two they produced the other 30 sentences not produced in session one. Participant engagement was monitored with a camera over one eye.

Functional and anatomical images were acquired on a 1.5T Siemens MAGNETOM Avanto with a 32 channel head coil (Siemens Healthcare, Erlangen, Germany). We used multiband

EPI<sup>11,12</sup> (TR = 700 ms, TE = 54.8 ms, flip angle of 75°, 28 slices, resolution = 3 x 3 x 4 mm), with 4x multiband factor and no in-plane acceleration. Slices were manually obliqued to include the entire brain. The six listening EPI runs were each 590 volumes/TRs and the speech production run had 700 volumes/TRs. Two six minute T1-weighted high-resolution MPAGE anatomical MRI scans followed the functional scans (TR = 2.73 s, TE = 3.57 ms, 176 sagittal slices, resolution = 1.0 mm<sup>3</sup>). Imaging parameters were the same for both fMRI sessions (thus resulting in 12 listening, two production runs and four anatomical scans).

### Preprocessing

Unless otherwise noted, the AFNI software suite was used for preprocessing and analyses (<http://afni.nimh.nih.gov/afni>)<sup>13</sup> Individual AFNI programmes are indicated parenthetically in subsequent descriptions.

The four anatomical/structural MRI scans were corrected for image intensity non-uniformity (*'3dUniformize'*) and deskulled using *ROBEX*<sup>14</sup>. Within each session, the second anatomical image was aligned to the first and they were averaged. Then the resulting session one and two anatomical images were aligned and averaged. This was done using a procedure to reduce bias by moving both anatomical images, so that both are interpolated some amount rather than one session receiving all the interpolation (see <https://sscc.nimh.nih.gov/sscc/dglen/alignmentacross2sessions>).

The resulting anatomical image was nonlinearly aligned (using *'auto\_warp.py'*) to the MNI N27 template brain, an average of 27 anatomical scans from a single participant (*'Colin'*)<sup>15</sup>. The anatomical scan was inflated and registered with Freesurfer software using *'recon-all'* and default parameters (version 6.0, <http://www.freesurfer.net>)<sup>16,17</sup>. This included automatic parcellations of the anatomical image. These were used to create white matter and ventricle (i.e., cerebral spinal fluid containing) regions of interest that were used as noise regressors. Automatic parcellation was also used to generate 167 regions of interest for network analyses (i.e., using the *'Destrieux Atlas'*)<sup>17</sup>.

The first 10 TRs (7 seconds) were removed from the fMRI timeseries before they were corrected for slice-timing differences (*'3dTshift'*) and despiked (*'3dDespike'*). Next, volume registration was done by aligning each timepoint to the mean of run four (*'3dvolreg'*). The functional data were then aligned to the anatomical images (*'align\_epi\_anat.py'*). This used the less biased procedure described for anatomical alignment, moving functional data from both sessions so that no one session received all the interpolation. Finally, the volume-registered and anatomically-aligned functional data were (nonlinearly) aligned to the MNI template brain (*'3dNwarpApply'*).

Next, we created two sets of timeseries. The first, to be used in the deconvolution analysis described below, involved only normalising each run to have a sum of squares of one (*'3dTproject'*). The second set of time series were normalised and detrended using Legendre polynomials whose degree varied with run lengths (following the AFNI recommended formula of [number of timepoints \* TR]/150). These were then submitted to spatial independent component analysis (ICA) to detect and remove artifacts. Specifically, we concatenated the normalised and detrended listening and production timeseries from both sessions separately (*'3dTcat'*). We did ICA on the resulting listening timeseries with 300 dimensions and on the production timeseries with 100 dimensions using *'melodic'* (version



3.14) from FSL<sup>118</sup>. Next, we labelled and removed artifacts from the timeseries, following recommendations from an existing guide for manual classification<sup>119</sup>. One of two trained authors went through all components and associated timecourses, labelling the components as ‘good’, ‘maybe’ or ‘artifact’. Our strategy was to preserve signal by not removing components classified as ‘maybe’.

Finally, we made a third timeseries using the concatenated listening runs for the regional homogeneity analysis described below. Specifically, the timeseries were normalised to have a sum of squares of one and detrended (*3dTproject*) with the following regressors: 1) Legendre polynomials whose degree varied with run lengths (following the previously described formula); 2) Six demeaned motion regressors from the volume registration (roll, pitch, yaw and changes in the inferior/superior, left/right and anterior/posterior directions); 3) A demeaned white matter activity regressor from the averaged timeseries in white matter regions; 4) A demeaned cerebrospinal fluid regressor from the averaged timeseries activity in ventricular regions; and 5) the ICA artifact component timecourses.

### Individual Deconvolutions

After preprocessing, two individual participant deconvolutions were conducted to get an estimation of the system impulse response function for the 1) overlearned and novel sentences from the listening runs and the 2) produced novel sentences from production run from both fMRI sessions (*3dDeconvolve*)<sup>120,121</sup>. In the first deconvolution, regressors of interest included one each for overlearned sentence one in session one, overlearned sentence two in session one, novel sentences in session one, overlearned sentence one in session two, overlearned sentence two in session two and novel sentences in session two. For each of these, the hemodynamic response was estimated using a cubic spline basis function that covered an 18 second period after each stimulus onset, using 20 tent functions to generate the impulse response function for every voxel. In a second deconvolution, regressors of interest included speech production of novel sentences in session one and speech production of novel sentences in session two. Again, the cubic spline basis function was used with the difference that the period covered was 20 seconds (to account for the extra time involved in producing the sentences), using 21 tent functions.

In both deconvolutions, regressors of noninterest included an automatically estimated number of polynomials (again following the  $[\text{number of timepoints} * \text{TR}]/150$  formula), six motions regressors from the volume registration step (roll, pitch, yaw and changes in the inferior/superior, left/right and anterior/posterior directions), two regressors from the average timeseries in the white matter and ventricles and all timecourses from ICA components labelled as artifacts.

### Novel LMM

Following individual participant deconvolutions, the resulting impulse response functions were spatially smoothed to achieve a level of smoothness of six mm full-width half maximum, regardless of the smoothness it had on input (*3dBlurToFWHM*)<sup>122</sup>. These were then used in four linear mixed-effects models (LMMs; *3dLME*)<sup>123</sup>.

First, we did a novel sentence listening LMM. Factors were session (one and two) and timepoint (0-25). This allowed us to test the prediction that learning was specific to overlearned sentences. Though we did not expect differences in activity for the novel

sentences across sessions, participants might have learned something about the talker's voice as the same talker made all sentences used in the study. Time was included as a factor to compare to the overlearning LMM (see next paragraph) though, again, we did not expect that there would be differences in the shapes of the hemodynamic response between sessions one and two. There were 26 timepoints because our TR was 700 ms and there are, therefore, 26 TRs covering the 18 second period from the deconvolution. Finally, collapsing over session and time, the novel LMM served to identify 'language regions' which were expected to encompass the bilateral inferior frontal gyrus and superior temporal plane.

### **Overlearning LMM**

Second, we conducted an overlearning sentence listening LMM with sentence (overlearned sentence one and two), session (one and two) and timepoint (0-25) as factors. This allowed us to understand the effect of learning on overlearned sentence listening between sessions one and two. Time is included as a factor because we expect differences in the shapes of the hemodynamic response across sessions though we did not make a priori predictions about the direction of those differences in individual brain regions. We also visualised the timepoints from the results to better understand if responses for overlearned sentences are a simple redistribution of activity (i.e., a relative modulation of activity from session one to two) or a reorganization of brain responses (i.e., activity in regions in session one or two that was not previously present).

### **Overlearning-*Novel* LMM**

Third, a followup analysis was conducted by subtracting the novel from the overlearned impulse response functions at each timepoint in each participant. We then ran a LMM with the same factors as the overlearning sentence listening LMM (i.e., session\*sentence\*timepoint). This allowed us to formally test whether overlearned sentences produced significantly more activity than novel sentences in sensorimotor regions and less activity than novel sentences in 'language regions'. We present the results of each session separately so that the direction of effect can be interpreted.

### **Production LMM**

Finally, we did a novel speech production LMM with session (one and two) and timepoint (0-28) as factors. This analysis was used to demonstrate regions involved in producing the novel sentences that participants heard during both sessions.

To correct for multiple comparisons in all LMMs, we used a multi-threshold approach rather than choosing an arbitrary p-value at the individual voxel level as is customary. In particular, we used a cluster simulation method to estimate the probability of noise-only clusters using the spatial autocorrelation function from the residuals in each LMM ('*3dFWHMx*' and '*3dClustSim*'). This resulted in the cluster sizes to achieve a corrected alpha value of 0.01 at nine different p-values (i.e., 0.05, 0.02, 0.01, 0.005, 0.002, 0.001, 0.0005, 0.0002 and 0.0001). We thresholded each map at the corresponding z-value for each of these nine p-values and associated cluster size. We then combined the resulting maps, leaving each voxel with its original z-value.

## Regional Homogeneity

The described deconvolution approach uses a linear model to derive an estimate of the hemodynamic response from multiple stimulus presentations. We reasoned that a strong case for the hypothesis could be made if a similar set of results were obtained using a more model-free approach across the whole timeseries. To do this, we used regional homogeneity which calculates the Kendall's coefficient of a concordance for each voxel within a neighbourhood of voxels (*'3dReHo'*)<sup>124,125</sup>. We chose this particular approach because it is also a measure of local interactions, synchronisation and connectivity<sup>125</sup>. Corresponding to our hypothesis, we expected an increase in local connectivity in sensorimotor regions and a decrease in superior temporal plane and inferior frontal regions after overlearning.

To do this analysis, we first constructed three timeseries that theoretically reflect only timepoints for processing overlearned sentences from session one, overlearned sentences from sessions two or novel sentences from both sessions. Specifically, we modelled expected hemodynamic responses by convolving stimulus onsets with a canonical hemodynamic response function (using 'WAV', a.k.a the 'Cox special' from *'waver'*). We then cut up the third timeseries described in 'Preprocessing' above by taking the relevant timepoints under the canonical response starting from the timepoint that the response starts to rise (a delay of 2.1 seconds or three TRs) and ending when the response returns to baseline (*'3dTcat'*). We removed any timepoints that overlapped in any of the timeseries. Because this was a slow event related design with jitter, this amounted to only 7.02% of the data that was about equally distributed across the three sentence timeseries.

We then did the regional homogeneity analysis for each of the three resulting timeseries using a radius of 2.3 which equals 57 voxels (we also tested 2.0 or 33 voxels and 2.9 or 93 voxels and it makes little difference to the results). The resulting maps were then blurred to achieve a level of smoothness of six mm full-width half maximum (*'3dBlurToFWHM'*). After this, we conducted three group paired t-tests to compare 1) the overlearned sentences from session one to those from session two; 2) the novel sentences from sessions one and two to the overlearned sentences from session one; and 3) the novel sentences from sessions one and two to the overlearned sentences from session two (*'3dttest++'*).

We saved the results as z-scores and used the residuals from the t-tests to do the same multi-thresholding procedure described for the LMM analyses to correct for multiple comparisons. We note that *3dttest++* has a built-in and more sophisticated equitable thresholding procedure (called ETAC)<sup>126</sup>. However, at the time of analysis, there was no obvious way to use this approach with, e.g., the LMM results. As such, we opted to use our described approach for consistency across analysis. That said, we looked at the results using the ETAC approach for comparison and they look similar to our multi-thresholding approach.

## Network

To do network analyses, we first concatenated the two overlearned sentence impulse response functions from the deconvolution for the two sessions separately. We blurred these as in previous analyses to a level of smoothness of six mm full-width half maximum (*'3dBlurToFWHM'*). Using the Freesurfer 'Destrieux Atlas' parcellation, 167 regions of interest were extracted from these timeseries for each participant. Pairwise Pearson's correlations were used to build two unweighted, undirected adjacency matrices for each

participant, one each for sessions one and two overlearned sentences. Absolute thresholding of  $r = 0.1$  was applied to correlation values, in order to build adjacency matrices for group comparisons<sup>127,128</sup>.

To test whether global network reorganization took place after overlearning, the distance measures ‘edit distance’ and ‘Deltacon’ were calculated. Edit distance computes additions or deletions of connections between two graphs<sup>129</sup>. The edit distance matrix was defined as:

$$(G, G') = \|A - A'\|$$

Where  $A$  and  $A'$  are the adjacency matrices for graphs  $G$  (session one) and  $G'$  (session two) respectively, and  $\delta$  is the pairwise edit distance<sup>129</sup>. Since session one and two shared node identity, this pairwise application was applicable. Change in connectivity was calculated as the ratio between the total number of lost and gained connections in the edit distance matrix and the total number of connections in both session one and two adjacency matrices for a single participant. A one-sample t-tests across participants was performed on the change in connectivity values to determine if they differed from the null, i.e. no change in connectivity.

In order to describe and visualise connections involved in edit distance differences from session one to two, a two-way chi-square test was performed on each region of interest pair across participants. Lost connections were defined as those whose connectivity in session two was lower while observed connectivity in session one was higher than expected. Gained connections were defined as the converse. We set a threshold at  $p < .01$  to afford some protection for multiple comparisons.

Though the edit distance matrix is simple to compute, it suffers from limitations. It only determines specific connection changes, but it does not interpret the change in the context of the rest of the network and its neighbours. Moreover, it does not differentiate between network densities: if a connection is lost in a very sparse network the result would be a large disruption, but if a connection is lost in a highly dense network the outcome on the global scale will be minimal<sup>130</sup>. For these reasons we also calculated Deltacon, a more robust similarity measure that determines the level of isomorphism between two networks with node correspondence, using Matusita’s distance<sup>130</sup>. We compared the results using a one-sample t-test across participants on the deltacon dissimilarity value (i.e. 1-Deltacon).

To further examine possible changes between session one and two, we explored a number of other global network measures. These included efficiency (average inverse shortest path length)<sup>131</sup>, diffusion, density and flow using the Brain Connectivity Toolbox in MATLAB<sup>132</sup>. Density measures how ‘connected’ a network is, diffusion measures how quickly information can get from point  $A$  to  $B$  and flow measures how centralised a network is for transfer of information<sup>132</sup>. We used a two-sample t-tests for these global measures, comparing sessions one and two. Finally, we computed two measures of local connectivity, centrality and community partitioning. Centrality (degree, eigenvector, closeness and betweenness) measures the importance of a node in a network, whilst community detection partitions the network into distinct sub-components or modules<sup>132</sup>.

## Results

### *Behavioural Tasks*

We hypothesized that participants would show behavioural markers of overlearning for the sentences they repeated at home. We assessed this with three behavioural tasks (Figure 2).

#### **Sentence Completion**

Figure 2A shows sentence completion times (in which participants completed the final word of a sentence) across testing sessions for both overlearned and novel sentences. There was an interaction between completion time measured at each session and whether the sentence was overlearned or not ( $F(3,33)=22.28$ ,  $p < .001$ ). For overlearned sentences, a decrease in sentence completion times was observed following training (session two versus session 3;  $t(11) = 3.90$ ,  $p < .01$ ). A similar decrease was not observed for novel sentences ( $t(11) = -1.04$ ,  $p = .32$ ). By the end of the study, participants completed the final word in the overlearned sentences 2.10 seconds faster than they did at the start of the experiment. By comparison, the time it took to complete the final word in novel sentences improved by 1.0 seconds.

-----  
**Figure 2**  
-----

#### **Lexical Decision**

Figure 2B shows changes in lexical decision times following overlearning (testing sessions one and two versus sessions three and four), for words drawn from overlearned sentences (red bar), words similar to overlearned words (blue bar), words dissimilar to overlearned words (grey bar) and nonwords (black bar). Changes in lexical decision time between the four word types were not statistically significantly different ( $F(3,33) = 2.32$ ,  $p = .094$ ). However, words drawn from overlearned sentences were identified faster following overlearning than they were prior to training ( $t(11) = 2.59$ ,  $p = .025$ ). A comparable result was not observed in the cases of similar ( $t(11) = 1.49$ ,  $p = .16$ ) and dissimilar words ( $t(11) = .49$ ,  $p = .63$ ).

#### **Sentence Recall**

Finally, Figure 2C gives a measure of sentence recall (i.e., memory) for overlearned sentences versus novel sentences following overlearning. Participants recalled overlearned sentences heard during scanning with significantly greater accuracy than novel sentences heard during scanning ( $t(11) = 6.54$ ,  $p < .001$ ). In fact, half of the participants recalled the overlearned sentences verbatim. By comparison, none of the participants recalled the novel sentences verbatim even though they had just heard them during scanning.

### *fMRI*

#### **Novel LMM**

We expected that learning would be specific to overlearned sentences. To help determine if this was the case, we conducted a novel sentence listening linear mixed-effects model (LMM) with session (one and two) and timepoint (0-25) as factors. There were no discernible effects of session and no session by time interaction. Nonetheless, we used general linear tests (GLTs) to directly contrast novel sentences for sessions one and two at all timepoints in a manner used in subsequent analyses (see Figure 3). Collapsed over all 26 timepoints, there

were a small number of differences from session one to two (Figure S2). This included seven clusters with more than 20 voxels, with decreases in activity in the cerebellum (x/y/z = 31/-84/-41; 890 voxels), thalamus (x/y/z = -5/-15/7; 42 voxels) and lingual gyrus (x/y/z = -29/-48/-2; 25 voxels) and increases in the left dorsal postcentral gyrus (x/y/z = -41/-39/67; 657 voxels), superior frontal gyrus (x/y/z = -1/32/36; 161 voxels), right superior parietal lobule (x/y/z = 19/-57/70; 99 voxels) and right middle anterior cingulate gyrus (x/y/z = 4/18/25; 20 voxels). We calculated the GLTs for novel sentences for sessions one and two independently and used these for analysis of the impulse response function (described in the next section; see Figure 4). We also calculated the GLTs for novel sentences, collapsing over sessions one and two and all 26 timepoints and used this as a guide to ‘language regions’ (see Figures 3 and 5).

### Overlearning LMM

To test the hypothesis that sentence listening after overlearning involves more sensorimotor and less activity in ‘language regions’, a LMM was done with session (one and two), sentence (overlearning sentences one and two) and timepoint (0-25) as factors. There were main effects of session and time that encompassed most of the brain at a corrected threshold. In contrast, the main effect of sentence at  $p < .01$  corrected resulted in only four clusters in the primary visual cortex (x/y/z = -11/-102/4; 615 voxels), middle occipital cortex around motion area MT+ (x/y/z = -47/-72/7; 123 voxels), superior parietal cortex (x/y/z = 28/-51/61; 121 voxels) and middle frontal gyrus (x/y/z = 43/33/34; 117 voxels). Given this relatively small amount of activity and that these regions are mostly centred around visual cortices, suggesting minor differences in sensory semantic properties of the overlearned sentences (recall that there were three sets of sentences), we collapsed over sentences one and two for all subsequent analyses.

Compared to baseline and independent of time, GLTs for sentences in sessions one and two both involve processing in the superior temporal plane. However there was an increase in processing in sensorimotor cortices and a large reduction in the spread of activity in the superior temporal plane in session two (Figure S1A and S1B). A direct contrast reveals that most of the brain differs at corrected threshold and this confirms a bilateral increase in sensorimotor regions and a decrease in the superior temporal plane (Figure S1C). Subcortically, the hippocampus, caudate and dorsal cerebellum increase while the thalamus, putamen and ventral cerebellum decrease from sessions one to two. These results are present even at voxel-wise corrected threshold of  $p < 1 \times 10^{-10}$ .

There was an interaction between session and time for overlearned sentences that involved many of these same regions, suggesting that the timing of activity also changes. To visualize changes, the GLTs for the contrast of sessions one and two at each timepoint are presented (Figure 3). To make visualisation easier, we created four time bins and illustrated approximately when activity occurs with regard to a ‘canonical’ hemodynamic response function (Figure 3, middle line). The previously described results can be seen, with an increase in sensorimotor and a decrease in superior temporal plane regions in session two. Also apparent is that processing began earlier in sensorimotor regions in session two (Figure 3, first column). This includes activity in primary motor and somatosensory cortex in the central sulcus (x/y/z = -44/-24/55; 376 voxels) and the supplementary motor area (x/y/z = 4/21/49; 40 voxels) and no subcortical structures. Differences in subcortical activity began in the second time bin and is as described above (i.e., an increase in activity in the hippocampus, caudate and dorsal cerebellum and a decrease in the thalamus, putamen and ventral

cerebellum). Using the separately conducted novel sentence listening and production LMMs as a guide, it appears that activity for overlearned sentences is increased in sensorimotor regions involved in producing speech (Figure 3, white outline) and reduced in regions active during novel sentence listening (Figure 3, dark grey).

---

### Figure 3

---

Next, we further explored overlearning LMM results by visualising session one and two timecourses for overlearned sentences. This allows us to better understand if responses for overlearned sentences are a simple redistribution or a reorganization of brain responses. Specifically, we thresholded the previously described GLT contrasting overlearned sentences between sessions one and two (see Figure S1C) at an arbitrary high z-value of 10. This resulted in 19 small clusters and we selected 15 of these for presentation, five frontal or parietal, five superior temporal and five subcortical regions. We averaged and plotted the coefficients in each of these regions for all timepoints. For comparison, we also plot timepoints for the separately conducted novel sentence listening LMM (see prior section). Overlearned sentence responses in frontal and parietal, including sensorimotor regions, all showed a new response in session two from something resembling a below baseline or lack of activity in session one (Figure 4, left). In contrast, overlearned sentence responses in superior temporal regions all showed a reduction in, lack of or below baseline response in session two from a state of heightened activity in session one (Figure 4, middle). Finally, subcortical regions showed a similar set of patterns, with an increase in the caudate and a decrease in the cerebellum and putamen for overlearned sentences from session one to two (Figure 4, right).

---

### Figure 4

---

#### Overlearning-**Novel LMM**

To more directly test the hypothesis that sentence processing after overlearning involves more activity in sensorimotor regions and less activity in ‘language regions’ compared to typical sentence processing, we did the same LMM as in the prior section but after first subtracting the coefficients for novel from overlearned sentences. We present the effects of each session separately because of the difficulty of interpreting the sign for this more complicated comparison. GLTs for overlearned minus novel sentences in both sessions, independent of time, show that overlearned sentences resulted in greater activity in sensorimotor regions. In contrast, novel sentences produce more activity in the superior temporal plane and inferior frontal gyrus in both sessions (Figure S3A and S3B). Subcortically, in session one, overlearned sentences produced more brainstem, nucleus accumbens and dorsal cerebellar activity while novel sentences produced more hippocampal activity. In session two, overlearned sentences produced more thalamus and more dorsal cerebellum activity while novel sentences produced more ventral cerebellar activity.

There were interactions between session and time for overlearned minus novel sentences that involved much of the brain though we do not attempt to interpret these here (though Figures 3 and 5 together suggest the direction of these effects). Nonetheless, to visualise changes over

time as is Figure 3, the GLTs for each time point are presented separately for overlearned minus novel sentences in session one (Figure 5A) and session two (Figure 5B). The pattern of results clearly shows more sensorimotor activity for overlearned sentences in both sessions (Figure 5A and 5B, reds). The speech production LMM shows that this occurs in similar sensorimotor regions used to produce novel sentences. Conversely, there was less superior temporal plane and inferior frontal gyrus activity for overlearned sentences (Figure 5A and 5B, blues). These regions closely overlapped those for novel sentences alone (Figure 5, dark grey).

---

## Figure 5

---

### Regional Homogeneity

The deconvolution and subsequent LMM results are model based, aggregating over stimuli, though without a priori assumptions about the shape of the hemodynamic response. To test whether there is support for hypotheses in a more model-free manner and to test whether there is more local sensorimotor connectivity after overlearning, we conducted a regional homogeneity analysis on preprocessed timeseries. There was an increase in sensorimotor cortices and a decrease in superior temporal plane local connectivity from session one to two for overlearned sentences (among other regions; Figure 6, top). We then directly contrasted overlearned and novel sentences for sessions one and two separately (Figure 6, bottom). In session one, there were few regions more active for overlearned sentences whereas novel sentences resulted in significantly greater regional connectivity, mostly throughout superior temporal plane and inferior parietal regions, bilaterally. In contrast, the overlearned sentences produced greater mostly inferior parietal and bilateral sensorimotor local connectivity in session two and the local superior temporal plane connectivity remained greater for novel sentences (Figure 6, bottom).

---

## Figure 6

---

### Network

To further test the hypothesis that the brain reorganises after overlearning, we analysed the global network variation between session one and two for overlearned sentences using edit distance. The average change in connectivity was 45.4% (SD = 5.50%; Minimum = 32.26%; Maximum = 55.37%), significantly different from the null hypothesis that there was no change from session one to two ( $t(11) = 28.69$ ,  $p = 1.08 \times 10^{-11}$ ). To visualise some of these changes in connectivity, we did chi-square ( $\chi^2$ ) tests on the binarised connections, using a threshold of 6.63 ( $p < .01$ ; Figure 7). This resulted in 90 changes in connections, with 25 connections gained (27.77%) and 65 connections lost (72.22%). This large reduction in connections was across all major brain subdivisions. Nonetheless, a full 80% of the changes included medial / midline (42) and/or subcortical structures (34, with four overlapping medial-subcortical connections). Of the 42 medial regions, 27 were lost connections (64.29%). Of the 34 subcortical regions, 28 were lost connections (82.35%). New subcortical connections involved the amygdala (1x), cerebellum (1x), pallidum (1x), hippocampus (2x)



and nucleus accumbens (2x). Lost subcortical connectivity involved the caudate (1x), amygdala (2x), cerebellum (2x), hippocampus (2x), pallidum (2x), brainstem (4x), putamen (5x), ventral diencephalon (6x) and thalamus (6x). At a more stringent threshold ( $p < .005$ ), the disproportionate number of lost connections remains similar at 75.93%, with a disproportionate number of lost connections containing medial and subcortical connections (85.19%; see Table S1 for more specific information about changes in connectivity).

## Figure 7

To characterise global connectivity changes with a more robust method, Deltacon similarity was computed. The average dissimilarity was 46.1%, (SD = 3.90%; Minimum = 43.20%; Maximum = 52.33%), with a significant change between sessions one and two ( $t(11) = 41.43$ ,  $p = 1.97 \times 10^{-13}$ ). To try and further understand network differences between sessions, we explored a number of other global and local measures. None of density, efficiency, diffusion or flow were significantly different from sessions one and two ( $ps > .05$ ). There were also no differences between measures of centrality or community structure ( $ps > .05$ ).

## Discussion

### *Summary*

We had participants overlearn sentences (Figure 1; Table 1) in an effort to evaluate the hypothesis that, as speech becomes more formulaic, it is processed more by sensorimotor cortices and not ‘language regions’. Behavioural results suggest that participants overlearned sentences. Reaction times to complete the final word of overlearned sentences decreased by 2.1 seconds (Figure 2A), participants' identification of words drawn from overlearned sentences was faster (Figure 2B) and overlearned sentences were remembered with high accuracy (Figure 2C).

Correspondingly, there was a considerable and local increase after overlearning in sensorimotor regions centred around the central sulcus, including primary motor and somatosensory cortices in both model based and model-free analyses (Figures 3, 5 and 6, oranges). These regions overlapped with those involved in producing novel sentences (Figures 3, 5 and 6, white outline). After overlearning, speech perception begins earlier in sensorimotor regions, before the hemodynamic response is typically expected to rise (Figures 3 and 5, left) and before superior temporal regions (compare Figure 4, left and middle onsets). Consistent with reorganisation, some of these motor regions had a hemodynamic response that was centred around zero before overlearning (Figure 4, left).

Concomitant with these increases, we found a dramatic bilateral reduction in activity and local processing in the superior temporal plane when comparing overlearned sentences from sessions one to two (Figures 3, blues; Figure 6, blues, top). When overlearned sentences are compared to novel sentences, there was a reduction in activity and local processing in superior temporal plane and the inferior frontal gyrus ‘language regions’ (Figure 5, blues and grey; Figure 6, blues, bottom). Consistent with reorganisation, some superior temporal plane regions had a hemodynamic response around zero after overlearning, suggesting they were not active on average (Figure 4, middle).

Global network measures indicate that about 45% of the brain reorganised after overlearning. Much of this reorganisation involves a small number of new connections and a profound drop in the number of connections across the whole brain (Figure 7). The latter most saliently involved medial cortical and subcortical brain regions. Lost subcortical connections involved the amygdala, basal ganglia (mostly the putamen but also the caudate), brainstem, cerebellum, hippocampus, thalamus and ventral diencephalon. These subcortical results parallel those in voxel-based analyses that showed less increases and more decreases in activity (e.g., Figure 4, right).

## **Overview**

Results suggest that the brain mechanisms associated with perceiving overlearned speech are qualitatively different from less formulaic or more compositional language. They suggest that when speech segments become sufficiently overlearned, they are processed by a much more circumscribed and cortically isolated set of sensorimotor regions involved in producing speech and little, if at all, by ‘language regions’. As reviewed in the Introduction, the hypothesis predicting these results was predicated on 1) the relative preservation of both sensorimotor regions and formulaic language in aphasia and Alzheimer’s disease<sup>45–49,63,65,66,75</sup> and 2) predictive models of the role of speech production related regions in speech perception<sup>12,79,101</sup>. Both suggest an increased role of sensorimotor and a decreased role of ‘language regions’ as learning increases. We review studies of formulaic language and motor sequence learning that mostly supports this proposal. Finally, we discuss implications for neurobiological models of language and therapy.

## **Formulaic Language**

### **Production**

It has been proposed that formulaic language production is supported by right hemisphere and subcortical interactions<sup>45</sup>. This conclusion was based on research showing that formulaic expressions are more common in left compared to right hemispheric damage<sup>51–54</sup> (though see<sup>49,133</sup>). It was additionally based on results suggesting that individuals with Alzheimer’s disease produce more formulaic language than people with basal ganglia strokes<sup>54,134</sup> and Parkinson’s disease<sup>66,74</sup>. This is attributed to relative preservation of the basal ganglia in Alzheimer’s disease. A region of interest based neuroimaging study in healthy people supports both arguments, showing that increased formulaic language production is correlated with increased right inferior frontal gyrus and decreased left caudate activity<sup>135</sup>.

Though we show some similar right hemisphere and subcortical effects, our results are more bilateral and centred more around sensorimotor rather than subcortical regions. To the extent our results can generalise to formulaic language production, they suggest that the postulated right hemisphere locus is due to a stronger weighting of right hemisphere regions after aphasia. Our results also suggest that the presumed subcortical locus might be less about the preservation or deterioration of the basal ganglia in Alzheimer’s and Parkinson’s disease and more about preserved sensorimotor regions in the former. This is more consistent with research suggesting that the basal ganglia is significantly impacted in Alzheimer’s disease, even in early stages<sup>136–138</sup>. Collectively, our results suggest that the more preserved right sensorimotor and not necessarily right ‘language regions’ or subcortical structures are the locus of formulaic speech production in aphasia and Alzheimer’s disease.

## **Comprehension**

Our results are more consistent with neuroimaging studies of formulaic related language comprehension. Much of this work centres around figurative language, like idioms and metaphors. These are processed bilaterally in ‘language regions’, with varying contribution of the left and right hemisphere as a function of, among other things, familiarity<sup>139–145</sup>. Furthermore, the more familiar<sup>146–152</sup>, frequent<sup>153–157</sup>, compositional (‘red boat’ vs ‘cup boat’) or coherent multiword expressions are<sup>158,159</sup>, the less ‘language regions’ tend to be active. Some studies show that sensorimotor activity is more strongly associated with high frequency words, word composition and impairments of word composition<sup>160,161</sup> while the basal ganglia are less active for higher frequency words<sup>157,162</sup>. One study concluded that ‘areas canonically implicated in traditional neurophysiological models of language processing appear to play a lesser role in basic composition’ (p. 2802)<sup>160</sup>. Another concluded that multiword expressions rely on regions other than ‘traditional frontal and temporal nodes of the language network’ (p. 12)<sup>158</sup>.

## ***Sequence Learning***

The strongest similarities to our results derive from motor sequence learning research. This work shows that becoming or being an expert motor performer involves a well documented set of decreases and increases in brain activity that depend on the length of learning<sup>163–173</sup>. Specifically, there are two learning stages associated with relative duration: fast online learning, described as more explicit and by repetition suppression and slow learning, described as more implicit and by repetition enhancement and sleep related or offline consolidation. Fast learning is frequently linked to more associative cortical and subcortical regions. In contrast, slow learning is associated with a global decrease in activity in most regions, including prefrontal, premotor, parietal, sensorimotor and subcortical structures like more associative basal ganglia and cerebellar regions. Amongst these widespread decreases, there is a selective increase in some sensorimotor and subcortical regions, including less associative aspects of the basal ganglia and cerebellum. Generally, these changes might be described as a shift away from cognitive systems (involving attentional, inhibition, control, etc.) and towards more ‘automatic’ sensorimotor brain regions.

Our study incorporated early and late learning phases, with an early online perceptual learning period (measured in fMRI session one) and a late offline period of production based learning with motor memory consolidation (measured in fMRI session two). Consistent with slow learning, we show global decreases in activity in most cortical and subcortical regions with a selective focal increase in some sensorimotor and subcortical regions. Also consistent with the distinction between fast and slow learning, our results do not simply constitute a redistribution of activity patterns in the same ‘language regions’ but, rather, are a reorganization to sensorimotor regions, suggesting different processes underlying the perception of novel and overlearned sentences<sup>165</sup>.

## ***Neurobiological Models***

More generally, our results suggest a model by which the perception of overlearned speech may not rely on ‘language regions’ or ‘the language network’. Like test-retest reliability and lexical processing studies (discussed in the Introduction), our results suggest that the

neurobiology of speech perception and language comprehension is more variable and distributed than posited by classical and contemporary models<sup>1-11</sup>.

If so, an open question becomes what differentiates the more distributed regions found in some studies from the fixed set of regions in popular models? One hypothesis is that the whole brain variously participates in speech perception and language comprehension and that 'language regions' as we know them are mostly connectivity hubs coordinating this more distributed system<sup>12</sup>. Because of the reliance on measures of central tendency, these distributed regions are averaged out in most studies<sup>12</sup>. They only become obvious when specific categories are under examination, e.g., individual differences, action words or, as here, formulaic language.

Indeed, most 'language regions' are structural and functional connectivity hubs in humans<sup>174-177</sup>, as are homologous regions in monkeys<sup>178-181</sup>. As befitting connectivity hubs, 'language regions' are some of the most 'reused' regions in the human brain, participating in many perceptual, motor and cognitive domains beyond language<sup>182</sup>. Given how central and diverse they are, it is perhaps not surprising that damage to these hubs predict neuropsychological outcomes beyond language problems<sup>183,184</sup>.

If they are mostly hubs, the focus on 'language regions' or their 'homologues' in therapeutic interventions runs the risk of overemphasizing the importance of less specific regions and neglecting more behaviourally relevant network nodes. For example, given the preservation of formulaic language and corresponding sensorimotor regions in some individuals with aphasia and Alzheimer's disease, it makes sense to focus on those expressions and regions as they might be used as a scaffold for language recovery. Indeed, use of formulaic song and language shows promise in therapy<sup>185-190</sup>. In contrast, the relative preservation of some other aspect of language and associated distributed set of network nodes might benefit from a different intervention.

To the extent models guide therapy, this proposal requires us to move beyond current neurobiological models with static regions<sup>2,12,79,191,192</sup>. It begs for a more detailed neurobiological account of overlearned expressions and other factors that result in differently distributed language processes. It also suggests a greater focus on item analysis and individual differences and a less strong reliance on measures of central tendency to understand the organisation of language and the brain<sup>193</sup>.

### ***Limitations***

Overlearning often occurs through repetition in short sessions over a relatively small period of time (e.g., as in learning at home over a period of weeks for school tests). However, much of formulaic language is picked up over a lifetime of repetitions. Similarly, though many overlearned expressions are not particularly meaningful (e.g., 'whats up'), most do have more specific meanings. Our repetition learning task did not unfold over years nor did it emphasise semantic content per se. Thus, the neurobiology of more extensive overlearning, with more semantically meaningful content might differ from what was observed here.

As with any longitudinal design, it might be argued that participants were not attending to overlearned stimuli or attending as much in session two. Participants were monitored for wakefulness during scanning and all appeared to be attending in both sessions. Post scan recall

indicates that participants paid attention enough to recall novel sentences with relatively high accuracy.

## ***Conclusions***

Results suggest that the brain regions supporting speech perception are not fixed but, rather, dramatically reorganise as a function of individual experience with speech production. Specifically, repeated experience speaking the same word sequence seems to change the memory representation of those words to be more formulaic. This trace is subsequently used by the brain in the process of speech perception, perhaps in a predictive manner. This involves a different set of regions than is said to support more compositional language. Given how frequently formulaic expressions occur, how fast they are processed and important they are in learning, results call for more research on how overlearning occurs and formulaic speech is processed. They also suggest why formulaic expressions might be preserved in aphasia and Alzheimer's disease. Therapy for such disorders, if they are neurobiologically informed, are presumably informed by prevailing beliefs. One of these is the (often implicit) belief that language is supported by a static set of 'language regions' or 'the language network', formalised in classical and contemporary neurobiological models. Given that this is not the case, as shown here and elsewhere, future interventions might benefit from adopting more dynamic and distributed network models of language and the brain<sup>12</sup>.

## **Acknowledgments**

JIS would like to thank the fine people at the Birkbeck-UCL Centre for Neuroimaging (BUCNI) and acknowledge support from EPSRC EP/M026965/1 'New pathways to hearing: A multisensory noise reducing and palate based sensory substitution device for speech perception.' JIS would like to thank Marple and BB. Hair are your aerials. DRL would like to acknowledge support from the British Academy and the Natural Sciences and Engineering Research Council of Canada.

## **Author Contributions**

DL and JIS conceived of the study. SB, DL, EM and JIS designed the study. All authors helped with stimulus pretesting and EM selected overlearned sentences. JIS wrote the manuscript with help from SA, SB, YJJ and DL. EM, SB, SL and YJJ collected the data. DL did behavioural analyses. SL and YJJ did the ICA artifact labelling for fMRI preprocessing, checked by JIS. JIS did all fMRI preprocessing and analyses except the network analyses, done by SA.

## **Competing Interests**

The authors declare no competing financial interests.

## **Data Availability**

All fMRI data will be made available on OpenNeuro (<https://openneuro.org/>). Scripts used to analyse data will be made available on GitHub (<https://github.com/lab-lab>). Images used to make all figures will be made available on NeuroVault (<https://neurovault.org/>).

## Tables

**Table 1.** Overlearned sentence information. All sentences were 2.5 seconds long and the final word was 500 ms. Cloze probability from Block and Baldwin (2010)<sup>107</sup>. Complexity and word frequency are from Subtlex-UK<sup>108</sup>.

Set	Sentence	Words	Cloze Probability	Complexity	Average Frequency
1	Instead of dressing, I prefer vinegar and oil.	8	0.76	1.2	7716.87
	At the pub, he ordered another mug of beer.	9	0.87	1.2	8645.02
2	She was tired of her life and felt ready for a change.	12	0.83	1.3	7134.33
	The jury found him innocent and set him free.	9	0.85	1.3	7996.94
3	They went to the video store to rent a movie.	10	0.54	1.4	13093.97
	She followed the recipe correctly to cook the meal.	9	0.56	1.4	13651.17

## Figure Captions

**Figure 1.** Study overview. We conducted functional magnetic resonance imaging (fMRI) over two sessions. In both sessions, participants ‘passively’ listened to two sentences repeated 60 times each and 60 novel sentences. In a final run, participants produced the 60 novel sentences. The sessions were separated by 15 days. During this time period, participants trained at home by producing the two repeated sentences from fMRI 30 times each, twice a day. To assess learning, participants performed sentence completion, lexical decision and sentence recall tasks before and after training.

**Figure 2.** Behavioural measures of overlearning. A) For each testing session (1-4), sentence completion is the length of time (in ms) that it took participants to type in the final word of overlearned (red line) and novel sentences (blue line). Completion times decreased following overlearning. B) Lexical decisions (word / not word) were made for words drawn from the overlearned sentences (red bar), similar words (blue bar), dissimilar words (grey bar) and reversed words (black bar). Change in lexical decision time reflects decision times (in ms) measured in sessions one and two (pre overlearning) versus decision times measured in sessions three and four (post overlearning). C) Participants recalled sentences they heard during scanning. The accuracy of sentence recall was measured as the cosine similarity between recalled and heard sentences. Participants were more likely to recall sentences they had overlearned (red bar) versus novel sentences (blue bar).

**Figure 3.** Direct contrast of overlearned sentence listening from session one and two across time. General linear tests (GLTs) between session one and two overlearned sentences were done at each of 26 timepoints in the estimated impulse response functions following a linear mixed-effects model (LMM). For ease of visualisation, these contrasts are collapsed into four time bins between the indicated seconds. The orange line separating the left (LH) and right (RH) hemisphere surfaces roughly represents the relative position in a ‘canonical’ hemodynamic response function. Listening to sentences after overlearning results in an increase in activity in sensorimotor regions centred around the central sulcus (among other regions, reds) and a reduction of activity in the superior temporal plane (blues). Increased sensorimotor activity overlaps with producing novel sentences as determined by a separate LMM (white outline). Note that the sensorimotor regions are first engaged after overlearning in the delay period before the canonical response rises. Decreased superior temporal plane activity overlaps with listening to novel sentences as determined by another LMM (dark grey, presented on only the far left column though it represents all 26 timepoints). Each timepoint was cluster size corrected for multiple comparisons at alpha ( $\alpha$ ) < .01. The colour bar represents z-scores.

**Figure 4.** Estimated impulse response functions for overlearned and novel sentence listening in different regions. The direct contrast of overlearned sentences from sessions one and two was thresholded at  $z = 10$  ( $p < 1.52 \times 10^{-23}$ ), resulting in 19 clusters of activity. Fifteen of these, five each from frontal / parietal, temporal and subcortical regions were selected for display. The impulse response function from the voxels in each of these clusters from the overlearning linear mixed-effects model (LMM) were averaged and plotted for each session. Also plotted are the averaged novel sentence listening LMM impulse response functions for comparison. Note that in most cases, the change in the impulse response functions before or after learning are from a state of below baseline or inactivity. Abbreviations: CS = central

sulcus; LH = left hemisphere; PoCG = postcentral gyrus; POLE = temporal pole; PreCG = precentral gyrus; PTr = pars triangularis of the inferior frontal gyrus; RH = right hemisphere; SMA = supplementary motor area; STSa = anterior superior temporal sulcus; and STSp = posterior superior temporal sulcus.

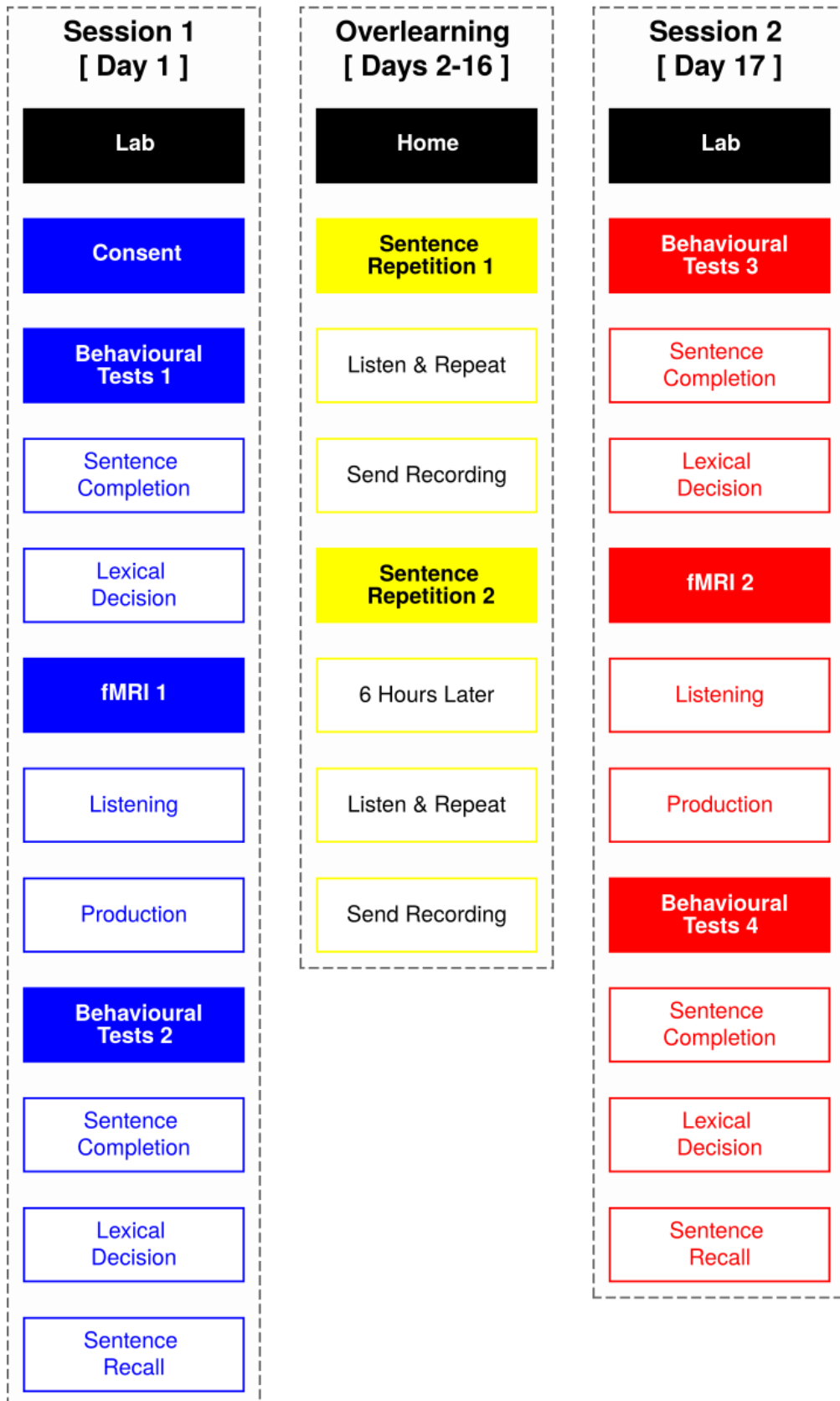
**Figure 5.** Direct contrast of overlearned minus novel sentence listening for session one and two across time. General linear tests (GLTs) for session one (top) and two (bottom) for the subtraction of novel from overlearned sentences, done at each of 26 timepoints in the estimated impulse response functions. In both sessions, listening to overlearned sentences results in significantly more activity in sensorimotor regions (reds) also involved in producing speech (white outline). Conversely, overlearned sentences result in less activity in the superior temporal plane and inferior frontal gyrus (blues), specifically in regions that are involved in processing novel sentences (dark grey, left hand column). Everything else is as in Figure 3.

**Figure 6.** Local connectivity changes after overlearning. A regional homogeneity analysis was conducted to provide a more data driven validation of more model based results (shown in Figures 3-5) and to estimate changes in local synchronization or connectivity. In particular, we directly contrasted overlearned sentences between sessions one (blues) and two (reds; top row). We also contrasted novel (blues) and overlearned sentences (reds) for sessions one (middle row) and two (bottom row). Results confirm that listening to overlearned sentences after learning results in significantly more local connectivity in sensorimotor regions (reds) and significantly less local connectivity in the superior temporal plane.

**Figure 7.** Changes in network connectivity after overlearning. The chord plot shows all 90 lost (blue) and gained (red) connections between sessions one and two as determined by chi-square tests on binarised connections between pairs of 167 regions of interest ( $\chi^2 > 6.63$ ;  $p < .01$ ). For ease of visualisation, these were grouped into frontal (Fr), insula (In), medial, occipital (Oc), parietal (Pa), sensorimotor (SM), subcortical (SC) and temporal (Te) regions. Results suggest that sentence overlearning results in significantly less connectivity, particularly in medial cortical and subcortical regions.



**Figure 1**



**Figure 2**

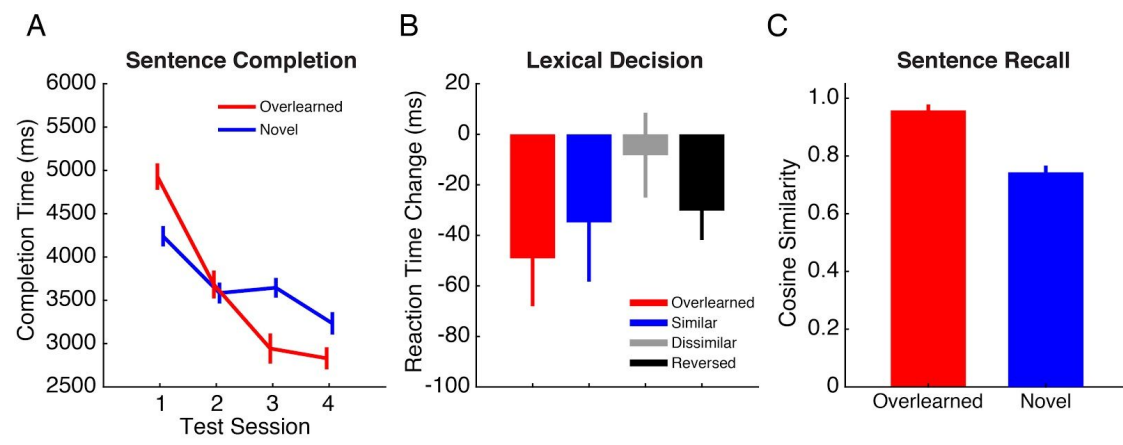
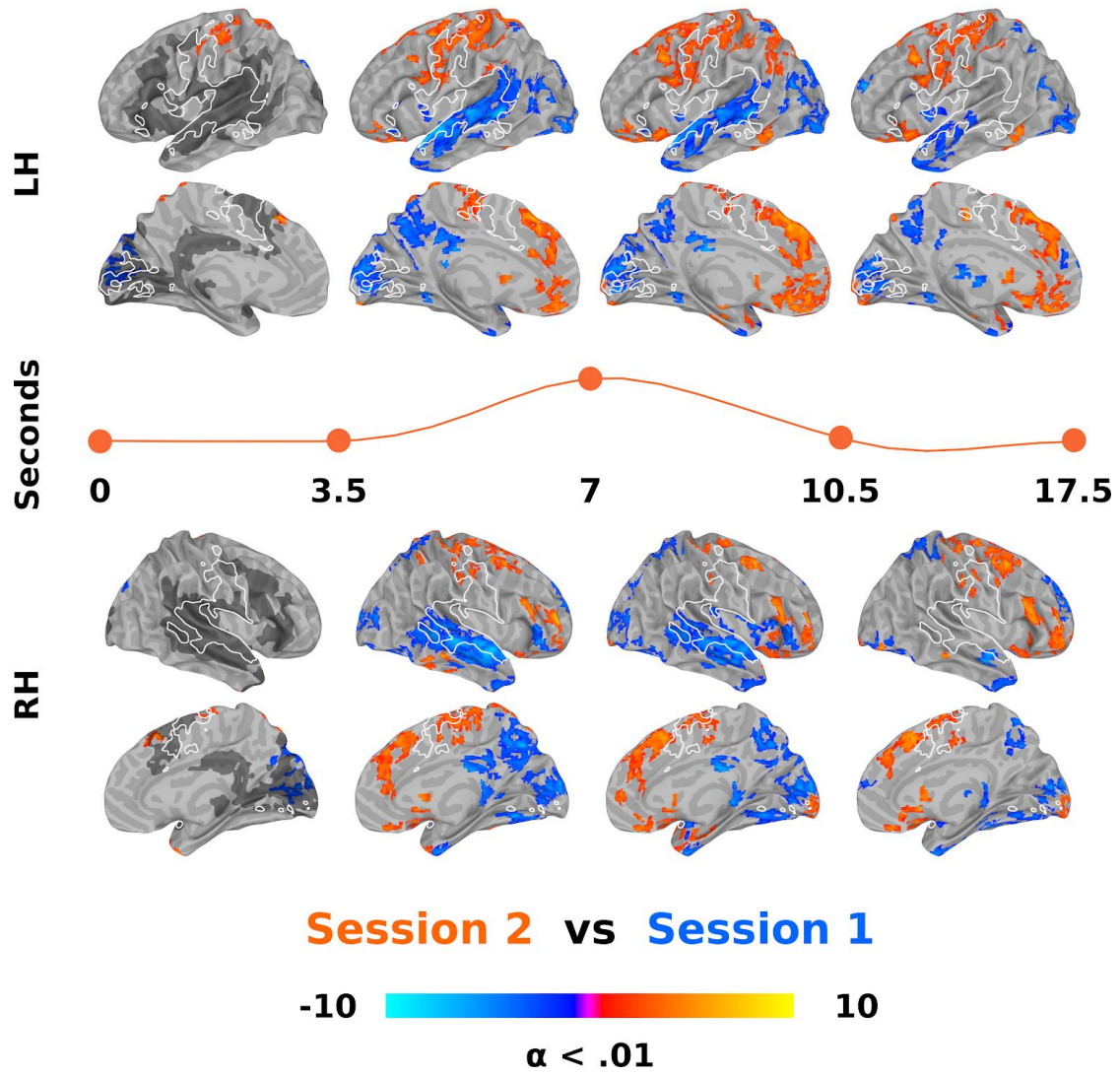


Figure 3

## Overlearned Sentences



**Figure 4**

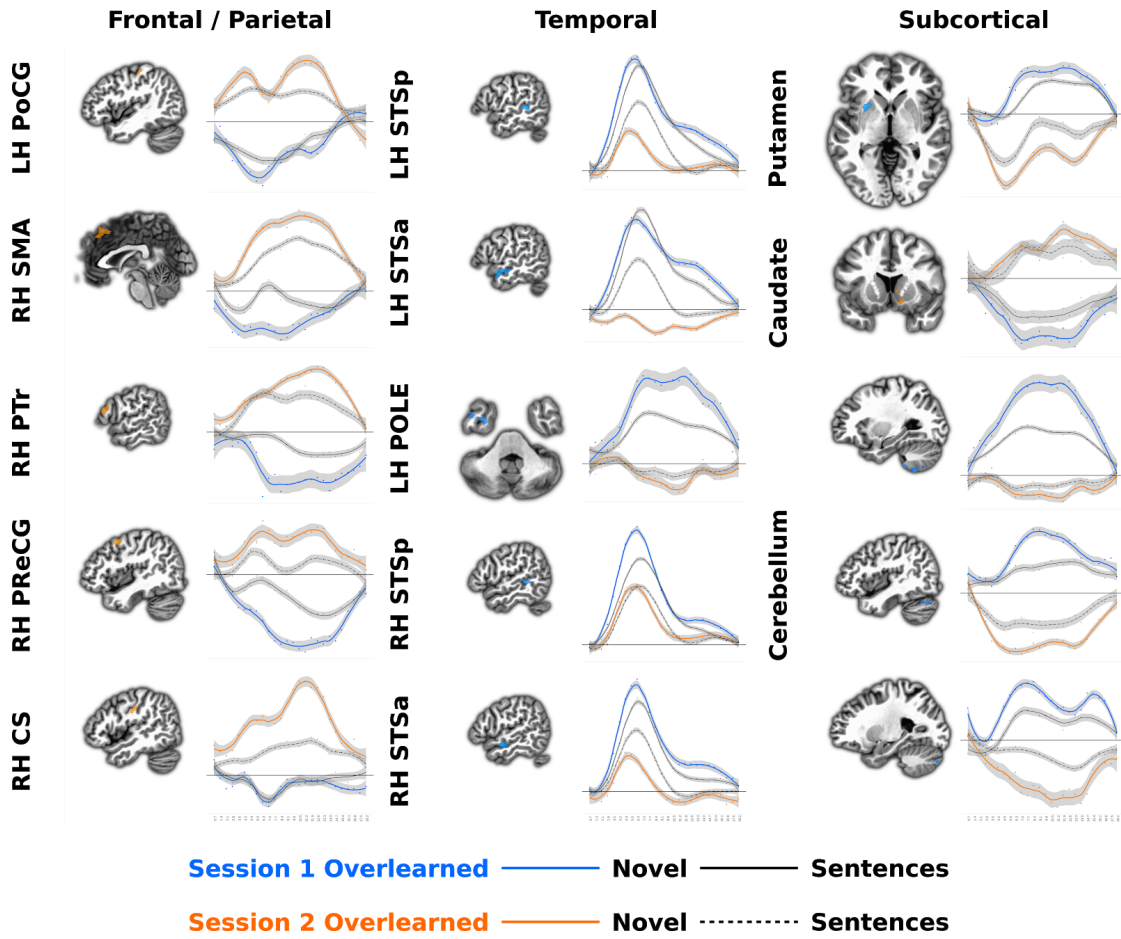


Figure 5

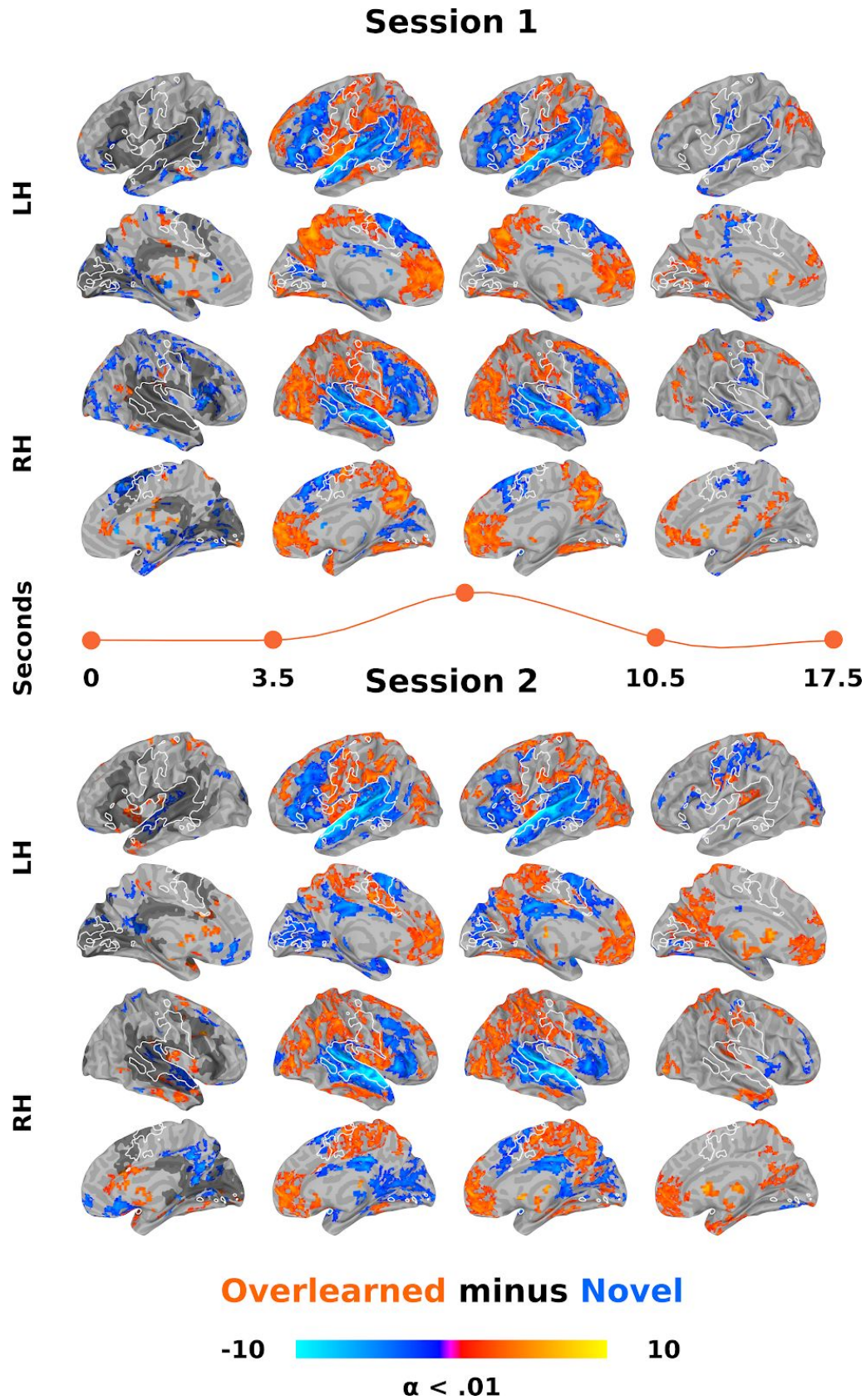
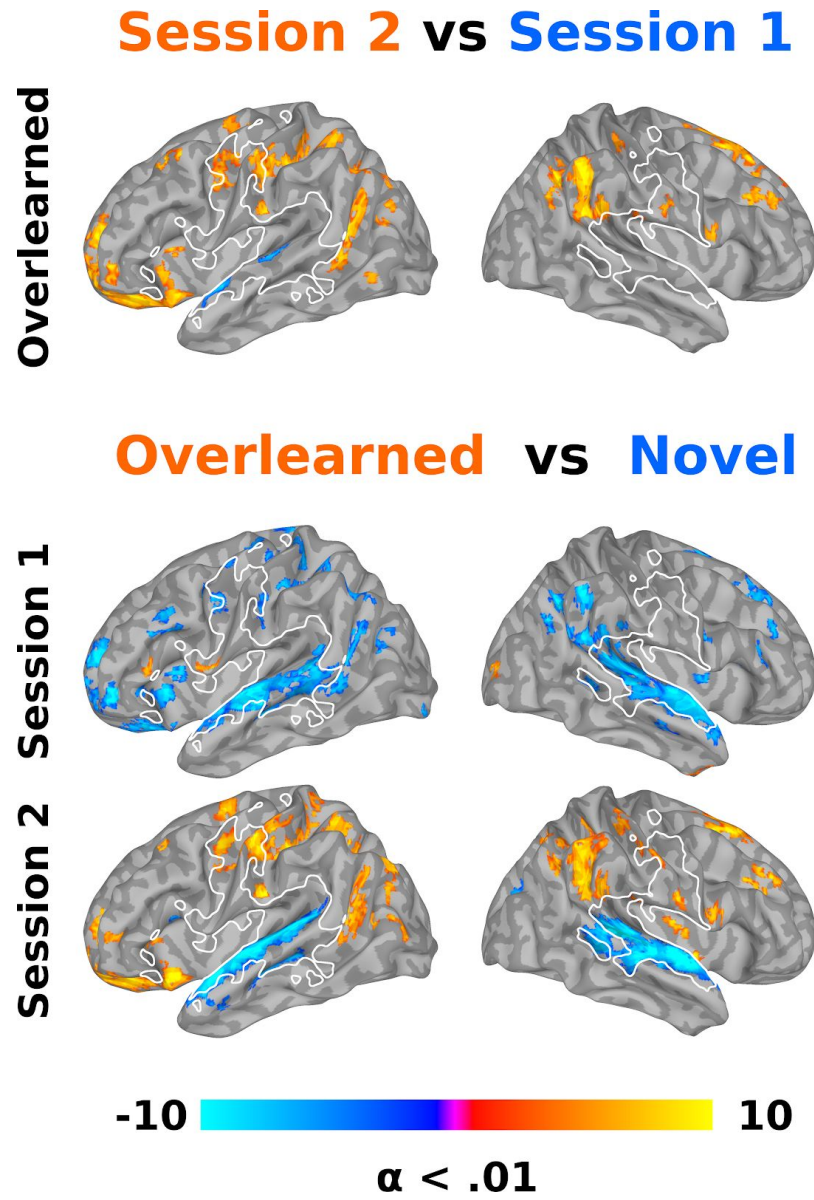
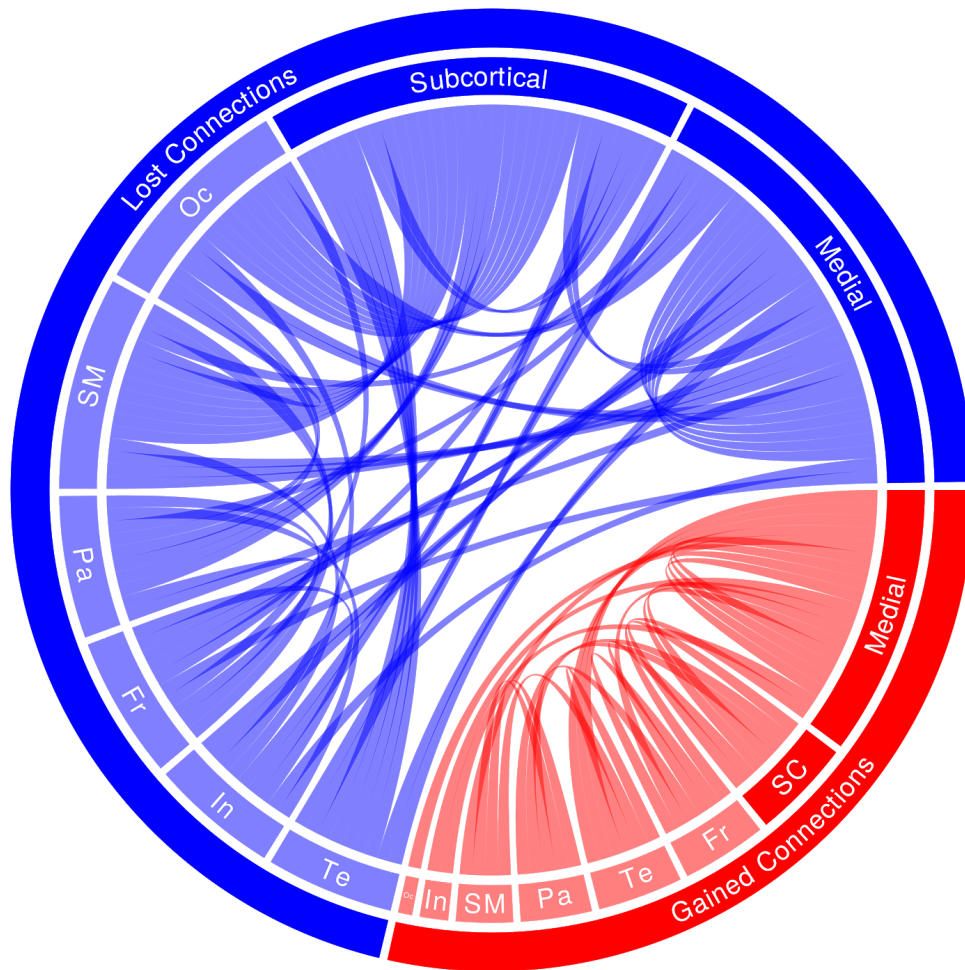


Figure 6



**Figure 7**



## References

1. Geschwind, N. The organization of language and the brain. *Science* **170**, 940–944 (1970).
2. Tremblay, P. & Dick, A. S. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain Lang.* **162**, 60–71 (2016).
3. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
4. Hickok, G. & Poeppel, D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99 (2004).
5. Hickok, G. & Poeppel, D. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* **4**, 131–138 (2000).
6. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).
7. Hickok, G. Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* **13**, 135–145 (2012).
8. Friederici, A. D. & Gierhan, S. M. E. The language network. *Curr. Opin. Neurobiol.* **23**, 250–254 (2013).
9. Fedorenko, E. & Thompson-Schill, S. L. Reworking the language network. *Trends Cogn. Sci.* **18**, 120–126 (2014).
10. Fedorenko, E. The role of domain-general cognitive control in language comprehension. *Front. Psychol.* **5**, 335 (2014).
11. Hickok, G., Houde, J. & Rong, F. Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron* **69**, 407–422 (2011).
12. Skipper, J. I. The NOLB model: a model of the natural organization of language and the brain. in *Cognitive Neuroscience of Natural Language Use* (eds. Willems, R. M. & Willems, R. M.) 101–134 (Cambridge University Press, 2015).
13. Pulvermüller, F. Neural reuse of action perception circuits for language, concepts and



- communication. *Prog. Neurobiol.* **160**, 1–44 (2018).
14. Pulvermüller, F. How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.* **17**, 458–470 (2013).
  15. Frost, M. A. & Goebel, R. Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* **59**, 1369–1381 (2012).
  16. Wilson, S. M., Bautista, A., Yen, M., Lauderdale, S. & Eriksson, D. K. Validity and reliability of four language mapping paradigms. *NeuroImage: Clinical* (2016) doi:10.1016/j.nicl.2016.03.015.
  17. Burton, M. W., Noll, D. C. & Small, S. L. The Anatomy of Auditory Word Processing: Individual Variability. *Brain Lang.* **77**, 119–131 (2001).
  18. Voyvodic, J. T. Reproducibility of single-subject fMRI language mapping with AMPLE normalization. *J. Magn. Reson. Imaging* **36**, 569–580 (2012).
  19. Miller, M. B. *et al.* Unique and persistent individual patterns of brain activity across different memory retrieval tasks. *Neuroimage* **48**, 625–635 (2009).
  20. Seitzman, B. A. *et al.* Trait-like variants in human functional brain networks. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22851–22861 (2019).
  21. Miller, M. B., Donovan, C.-L., Bennett, C. M., Aminoff, E. M. & Mayer, R. E. Individual differences in cognitive style and strategy predict similarities in the patterns of brain activity between individuals. *Neuroimage* **59**, 83–93 (2012).
  22. Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J. & Hoenig, K. The Sound of Concepts: Four Markers for a Link between Auditory and Conceptual Brain Systems. *Journal of Neuroscience* **28**, 12224–12230 (2008).
  23. Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L. & Ungerleider, L. G. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* **270**, 102–105 (1995).

24. Hauk, O., Johnsrude, I. & Pulvermüller, F. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* **41**, 301–307 (2004).
25. González, J. *et al.* Reading cinnamon activates olfactory brain regions. *Neuroimage* **32**, 906–912 (2006).
26. García, A. M. *et al.* How meaning unfolds in neural time: Embodied reactivations can precede multimodal semantic effects during language processing. *Neuroimage* **197**, 439–449 (2019).
27. Shtyrov, Y., Butorina, A., Nikolaeva, A. & Stroganova, T. Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1918–23 (2014).
28. MacGregor, L. J., Pulvermüller, F., van Casteren, M. & Shtyrov, Y. Ultra-rapid access to words in the brain. *Nat. Commun.* **3**, 711 (2012).
29. Wray, A. & Perkins, M. R. The functions of formulaic language: An integrated model. *Lang. Commun.* **20**, 1–28 (2000).
30. Constant, M. *et al.* Multiword Expression Processing: A Survey. *Comput. Linguist.* **43**, 837–892 (2017).
31. Van Lancker, D. & Cummings, J. L. Expletives: neurolinguistic and neurobehavioral perspectives on swearing. *Brain Res. Brain Res. Rev.* **31**, 83–104 (1999).
32. Liu, D. The Most Frequently Used Spoken American English Idioms: A Corpus Analysis and Its Implications. *TESOL Quarterly* **37**, 671 (2003).
33. N-grams data from the 14 billion word iWeb corpus. <https://www.ngrams.info/iweb.asp>.
34. Robinson, B. *Withnail & I.* (1987).
35. Bannard, C. & Lieven, E. Formulaic language in L1 acquisition. *Annu. Rev. Appl. Linguist.* **32**, 3 (2012).
36. Ellis, N. C. Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annu. Rev. Appl. Linguist.* **32**, 17 (2012).

37. Christiansen, M. H. & Arnon, I. More Than Words: The Role of Multiword Sequences in Language Learning and Use. *Top. Cogn. Sci.* **9**, 542–551 (2017).
38. Conklin, K. & Schmitt, N. The Processing of Formulaic Language. *Annu. Rev. Appl. Linguist.* **32**, 45–61 (2012).
39. Janssen, N. & Barber, H. A. Phrase frequency effects in language production. *PLoS One* **7**, e33202 (2012).
40. Arnon, I. & Snider, N. More than words: Frequency effects for multi-word phrases. *J. Mem. Lang.* **62**, 67–82 (2010).
41. Conklin, K. & Schmitt, N. Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Appl Linguist* **29**, 72–89 (2008).
42. Bannard, C. & Matthews, D. Stored word sequences in language learning: the effect of familiarity on children’s repetition of four-word combinations. *Psychol. Sci.* **19**, 241–248 (2008).
43. Tabossi, P., Fanari, R. & Wolf, K. Why are idioms recognized fast? *Mem. Cognit.* **37**, 529–540 (2009).
44. Rammell, C. S., Van Lancker Sidtis, D. & Pisoni, D. B. Perception of formulaic and novel expressions under acoustic degradation. *Ment. Lex.* **12**, 234–262 (2017).
45. Van Lancker Sidtis, D. & Sidtis, J. J. Cortical-subcortical production of formulaic language: A review of linguistic, brain disorder, and functional imaging studies leading to a production model. *Brain Cogn.* **126**, 53–64 (2018).
46. Van Lancker, D. & Klein, K. Preserved recognition of familiar personal names in global aphasia. *Brain Lang.* **39**, 511–529 (1990).
47. Bruns, C. *et al.* ‘I don’t know’: a usage-based approach to familiar collocations in non-fluent aphasia. *Aphasiology* **33**, 140–162 (2019).
48. Jaecks, P., Hielscher-Fastabend, M. & Stenneken, P. Diagnosing residual aphasia using

- spontaneous speech analysis. *Aphasiology* **26**, 953–970 (2012).
49. Zimmerer, V. C., Newman, L., Thomson, R., Coleman, M. & Varley, R. A. Automated analysis of language production in aphasia and right-hemisphere damage: frequency and collocation strength. *Aphasiology* **32**, 1267–1283 (2018).
  50. Code, C. Neurolinguistic analysis of recurrent utterance in aphasia. *Cortex* **18**, 141–152 (1982).
  51. Van Lancker Sidtis, D. & Postman, W. A. Formulaic expressions in spontaneous speech of left- and right-hemisphere-damaged subjects. *Aphasiology* **20**, 411–426 (2006).
  52. Yang, S.-Y. & Van Lancker Sidtis, D. Production of Korean Idiomatic Utterances Following Left- and Right-Hemisphere Damage: Acoustic Studies. *J. Speech Lang. Hear. Res.* **59**, 267–280 (2016).
  53. Van Lancker, D. R. & Kempler, D. Comprehension of familiar phrases by left- but not by right-hemisphere damaged patients. *Brain Lang.* **32**, 265–277 (1987).
  54. Sidtis, D., Canterucci, G. & Katsnelson, D. Effects of neurological damage on production of formulaic language. *Clin. Linguist. Phon.* **23**, 270–284 (2009).
  55. Pillay, S. B., Binder, J. R., Humphries, C., Gross, W. L. & Book, D. S. Lesion localization of speech comprehension deficits in chronic aphasia. *Neurology* **88**, 970–975 (2017).
  56. Rogalsky, C. *et al.* The Neurobiology of Agrammatic Sentence Comprehension: A Lesion Study. *J. Cogn. Neurosci.* **30**, 234–255 (2018).
  57. Sul, B. *et al.* Association of Lesion Location With Long-Term Recovery in Post-stroke Aphasia and Language Deficits. *Front. Neurol.* **10**, 776 (2019).
  58. Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B. & Jaeger, J. J. Lesion analysis of the brain areas involved in language comprehension. *Cognition* **92**, 145–177 (2004).
  59. Fridriksson, J., Fillmore, P., Guo, D. & Rorden, C. Chronic Broca’s Aphasia Is Caused

- by Damage to Broca's and Wernicke's Areas. *Cereb. Cortex* (2014)
- doi:10.1093/cercor/bhu152.
60. Forkel, S. J. & Catani, M. Lesion mapping in acute stroke aphasia and its implications for recovery. *Neuropsychologia* **115**, 88–100 (2018).
  61. Baldo, J. V., Arévalo, A., Patterson, J. P. & Dronkers, N. F. Grey and white matter correlates of picture naming: evidence from a voxel-based lesion analysis of the Boston Naming Test. *Cortex* **49**, 658–667 (2013).
  62. Faulkner, J. W. & Wilshire, C. E. Mapping eloquent cortex: A voxel-based lesion-symptom mapping study of core speech production capacities in brain tumour patients. *Brain Lang.* **200**, 104710 (2020).
  63. Bridges, K. A. & Van Lancker Sidtis, D. Formulaic Language in Alzheimer's Disease. *Aphasiology* **27**, (2013).
  64. Nicholas, M., Obler, L. K., Albert, M. L. & Helm-Estabrooks, N. Empty speech in Alzheimer's disease and fluent aphasia. *J. Speech Hear. Res.* **28**, 405–410 (1985).
  65. Zimmerer, V. C., Wibrow, M. & Varley, R. A. Formulaic Language in People with Probable Alzheimer's Disease: A Frequency-Based Approach. *J. Alzheimers. Dis.* **53**, 1145–1160 (2016).
  66. Van Lancker Sidtis, D., Choi, J., Alken, A. & Sidtis, J. J. Formulaic Language in Parkinson's Disease and Alzheimer's Disease: Complementary Effects of Subcortical and Cortical Dysfunction. *J. Speech Lang. Hear. Res.* **58**, 1493–1507 (2015).
  67. Thompson, P. M. *et al.* Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb. Cortex* **11**, 1–16 (2001).
  68. Thompson, P. M. *et al.* Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci.* **23**, 994–1005 (2003).
  69. Halliday, G. M., Double, K. L., Macdonald, V. & Kril, J. J. Identifying severely atrophic cortical subregions in Alzheimer's disease. *Neurobiol. Aging* **24**, 797–806 (2003).

70. Lehmann, M. *et al.* Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. *Neurobiol. Aging* **32**, 1466–1476 (2011).
71. Arnold, S. E., Hyman, B. T., Flory, J., Damasio, A. R. & Van Hoesen, G. W. The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cereb. Cortex* **1**, 103–116 (1991).
72. Damoiseaux, J. S., Prater, K. E., Miller, B. L. & Greicius, M. D. Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiol. Aging* **33**, 828.e19–30 (2012).
73. Lehmann, M. *et al.* Loss of functional connectivity is greater outside the default mode network in nonfamilial early-onset Alzheimer's disease variants. *Neurobiol. Aging* **36**, 2678–2686 (2015).
74. Bridges, K. A., Van Lancker Sidtis, D. & Sidtis, J. J. The role of subcortical structures in recited speech: Studies in Parkinson's disease. *J. Neurolinguistics* **26**, 594–601 (2013).
75. Illes, J., Metter, E. J., Hanson, W. R. & Iritani, S. Language production in Parkinson's disease: acoustic and linguistic considerations. *Brain Lang.* **33**, 146–160 (1988).
76. Hacker, C. D., Perlmuter, J. S., Criswell, S. R., Ances, B. M. & Snyder, A. Z. Resting state functional connectivity of the striatum in Parkinson's disease. *Brain* **135**, 3699–3711 (2012).
77. Sharman, M. *et al.* Parkinson's disease patients show reduced cortical-subcortical sensorimotor connectivity. *Mov. Disord.* **28**, 447–454 (2013).
78. Díez-Cirarda, M. *et al.* Dynamic functional connectivity in Parkinson's disease patients with mild cognitive impairment and normal cognition. *Neuroimage Clin* **17**, 847–855 (2018).
79. Skipper, J. I., Devlin, J. T. & Lametti, D. R. The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain*

- Lang.* **164**, 77–105 (2017).
80. Archakov, D. *et al.* Auditory representation of learned sound sequences in motor regions of the macaque brain. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 15242–15252 (2020).
  81. Stephan, M. A., Lega, C. & Penhune, V. B. Auditory prediction cues motor preparation in the absence of movements. *Neuroimage* **174**, 288–296 (2018).
  82. Lahav, A., Saltzman, E. & Schlaug, G. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* **27**, 308–314 (2007).
  83. Lee, H. & Noppeney, U. Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E1441–50 (2011).
  84. Bangert, M. *et al.* Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage* **30**, 917–926 (2006).
  85. Zatorre, R. J., Chen, J. L. & Penhune, V. B. When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* **8**, 547–558 (2007).
  86. Du, Y. & Zatorre, R. J. Musical training sharpens and bonds ears and tongue to hear speech better. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 13579–13584 (2017).
  87. Zatorre, R. J. Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science* **342**, 585–589 (2013).
  88. Grahn, J. A. & Rowe, J. B. Feeling the Beat: Premotor and Striatal Interactions in Musicians and Nonmusicians during Beat Perception. *Journal of Neuroscience* **29**, 7540–7548 (2009).
  89. Grahn, J. A. Neural mechanisms of rhythm perception: current findings and future perspectives. *Top. Cogn. Sci.* **4**, 585–606 (2012).
  90. Freitas, C. *et al.* Neural Correlates of Familiarity in Music Listening: A Systematic Review and a Neuroimaging Meta-Analysis. *Front. Neurosci.* **12**, 686 (2018).
  91. Bakker-Marshall, I. *et al.* Theta-band Oscillations in the Middle Temporal Gyrus Reflect

- Novel Word Consolidation. *J. Cogn. Neurosci.* **30**, 621–633 (2018).
92. Takashima, A., Bakker, I., van Hell, J. G., Janzen, G. & McQueen, J. M. Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *Neuroimage* **84**, 265–278 (2014).
  93. De Diego-Balaguer, R. *et al.* Striatal degeneration impairs language learning: evidence from Huntington’s disease. *Brain* **131**, 2870–2881 (2008).
  94. Bahlmann, J., Schubotz, R. I. & Friederici, A. D. Hierarchical artificial grammar processing engages Broca’s area. *Neuroimage* **42**, 525–534 (2008).
  95. López-Barroso, D. *et al.* Multiple brain networks underpinning word learning from fluent speech revealed by independent component analysis. *Neuroimage* **110**, 182–193 (2015).
  96. Callan, D. E. *et al.* Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage* **19**, 113–124 (2003).
  97. Wilson, S. M. & Iacoboni, M. Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage* **33**, 316–325 (2006).
  98. Schmitz, J. *et al.* Motor cortex compensates for lack of sensory and motor experience during auditory speech perception. *Neuropsychologia* (2018)  
doi:10.1016/j.neuropsychologia.2018.01.006.
  99. Assaneo, M. F. *et al.* Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nat. Neurosci.* **22**, 627–632 (2019).
  100. Skipper, J. I., Nusbaum, H. C. & Small, S. L. Lending a helping hand to hearing: another motor theory of speech perception. *Action to language via the mirror neuron system* 250–285 (2006).
  101. Skipper, J. I. Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130297 (2014).



102. Skipper, J. I. & Zevin, J. D. Brain reorganization in anticipation of predictable words. *bioRxiv* 101113 (2017) doi:10.1101/101113.
103. Skipper, J. I. & Lamett, D. R. Speech perception under the tent: A domain-general predictive role for the cerebellum. *bioRxiv* 2020.06.05.136804 (2020) doi:10.1101/2020.06.05.136804.
104. Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C. & Small, S. L. Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* **101**, 260–277 (2007).
105. Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C. & Small, S. L. Gestures orchestrate brain networks for language understanding. *Curr. Biol.* **19**, 661–667 (2009).
106. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
107. Block, C. K. & Baldwin, C. L. Cloze probability and completion norms for 498 sentences: behavioral and neural validation using event-related potentials. *Behav. Res. Methods* **42**, 665–670 (2010).
108. Van Heuven, W. J. B., Mandera, P., Keuleers, E. & Brysbaert, M. SUBTLEX-UK: A new and improved word frequency database for British English. *Q. J. Exp. Psychol.* **67**, 1176–1190 (2014).
109. Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. in *Soviet physics doklady* vol. 10 707–710 (1966).
110. Saberi, K. & Perrott, D. R. Cognitive restoration of reversed speech. *Nature* **398**, 760 (1999).
111. Feinberg, D. A. *et al.* Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS One* **5**, e15710 (2010).
112. Feinberg, D. A. & Setsompop, K. Ultra-fast MRI of the human brain with simultaneous multi-slice imaging. *J. Magn. Reson.* **229**, 90–100 (2013).
113. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic

- resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
114. Iglesias, J. E., Liu, C.-Y., Thompson, P. M. & Tu, Z. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Trans. Med. Imaging* **30**, 1617–1634 (2011).
115. Holmes, C. J. *et al.* Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* **22**, 324–333 (1998).
116. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
117. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).
118. Smith, S. M. *et al.* Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144–168 (2013).
119. Griffanti, L. *et al.* Hand classification of fMRI ICA noise components. *Neuroimage* **154**, 188–205 (2017).
120. Glover, G. H. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* **9**, 416–429 (1999).
121. Ward, B. D. Deconvolution analysis of fMRI time series data. *Milwaukee, WI: Biophysics Research Institute, Medical College of Wisconsin* (2002).
122. Friedman, L., Glover, G. H., Krenz, D., Magnotta, V. & FIRST BIRN. Reducing inter-scanner variability of activation in a multicenter fMRI study: role of smoothness equalization. *Neuroimage* **32**, 1656–1668 (2006).
123. Chen, G., Saad, Z. S., Britton, J. C., Pine, D. S. & Cox, R. W. Linear mixed-effects modeling approach to FMRI group analysis. *Neuroimage* **73**, 176–190 (2013).
124. Zang, Y., Jiang, T., Lu, Y., He, Y. & Tian, L. Regional homogeneity approach to fMRI data analysis. *Neuroimage* **22**, 394–400 (2004).
125. Jiang, L. & Zuo, X.-N. Regional Homogeneity: A Multimodal, Multiscale Neuroimaging

- Marker of the Human Connectome. *Neuroscientist* (2015)
- doi:10.1177/1073858415595004.
126. Cox, R. W. Equitable Thresholding and Clustering: A Novel Method for Functional Magnetic Resonance Imaging Clustering in AFNI. *Brain Connect.* **9**, 529–538 (2019).
127. Garrison, K. A., Scheinost, D., Finn, E. S., Shen, X. & Constable, R. T. The (in)stability of functional brain network measures across thresholds. *Neuroimage* **118**, 651–661 (2015).
128. Wang, J. *et al.* Test--retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Hum. Brain Mapp.* **38**, 2226–2241 (2017).
129. Wills, P. & Meyer, F. G. Metrics for graph comparison: A practitioner's guide. *PLoS One* **15**, e0228728 (2020).
130. Koutra, D., Vogelstein, J. T. & Faloutsos, C. DELTACON: A Principled Massive-Graph Similarity Function. *arXiv [cs.SI]* (2013).
131. Ek, B., VerSchneider, C. & Narayan, D. A. Global efficiency of graphs. *AKCE Int. J. Graphs Comb.* **12**, 1–13 (2015).
132. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
133. Baldo, J. V., Kacinik, N. A., Moncrief, A., Beghin, F. & Dronkers, N. F. You may now kiss the bride: Interpretation of social situations by individuals with right or left hemisphere injury. *Neuropsychologia* **80**, 133–141 (2016).
134. Speedie, L. J., Wertman, E., Ta'ir, J. & Heilman, K. M. Disruption of automatic speech following a right basal ganglia lesion. *Neurology* **43**, 1768–1774 (1993).
135. Sidtis, J. J., Van Lancker Sidtis, D., Dhawan, V. & Eidelberg, D. Switching Language Modes: Complementary Brain Patterns for Formulaic and Propositional Language. *Brain Connect.* **8**, 189–196 (2018).
136. Pini, L. *et al.* Brain atrophy in Alzheimer's Disease and aging. *Ageing Res. Rev.* **30**,

- 25–48 (2016).
137. Cho, H. *et al.* Shape changes of the basal ganglia and thalamus in Alzheimer’s disease: a three-year longitudinal study. *J. Alzheimers. Dis.* **40**, 285–295 (2014).
138. Tentolouris-Piperas, V., Ryan, N. S., Thomas, D. L. & Kinnunen, K. M. Brain imaging evidence of early involvement of subcortical regions in familial and sporadic Alzheimer’s disease. *Brain Res.* **1655**, 23–32 (2017).
139. Boulenger, V., Hauk, O. & Pulvermuller, F. Grasping Ideas with the Motor System: Semantic Somatotopy in Idiom Comprehension. *Cereb. Cortex* **19**, 1905–1914 (2008).
140. Hillert, D. G. & Buračas, G. T. The neural substrates of spoken idiom comprehension. *Lang. Cogn. Process.* **24**, 1370–1391 (2009).
141. Mashal, N., Faust, M., Hendler, T. & Jung-Beeman, M. Hemispheric differences in processing the literal interpretation of idioms: converging evidence from behavioral and fMRI studies. *Cortex* **44**, 848–860 (2008).
142. Bohrn, I. C., Altmann, U. & Jacobs, A. M. Looking at the brains behind figurative language—A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia* **50**, 2669–2683 (2012).
143. Reyes-Aguilar, A., Valles-Capetillo, E. & Giordano, M. A Quantitative Meta-analysis of Neuroimaging Studies of Pragmatic Language Comprehension: In Search of a Universal Neural Substrate. *Neuroscience* **395**, 60–88 (2018).
144. Yang, J. The role of the right hemisphere in metaphor comprehension: A meta-analysis of functional magnetic resonance imaging studies. *Hum. Brain Mapp.* **35**, 107–122 (2014).
145. Kasparian, K. Hemispheric differences in figurative language processing: Contributions of neuroimaging methods and challenges in reconciling current empirical findings. *J. Neurolinguistics* **26**, 1–21 (2013).
146. Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W. & Jacobs, A. M. Old proverbs

- in new skins - an fMRI study on defamiliarization. *Front. Psychol.* **3**, 204 (2012).
147. Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A. & Chatterjee, A. From novel to familiar: tuning the brain for metaphors. *Neuroimage* **59**, 3212–3221 (2012).
148. Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W. & Jacobs, A. M. When we like what we know--a parametric fMRI analysis of beauty and familiarity. *Brain Lang.* **124**, 1–8 (2013).
149. Schmidt, G. L. & Seger, C. A. Neural correlates of metaphor processing: the roles of figurativeness, familiarity and difficulty. *Brain Cogn.* **71**, 375–386 (2009).
150. Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M. & Pietrini, P. Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Res. Bull.* **86**, 203–216 (2011).
151. Diaz, M. T., Barrett, K. T. & Hogstrom, L. J. The influence of sentence novelty and figurativeness on brain activity. *Neuropsychologia* **49**, 320–330 (2011).
152. Subramaniam, K., Faust, M., Beeman, M. & Mashal, N. The repetition paradigm: enhancement of novel metaphors and suppression of conventional metaphors in the left inferior parietal lobe. *Neuropsychologia* **50**, 2705–2719 (2012).
153. Protopapas, A. *et al.* Evaluating cognitive models of visual word recognition using fMRI: Effects of lexical and sublexical variables. *Neuroimage* **128**, 328–341 (2016).
154. Carreiras, M., Riba, J., Vergara, M., Heldmann, M. & Münte, T. F. Syllable congruency and word frequency effects on brain activation. *Hum. Brain Mapp.* **30**, 3079–3088 (2009).
155. Fiez, J. A., Balota, D. A., Raichle, M. E. & Petersen, S. E. Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron* **24**, 205–218 (1999).
156. Desai, R. H., Choi, W. & Henderson, J. M. Word frequency effects in naturalistic reading. *Language, Cognition and Neuroscience* **35**, 583–594 (2020).

157. Blumenthal-Dramé, A. *et al.* Frequency and Chunking in Derived Words: A Parametric fMRI Study. *J. Cogn. Neurosci.* **29**, 1162–1177 (2017).
158. Bhattasali, S., Fabre, M. & Hale, J. Processing MWEs: Neurocognitive Bases of Verbal MWEs and Lexical Cohesiveness within MWEs. in *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)* 6–17 (2018).
159. Bhattasali, S. *et al.* Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension. *Language, Cognition and Neuroscience* **34**, 491–510 (2019).
160. Bemis, D. K. & Pykkänen, L. Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *J. Neurosci.* **31**, 2801–2814 (2011).
161. Price, A. R., Bonner, M. F., Peelle, J. E. & Grossman, M. Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *J. Neurosci.* **35**, 3276–3284 (2015).
162. Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L. & Seidenberg, M. S. Neural correlates of implicit and explicit combinatorial semantic processing. *Neuroimage* **53**, 638–646 (2010).
163. Dayan, E. & Cohen, L. G. Neuroplasticity subserving motor skill learning. *Neuron* **72**, 443–454 (2011).
164. Lohse, K. R., Wadden, K., Boyd, L. A. & Hodges, N. J. Motor skill acquisition across short and long time scales: a meta-analysis of neuroimaging data. *Neuropsychologia* **59**, 130–141 (2014).
165. Kelly, A. M. C. & Garavan, H. Human Functional Neuroimaging of Brain Changes Associated with Practice. *Cereb. Cortex* **15**, 1089–1102 (2005).
166. Doyon, J., Gabbitov, E., Vahdat, S., Lungu, O. & Boutin, A. Current issues related to

- motor sequence learning in humans. *Current Opinion in Behavioral Sciences* **20**, 89–97 (2018).
167. Doyon, J. *et al.* Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behav. Brain Res.* **199**, 61–75 (2009).
168. Graybiel, A. M. & Grafton, S. T. The striatum: where skills and habits meet. *Cold Spring Harb. Perspect. Biol.* **7**, a021691 (2015).
169. Yang, J. The influence of motor expertise on the brain activity of motor task performance: A meta-analysis of functional magnetic resonance imaging studies. *Cogn. Affect. Behav. Neurosci.* **15**, 381–394 (2015).
170. Caligiore, D., Arbib, M. A., Miall, R. C. & Baldassarre, G. The super-learning hypothesis: Integrating learning processes across cortex, cerebellum and basal ganglia. *Neurosci. Biobehav. Rev.* **100**, 19–34 (2019).
171. Ashby, F. G., Turner, B. O. & Horvitz, J. C. Cortical and basal ganglia contributions to habit learning and automaticity. *Trends Cogn. Sci.* **14**, 208–215 (2010).
172. Penhune, V. B. & Steele, C. J. Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. *Behav. Brain Res.* **226**, 579–591 (2012).
173. Albouy, G., King, B. R., Maquet, P. & Doyon, J. Hippocampus and striatum: dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. *Hippocampus* **23**, 985–1004 (2013).
174. Nijhuis, E. H. J., van Cappellen van Walsum, A.-M. & Norris, D. G. Topographic hub maps of the human structural neocortical network. *PLoS One* **8**, e65511 (2013).
175. Hagmann, P. *et al.* Mapping the structural core of human cerebral cortex. *PLoS Biol.* **6**, e159 (2008).
176. Cole, M. W., Pathak, S. & Schneider, W. Identifying the brain's most globally connected regions. *Neuroimage* **49**, 3132–3148 (2010).
177. van den Heuvel, M. P. & Sporns, O. An anatomical substrate for integration among

- functional networks in human cortex. *J. Neurosci.* **33**, 14489–14500 (2013).
178. Scholtens, L. H., Schmidt, R., de Reus, M. A. & van den Heuvel, M. P. Linking macroscale graph analytical organization to microscale neuroarchitectonics in the macaque connectome. *J. Neurosci.* **34**, 12192–12205 (2014).
179. Modha, D. S. & Singh, R. Network architecture of the long-distance pathways in the macaque brain. *Proceedings of the National Academy of Sciences* **107**, 13485–13490 (2010).
180. Harriger, L., van den Heuvel, M. P. & Sporns, O. Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS One* **7**, e46497 (2012).
181. Ghahremani, M., Hutchison, R. M., Menon, R. S. & Everling, S. Frontoparietal Functional Connectivity in the Common Marmoset. *Cereb. Cortex* (2016)  
doi:10.1093/cercor/bhw198.
182. Anderson, M. L. & Penner-Wilger, M. Neural reuse in the evolution and development of the brain: Evidence for developmental homology? *Dev. Psychobiol.* **55**, 42–51 (2013).
183. Warren, D. E. *et al.* Network measures predict neuropsychological outcome after brain injury. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14247–14252 (2014).
184. Crossley, N. A. *et al.* The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382–2395 (2014).
185. Stahl, B., Gawron, B., Regenbrecht, F., Flöel, A. & Kotz, S. A. Formulaic Language Resources May Help Overcome Difficulties in Speech-Motor Planning after Stroke. *PLoS One* **15**, e0233608 (2020).
186. Stahl, B., Henseler, I., Turner, R., Geyer, S. & Kotz, S. A. How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. *Front. Hum. Neurosci.* **7**, 35 (2013).
187. Stahl, B., Mohr, B., Dreyer, F. R., Lucchese, G. & Pulvermüller, F. Using language for social interaction: Communication mechanisms promote recovery from chronic



- non-fluent aphasia. *Cortex* **85**, 90–99 (2016).
188. Stahl, B. & Van Lancker Sidtis, D. Tapping into neural resources of communication: formulaic language in aphasia therapy. *Front. Psychol.* **6**, 1526 (2015).
189. Särkämö, T. & Sihvonen, A. J. Golden oldies and silver brains: Deficits, preservation, learning, and rehabilitation effects of music in ageing-related neurological disorders. *Cortex* **109**, 104–123 (2018).
190. Zumbansen, A., Peretz, I. & Hébert, S. Melodic intonation therapy: back to basics for future research. *Front. Neurol.* **5**, 7 (2014).
191. Upton, E. & Hope, T. M. H. Moving beyond the dual stream account of language. *Brain* **143**, 2336–2338 (2020).
192. Hula, W. D. *et al.* Structural white matter connectometry of word production in aphasia: an observational study. *Brain* **143**, 2532–2544 (2020).
193. Seghier, M. L. & Price, C. J. Interpreting and Utilising Intersubject Variability in Brain Function. *Trends Cogn. Sci.* **22**, 517–530 (2018).