



# Approaches to Measuring Language Lateralisation: An Exploratory Study Comparing Two fMRI Methods and Functional Transcranial Doppler Ultrasound

Dorothy V. M. Bishop<sup>1,2</sup>, Zoe V. J. Woodhead<sup>1,2</sup>, and Kate E. Watkins<sup>1,2</sup>

<sup>1</sup>Wellcome Centre for Integrative Neuroimaging, Oxford, United Kingdom

<sup>2</sup>Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom

**Keywords:** asymmetry, bilateral language, deactivation, functional transcranial Doppler ultrasound, laterality, region of interest

## ABSTRACT

In this exploratory study we compare and contrast two methods for deriving a laterality index (LI) from functional magnetic resonance imaging (fMRI) data: the weighted bootstrapped mean from the LI Toolbox (toolbox method), and a novel method that uses subtraction of activations from homologous regions in left and right hemispheres to give an array of difference scores (mirror method). Data came from 31 individuals who had been selected to include a high proportion of people with atypical laterality when tested with functional transcranial Doppler ultrasound (fTCD). On two tasks, word generation and semantic matching, the mirror method generally gave better agreement with fTCD laterality than the toolbox method, both for individual regions of interest, and for a large region corresponding to the middle cerebral artery. LI estimates from this method had much smaller confidence intervals (CIs) than those from the toolbox method; with the mirror method, most participants were reliably lateralised to left or right, whereas with the toolbox method, a higher proportion were categorised as bilateral (i.e., the CI for the LI spanned zero). Reasons for discrepancies between fMRI methods are discussed: one issue is that the toolbox method averages the LI across a wide range of thresholds. Furthermore, examination of task-related *t*-statistic maps from the two hemispheres showed that language lateralisation is evident in regions characterised by deactivation, and so key information may be lost by ignoring voxel activations below zero, as is done with conventional estimates of the LI.

## INTRODUCTION

It has been established for many years that language processing depends on specialised areas in the left side of the brain in most people. Initial evidence came from observations of aphasia after brain lesions; as early as the mid-19th century it was noted that aphasia was strongly associated with left-sided damage (see historical account by Berker et al., 1986). Neurosurgeons subsequently developed the Wada test (Wada & Rasmussen, 1960) in which each hemisphere was successively anaesthetised to determine which was dominant for language, again revealing the bias to left-sided language processing in most people. With the advent of modern brain imaging methods, it became possible to observe correlates of brain activation associated with different cognitive tasks, confirming the population bias to left hemisphere processing for generating

Citation: Bishop, D. V. M., Woodhead, Z. V. J., & Watkins, K. E. (2024). Approaches to measuring language lateralisation: An exploratory study comparing two fMRI methods and functional transcranial Doppler ultrasound. *Neurobiology of Language*, 5(2), 409–431. [https://doi.org/10.1162/nol\\_a\\_00136](https://doi.org/10.1162/nol_a_00136)

DOI:  
[https://doi.org/10.1162/nol\\_a\\_00136](https://doi.org/10.1162/nol_a_00136)

Supporting Information:  
[https://doi.org/10.1162/nol\\_a\\_00136](https://doi.org/10.1162/nol_a_00136)

Received: 5 September 2023  
Accepted: 8 January 2024

Competing Interests: The authors have declared that no competing interests exist.

Corresponding Author:  
Dorothy V. M. Bishop  
[dorothy.bishop@psy.ox.ac.uk](mailto:dorothy.bishop@psy.ox.ac.uk)

Handling Editor:  
Stefano Cappa

Copyright: © 2024  
Massachusetts Institute of Technology  
Published under a Creative Commons  
Attribution 4.0 International  
(CC BY 4.0) license

language. There is, however, a significant minority of people who depart from the population trend and have atypical lateralisation; this may take the form of reversal of the usual left-sided bias for language—right-hemisphere language—or an overall lack of bias—bilateral language.

Atypical lateralisation is of considerable theoretical interest to neuroscientists interested in the evolutionary origins and functional significance of laterality. For many years, developmental disorders have been linked with atypical language lateralisation, but there is little solid, replicable evidence, in part because of inconsistencies in methods for quantifying lateralisation (see Bishop, 2013, for review). In addition, language lateralisation is relevant for understanding how individuals recover from unilateral brain injury and for planning neurosurgical procedures involving language areas.

The current study aims to advance progress in developing reliable, sensitive, and valid methods for measuring functional brain lateralisation in individuals.

### Functional Transcranial Doppler Ultrasound

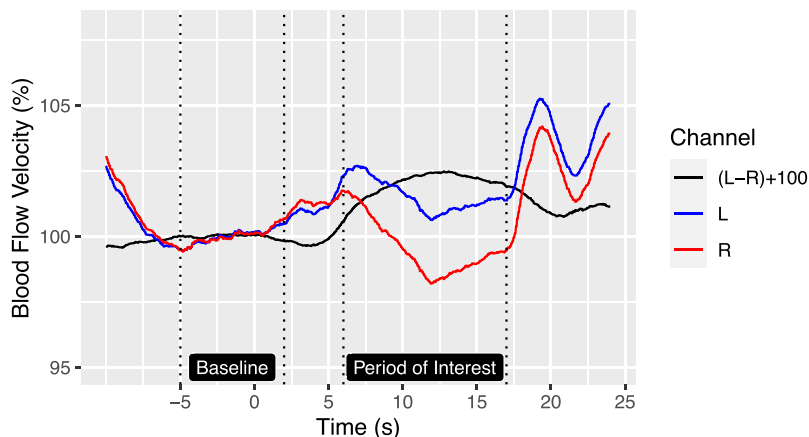
**Functional transcranial Doppler ultrasound (fTCD):**  
A method of measuring blood flow velocity in cerebral arteries via ultrasound probes positioned on the head.

**Laterality index (LI):**  
A measure of asymmetric task processing that is obtained by dividing the difference between left- and right-sided values by the sum of left- and right-sided values.

**Word generation:**  
Regarded as a gold-standard task for assessing language lateralisation; it is also known as verbal fluency or phonemic fluency, and it requires the participant to generate words that begin with a specific letter.

Functional transcranial Doppler ultrasound (fTCD) is an inexpensive and portable method for measuring lateralisation by comparing blood flow velocity in the left and right middle cerebral arteries (MCAs) during an activation task, relative to a resting baseline period. It is relatively insensitive to movement artefacts, and can be used with children aged 4 years and over (Bishop et al., 2014). In a recent study, Parker et al. (2022) compared lateralisation indices (LIs) for six different language tasks in left- and right-handers. Left-lateralisation characterised tasks involving language generation, but not tasks that predominantly involved receptive language, and lateralisation was generally weaker in left-handers. Split-half reliabilities were reasonable for all tasks (0.7 or more).

With fTCD, signals from left and right sides are normalised by dividing by the mean blood flow velocity of that side, and each trial is baseline corrected by subtracting the mean value from a rest period prior to stimulus presentation. A sample plot from averaged data on word generation is shown in Figure 1, with the black line showing the (rescaled) difference between left and right channels. When fTCD was first developed as a tool for investigating laterality, the LI was obtained by identifying a temporal window, or period of interest (POI), and taking the mean value in an interval defined as  $\pm 2$  s around the absolute peak difference between left and right signals in the POI (Deppe et al., 1997). This method had the disadvantage that it could give a biased LI estimate in weakly lateralised individuals where the peak difference



**Figure 1.** Averaged blood flow velocity over course of a word generation trial.

changed from left to right in the course of the POI. Selecting the larger peak could create a spurious bimodality in the distribution of LIs in a sample. As an alternative, Woodhead et al. (2020) proposed a simple subtraction method whereby the mean signal in the right MCA was subtracted from that on the left. The two methods were highly correlated, but the subtraction method gave a normal distribution of LIs in a sample. More recently, Thompson et al. (2023) proposed a method more analogous to that used in fMRI, where generalised models were used to estimate the interaction of side (left vs. right) with time interval (POI vs. non-POI) on the signal over a series of trials. This gave LIs that agreed well with the simpler subtraction method, but with a smaller error of measurement. For the current paper, we use the complex generalised additive model (GAM) method described by Thompson et al. (2023) to estimate LIs from fTCD data. This estimates the response function from the data, where predictors are side, POI (boxcar), plus terms for time relative to epoch start, plus a categorical index corresponding to individual epochs. LI is estimated as an interaction term: percentage difference in blood flow between left and right that is specific to the POI.

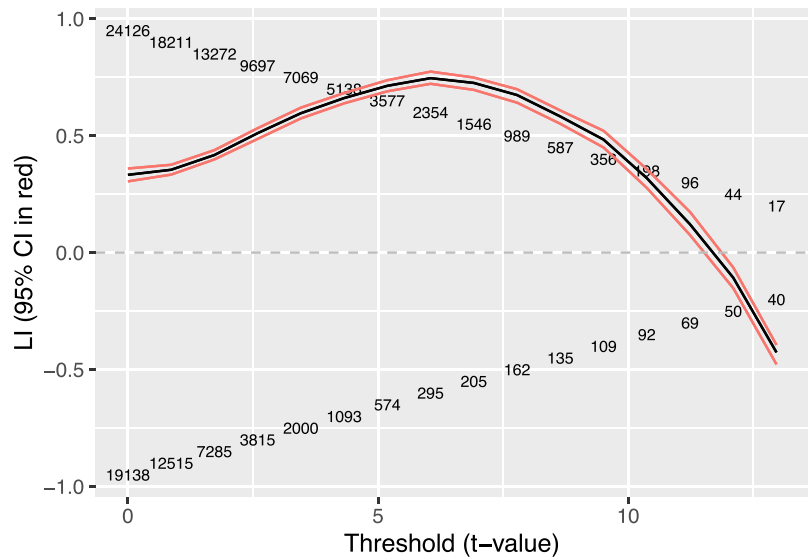
fTCD has excellent temporal resolution, but can only tell us about left–right differences, not where activations are occurring within a hemisphere. To answer questions about localisation, we need to turn to fMRI, which has complementary advantages and disadvantages to fTCD; it has poor temporal resolution but good spatial resolution and can measure changes in blood oxygenation in voxels as small as  $1 \text{ mm}^3$ . By comparing LIs obtained with fTCD ( $LI_{\text{fTCD}}$ ) and fMRI ( $LI_{\text{fMRI}}$ ) on comparable tasks, we can gain some insights into the sensitivity of fTCD to regional changes in blood flow. However, comparison is complicated by measurement issues.

### Functional Magnetic Resonance Imaging

A comprehensive review by Seghier (2008) described the range of methods that have been used to derive an LI from functional magnetic resonance imaging (fMRI) data. The optimal method for estimating laterality has been much debated (Bradshaw et al., 2017). Vingerhoets et al. (2023) used the Delphi approach (Hasson et al., 2000) to explore expert opinion on measurement of cerebral lateralisation. They concluded that there was still no agreement about the optimal way to measure lateralisation using fMRI, and the key concept of bilateral language was ill-defined. We need reliable and convenient methods for measuring functional lateralisation in individuals to advance research in this area. Here we compare a widely-used approach, the LI Toolbox, with an alternative approach, the mirror method.

### LI Toolbox

LI Toolbox is a freely available MATLAB toolbox that offers a flexible range of options for computing an LI (Wilke & Lidzba, 2007; Wilke & Schmithorst, 2006). The starting point is a  $t$ -statistic map derived in the conventional way from the general linear model, which is used to contrast the condition of interest with a comparison condition, which may be either rest or another condition that differs on key task demands. The next step is to identify voxels from the  $t$ -statistic map within the region of interest (ROI) that are above an activation threshold, and compute the LI by the standard formula  $(L - R) / (L + R)$  where L and R are either the number of voxels, or the summed activation of suprathreshold voxels on L and R sides. In practice, it makes little difference whether number of voxels or summed activation is used in the formula (Wilke & Lidzba, 2007). However, the choice of threshold can make a substantial difference to the LI. Figure 2 shows an illustrative example where the computed LI (based on summed  $t$  values in activated voxels) rises from 0.33 to 0.75 between thresholds of zero to 6 and then declines to become negative (i.e., right-biased) at thresholds of 12 or more.

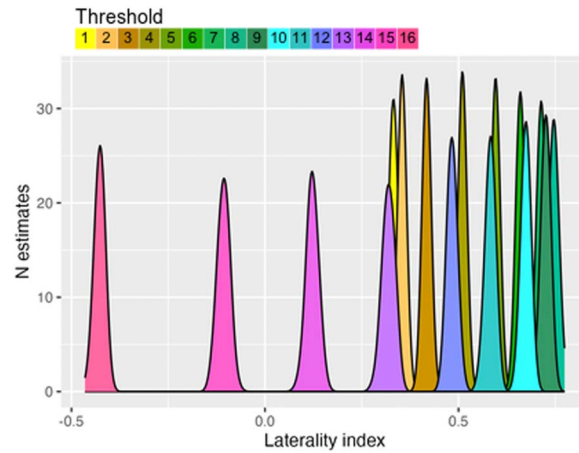


**Figure 2.** Laterality index (LI) estimates (based on summed activation) at different thresholds in a sample individual fMRI dataset using a word generation task for activation. The numbers in the body of the graph show the numbers of suprathreshold voxels on the left (above zero) and right (below zero). There are 16 threshold  $t$  values ranging from 0 to 12.96. Higher thresholds yielded too few voxels to meet criteria for inclusion.

One method from the LI Toolbox that has become popular as a way of taking into account information from a wide range of thresholds involves bootstrapping: The paper by Wilke and Schmithorst (2006) describing this approach has received 185 citations in Web of Science at the time of writing. This method (described in Bradshaw et al., 2017; Wilke & Schmithorst, 2006) takes an iterative approach. First a set of 20  $t$  thresholds is created using equally spaced steps from 0 to the maximum  $t$  value in the image for the ROI. Threshold values resulting in fewer than 10 activated voxels on either side are discarded. At each remaining threshold, and for each hemisphere, the bootstrapping procedure starts by defining the set of suprathreshold voxels. From this set, 100 samples (with replacement) are taken of the  $t$  statistics within the thresholded ROI; the default is to take 25% of values in each resample. So if in a given ROI there are 40 suprathreshold voxels in the left hemisphere and 24 in the right, 100 samples of 10  $t$  values are taken from the left, and 100 samples of 6  $t$  values from the right. An LI value is calculated based on the  $t$  statistics for every left + right combination of these re-samples ( $100 * 100 = 10,000$  LI values) at each threshold level. The bootstrapping approach has the advantage that a CI can be specified around the LI estimate for an individual. This makes it possible to have a statistical definition of bilateral language: Rather than relying on an arbitrary cutoff, we can designate a person as having bilateral language if the 95% CI around the LI includes zero.

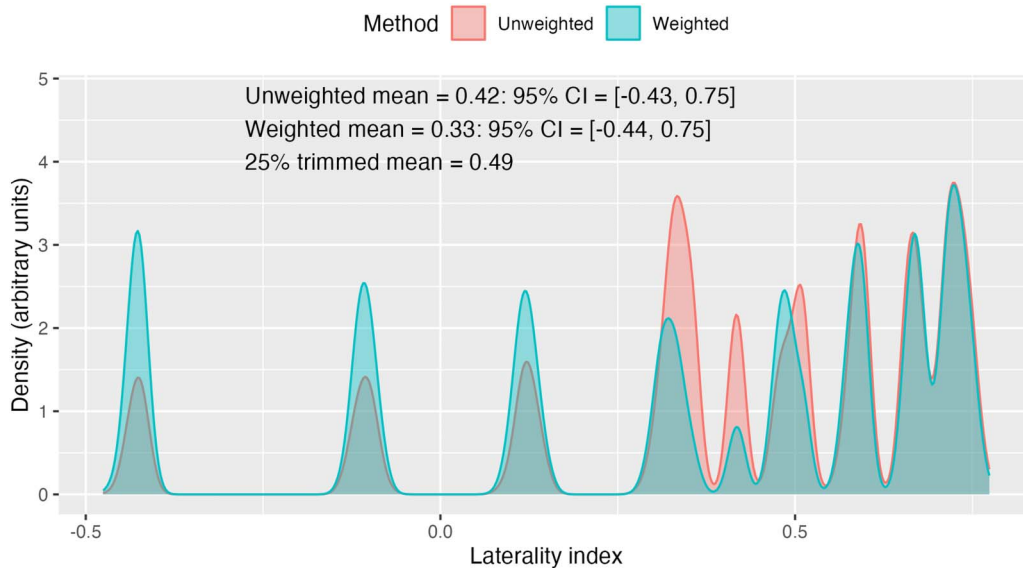
Sample data from the same participant as in Figure 2 are shown in Figure 3. Note that at any one threshold, the density plot of LIs is narrow, indicating little variation from iteration to iteration of the bootstrap, but the differences in LI between thresholds is substantial, and, as also shown in Figure 2, the LI for this participant becomes right-sided at very high thresholds.

Alternative options are available for computing the LI from these estimates: The simplest is the mean of all the estimates—to reduce the impact of outliers, this is based on the middle 50% of values at each threshold, that is, the 25% trimmed means. Further trimming can be conducted to give a trimmed mean based on the middle 50% of all of the trimmed mean LI values. Finally, one can compute a weighted mean that weights the trimmed mean LIs from each threshold by the threshold. Wilke and Schmithorst (2006) recommended this approach as



**Figure 3.** Distributions of bootstrapped LI estimates at 16 different thresholds in a sample individual fMRI dataset using a word generation task for activation. Threshold cutoffs are set in 20 steps from 0 to maximum activation. In this case the top four threshold cutoffs had insufficient voxels for inclusion.

it ensures that voxels with the highest correlation with the task will have most impact on the LI. All methods of mean estimation are intended to reduce the impact of outliers. Note that the 95% CI around these estimates is substantial; this does not indicate that the estimates are likely to be irreproducible. In fact, we know that the weighted mean LI from the LI Toolbox has excellent test–retest reliability (Johnstone et al., 2020). The large CI reflects the fact that it is based on a wide distribution of thresholds; within any one threshold, the estimates of LI are very similar, but across the range of thresholds, the LI may vary substantially. Although the LI Toolbox is well specified and justified, it still involves some arbitrary decisions. The method is described as “threshold-free,” but it is perhaps more accurate to say that it uses an algorithm for selecting a representative LI (either by mean, trimmed mean, or weighted mean), which will incorporate information from a wide range of thresholds (Figure 4).



**Figure 4.** Weighted and unweighted distributions of bootstrapped LI estimates from the same dataset as Figure 2. The weighting makes the peaks at high thresholds higher and those at low thresholds lower. The plot annotation shows LI estimates with different methods.

LI estimates will depend on the range of thresholds considered. For instance, if we had taken a threshold range of  $t$  values from 0 to 6, instead of 0 to 13 for the data in Figure 2, the LIs (mean, trimmed mean, or weighted mean) would have been higher, the range of LI values lower, and the CIs substantially smaller. The LI Toolbox uses a default threshold range extending to the maximum  $t$  value observed in the image. It follows that the threshold range can vary from person to person, task to task, and ROI to ROI. Furthermore, fewer voxels are included at high thresholds, and so the least reliable LI estimates contribute most to the weighted mean. Another point is that the method excludes voxels with negative  $t$  scores (i.e., below threshold of zero). It is generally assumed that only positive  $t$  values indicating task-related activation are of interest; however, as will be discussed further below, for some tasks and brain regions, functional lateralisation is evident when there are negative  $t$  values in both hemispheres, suggestive of task-related deactivation. These three considerations—range of thresholds, weighting by level of activation, and removal of voxels with negative  $t$  values—are not necessarily problems for the method, but they do confirm that decisions about which settings to use require careful thought and will have consequences for LI estimates.

#### **Alternative Approach to fMRI Laterality: Mirror Method**

We present here results from an alternative approach to measuring functional lateralisation, the mirror method, which was designed to be more comparable to the way in which laterality is assessed in fTCD. The logic is similar to the flip method, which was used by Watkins et al. (2001) to assess structural asymmetries in human MRI. In this method, rather than assessing how many voxels have suprathreshold  $t$  statistics in left and right hemispheres, the right-brain  $t$ -statistic map is flipped on the left-right axis and then subtracted from the left-brain  $t$ -statistic map to give a map of a single hemisphere showing left–right differences in  $t$  statistics for homologous regions.

This kind of method has occasionally been applied to functional MRI (Baciu et al., 2005; Cousin et al., 2007; Seghier et al., 2011), but there is no consensus on how to derive an individual LI using this approach. We term our method the *mirror method* to avoid confusion with other approaches that involve flipping one hemisphere before subtracting the activation maps.

Any method for determining laterality has to consider how to handle spatial dependencies between voxels. Smoothing and clustering are typically used (and options for clustering are provided in the LI Toolbox). Our approach with the mirror method was to minimise spatial dependencies between voxels within an ROI by repeated sampling of a small set of values. Pilot testing showed that correlations between adjacent voxels in a sample were minimised when a sample of 5% voxels was used, and so 1,000 samples (without replacement), each containing a random 5% of voxels in the region of interest, were taken, and the difference score computed at each voxel. The mean and the 2.5th and 97.5th percentiles were taken from the 1,000 samples to give the average difference score with a 95% CI. Note that this method is threshold-free, with all  $t$ -statistic map values being included, regardless of whether positive or negative.

#### **Prior Comparisons of fMRI and fTCD**

Thompson et al. (2023) compared estimates of LI from different analytic methods for a group of individuals who had completed similar language tasks with both fTCD and fMRI. A GAM approach was developed to be more comparable to the generalised linear model method used with fMRI. The individual LI values obtained with this method are closely comparable to those obtained by subtracting averaged left and right cerebral blood flow velocity (CBFV) during the

POI; however, the LI estimates have smaller standard errors than those from more traditional methods. The toolbox  $LI_{fMRI}$  was computed after applying a mask combining frontal, temporal, and parietal lobes, as an approximation of the MCA territory. For the standard toolbox  $LI_{fMRI}$  (weighted mean from bootstrap), the correlation with  $LI_{fTCD}$  was 0.66, indicating fair but far from perfect agreement between methods. This raises the question of how far differences relate to method of computing the LI (difference score for fTCD and laterality index for fMRI), or better sensitivity to regional laterality effects in fMRI.

### **Current Aims**

The main focus of Thompson et al. (2023) was on developing the GAM approach to analysing fTCD data. In contrast, our approach here is to consider in detail how  $LI_{fTCD}$  relates to  $LI_{fMRI}$ , using the same dataset from Bruckert (2016). Following Seghier et al. (2011), we also aimed to see how laterality varied across different brain regions. We used data from two tasks, word generation and semantic matching, and four ROIs, frontal, temporal, parietal, and cerebellar, and contrasted the toolbox  $LI_{fMRI}$  and the mirror  $LI_{fMRI}$  with the  $LI_{fTCD}$ .

As well as considering quantitative LIs, language laterality was categorised on each measure as left-lateralised if the lower CI exceeded zero, right-lateralised if the upper CI was below zero, and bilateral if zero was included in the CI.

## **MATERIALS AND METHODS**

### **Design**

A within-subjects design was used where participants were tested in two sessions: the first using fTCD, and the second using fMRI. On average, the interval between the two sessions was 11.8 days (range: 1–31 days).

In both sessions participants performed two tasks: a speech production task (word generation) and a semantic matching task based on the Pyramids and Palm Trees test (Howard & Patterson, 1992). An additional production task, auditory naming, was administered only in the fMRI session. The word generation and semantic matching tasks were designed to be as similar as possible in the two modalities (fTCD and fMRI), but some minor differences were required as noted below. Hence, the main within-subject independent variables were method (fTCD vs. fMRI) and task (word generation vs. semantic decision) and the dependent variable was LI. For the semantic matching task, we administered a comparison (baseline) perceptual judgement task in the fMRI session, but here we present LIs from semantic matching versus rest, as this is comparable to the procedure used with fTCD. Further information about results with the perceptual judgement baseline task is provided in Supplementary Material 1 of the Supporting Information, available at [https://doi.org/10.1162/nol\\_a\\_00136](https://doi.org/10.1162/nol_a_00136). The auditory naming task was added to the analysis for comparisons of methods of laterality assessment in fMRI.

### **Participants**

Participants were recruited from a sample of 231 individuals who were screened with fTCD. These participants were recruited from the Oxford Psychology Research Participant Recruitment Scheme (<https://opr.sona-systems.com>). All participants gave written, informed consent, and all procedures were approved by the Central University Research Ethics Committee of the University of Oxford (MSD-IDREC-C1-20014-003). fTCD data from the full sample have been described by Bruckert et al. (2021).

In order to obtain a sample with cases of both typical and atypical lateralisation, all 21 participants who showed right lateralised or bilateral activation during word generation in fTCD were invited to return for the second session with fMRI. This classification was based on whether or not the 95% CI for the LI index on word generation crossed zero, where the CI was derived from the standard error of the LIs obtained on individual trials. Sixteen of these participants agreed to take part and fulfilled all MRI safety criteria. Sixteen participants with left lateralisation for word generation in fTCD, matched for age, gender, and handedness, were also invited to return for the second session. The final sample that participated in both sessions comprised 32 participants (14 women, 18 men; 20 right handed, 12 left handed; mean age = 24.9 years, *SD* = 5.1 years). Due to time constraints, one participant missed the semantic matching task in the fMRI session, and another participant missed the auditory naming task, so the final sample size in the typical lateralisation group was 15 for those two tasks, and 16 for word generation.

### Functional Transcranial Doppler Ultrasound

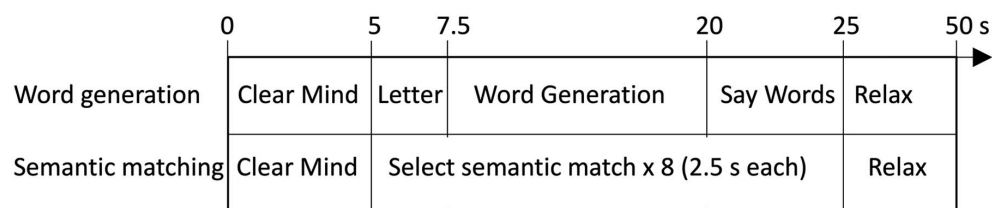
#### Procedure

fTCD was acquired during the first session. Participants were trained to perform the two tasks and completed some practice trials; then the fTCD headset and probes were fitted and a stable signal was found. The fTCD signal was then recorded, first for the word generation task, then for the semantic matching task, with a break in between.

The timings for the two tasks are shown in Figure 5. Both tasks had a trial structure that started with a “Clear Mind” instruction on screen (5 s), followed by the task for 20 s, and ending with 25 s of rest where the word “Relax” was presented on screen. Participants were instructed to stay still and think of as little as possible during the Relax period. The onset of the Clear Mind cue was taken as the start of the trial (peri-stimulus time = 0 s).

**Word Generation.** In this task (also known as verbal or phonemic fluency) a letter was presented in the centre of a screen and the participant was required to silently generate as many words as possible starting with that letter. There were 23 trials, each with a different letter (excluding Q, X, and Z) presented in a randomised order. After the Clear Mind cue, the letter was presented on screen for 2.5 s, followed by a blank screen for 12.5 s. Participants were required to covertly generate words beginning with the letter during this period. A “Say Words” cue was then presented for 5 s, during which participants were required to overtly report the words they had generated.

**Semantic Matching.** In the semantic matching task (based on the picture version of the Pyramids and Palm Trees test; Howard & Patterson, 1992) a triad of line drawings was presented,



**Figure 5.** Time course of trials in expressive (word generation) and receptive (semantic matching) tasks.



one at the top of the screen and two below. The participant was required to decide which of the two pictures below was the closest semantic match to the one at the top and respond by button press. There were 15 trials. For each trial, after the Clear Mind cue, the participant was presented with eight consecutive picture triads, each lasting 2.5 seconds. Participants reported their decision by keyboard button press using their left or right index fingers. The location of the target picture was counterbalanced so that an equal number of left or right button presses was required. The time course of trials is shown in Figure 5.

#### **Data Acquisition**

The fTCD data were recorded from the left and right MCAs simultaneously using two ultrasound monitoring probes held in place using an elastic headset. The signal was recorded using a Doppler-BoxX receiver and processed using QL software (Version 3.2) on a laptop PC. All equipment (the probes, headset, receiver box and software) were from Compumedics DWL (n.d.). The experimental tasks were presented on a PC monitor using Presentation software (Version 20.0; Neurobehavioral Systems, 2024) which sent marker pulses to the Doppler-BoxX system to denote the onset of each trial.

#### **Data Processing**

The CBFV data from left and right channels were analysed using custom scripts in R Studio (RStudio Team, 2015). The analysis followed the process described by Woodhead et al. (2019). In brief, this included downsampling from 100 Hz to 25 Hz; epoching from  $-12$  s to 30 s peri-stimulus time; artefact rejection (both manual for gross artefacts and automatic for signal intensities outside of the 0.0001–0.9999 quantiles); signal normalisation; heart cycle integration; baseline correction using ten s of rest immediately preceding each trial as a baseline level; a final artefact detection stage where trials containing signal below 60% or above 100% of the mean normalised CBFV were rejected; and averaging of all trials (excluding rejected trials) for each task.

#### **LI Calculation**

For LI calculation we departed from the method used by Bruckert (2016) and adopted the complex GAM approach described by Thompson et al. (2023), which is more comparable to the generalised linear model method used with fMRI. The new  $LI_{fTCD}$  values from the GAM method were used to recategorise individuals as left-lateralised, right-lateralised, or bilateral on each of the two language tasks, depending on whether the 95% CI around the LI estimate was above zero (left), below zero (right), or spanned zero (bilateral).

The LI estimates obtained this way are not confined to the range  $-1$  to  $1$ . To make it possible to visualise  $LI_{fTCD}$  on the same scale as  $LI_{fMRI}$ , values were scaled by dividing the  $LI_{fTCD}$  by 6; this meant that the the largest absolute scaled  $LI_{fTCD}$  value was close to 1.

#### **Data Quality**

Two prespecified criteria were used to check data quality. First, if any participant had more than 20% of trials rejected (i.e., more than five trials for word generation or more than three trials for semantic matching), they would be excluded from the analysis. Second, the trial-by-trial variability was assessed using the standard error of LI values for each trial. The fTCD data were previously analysed as part of a larger sample of 156 participants (Bruckert et al., 2021).

Outlier standard error values were identified in that dataset and excluded from the analysis. No participants were excluded from the current analysis on the basis of these two quality checks.

### **Functional Magnetic Resonance Imaging**

#### ***Procedure***

fMRI was acquired in the second session. Participants were first briefed on the imaging protocol and practiced the tasks outside of the scanner. They were then positioned in the scanner and a structural brain image was acquired. A block design was used. Three tasks were performed in separate runs, each lasting six min. As well as word generation and semantic matching, an auditory naming task was used. This task had no fTCD counterpart, but is considered when looking at agreement between toolbox LI and mirror LI for different ROIs. The order of task presentation was counterbalanced between participants.

**Word Generation.** This task was similar to the fTCD version, but owing to fMRI's greater susceptibility to motion artefacts there was no overt word reporting phase. A block design was used, where the task was performed for 15 s followed by 15 s of rest (with a fixation cross). There were 12 blocks, each with a different letter presented on the screen throughout the duration of the block. Participants were required to covertly think of as many words as they could starting with that letter. Twelve letters (A, B, E, G, J, K, M, N, O, S, U, V) were presented in a randomised order.

**Semantic Matching.** The presentation of picture triad stimuli was similar to the fTCD method. Each fMRI block comprised eight picture triads, each with a duration of 2.5 s (20 s in total). The participants were required to respond by button press using their left and right thumbs on an MRI-compatible button box. Unlike the word generation task, a comparison task (active perceptual baseline) was also acquired during the semantic matching fMRI run. To maximise similarity with fTCD, this comparison task was not included in the current analysis, but further details can be found in [Supplementary Material 1](#).

**Auditory Naming.** The auditory naming paradigm was based on the auditory responsive naming task (Bookheimer et al., 1998) and identical to the one used by Badcock et al. (2012), who adapted this task for the use in fMRI. Participants heard short definitions of a high frequency nouns through MRI compatible Model S14 insert earphones (Sensimetrics, 2024), and were required to silently generate the described word (e.g., the participant heard “shines in the sky” and thinks of “sun”). Because this task used an auditory presentation, which creates substantial activation in auditory cortex, a reversed speech condition (same recordings played backwards) was included, so the effect of auditory stimulation could be controlled for.

#### ***Data Acquisition***

Scanning was performed in a Siemens 3 T Trio scanner with a 32-channel head coil. The task stimuli were presented using Presentation Software (Neurobehavioural Systems) with stimulus onset synchronised with the scanner. The stimuli were projected via a mirror mounted on the head coil.

A high resolution T1-weighted MPRAGE was acquired for image registration (TR = 2,040 ms, TE = 4.7 ms, flip angle = 8°, 192 transverse slices, 1 mm isotropic voxels). Echo-planar images

(EPI) were acquired to measure change in blood oxygen levels during the behavioural tasks (TR = 3 s, TE = 30 ms, flip angle = 90°, 48 axial slices, slice thickness = 3 mm, in-plane resolution = 3 × 3 mm).

### **Data Processing**

Data processing was conducted using FEAT (the fMRI Expert Analysis Tool) in FSL (FMRIB; Analysis Group, n.d.). The preprocessing stages included head motion correction through realignment to the middle volume of the EPI dataset; skull stripping using FSL's Brain Extraction Tool (BET; Smith, 2002); spatial smoothing using a 6 mm full-width-half-maximum Gaussian kernel; high-pass temporal filtering at 90 s; and unwarping using fieldmaps in FSL's Phase Region Expanding Labeller for Unwrapping Discrete Estimates tool (PRELUDE) and FMRIB's Utility for Geometrically Unwarping EPI (FUGUE; Jenkinson, 2003).

The preprocessed data were entered into first-level (subject-specific) general linear models. The different task runs were analysed separately. The explanatory variables (EVs) in the general linear models were: the time course of the active tasks convolved with a double-gamma function; the temporal derivatives of the time course EV; and six motion correction parameters as covariates of no interest. For word generation and semantic matching tasks, the contrast of interest was the implicit (resting) baseline. For auditory naming, the contrast of interest was the backward speech condition. FSL FLIRT (Jenkinson et al., 2002) was used to linearly transform *t*-statistic maps into standard space using the MNI152\_T1\_2mm\_brain template (Muschelli & Fortin, n.d.).

### **LI Calculation**

The following methods for calculating  $LI_{fMRI}$  were compared:

**Conventional Weighted Mean LI From LI Toolbox.** The *t*-statistic maps were used to calculate LI values using the bootstrapping method from the LI Toolbox (Wilke & Lidzba, 2007; Wilke & Schmithorst, 2006) for each task and ROI. Here we used the weighted mean, as described above, which we refer to as the toolbox LI. The LI Toolbox software does not provide a CI for the weighted mean. We computed one by creating a new histogram of LIs, with the proportion of LIs weighted by threshold (see Figure 4 for an example). The mean of these values agrees with the weighted mean computed by the LI Toolbox. The 2.5th and 97.5th percentiles were determined for each weighted histogram to give a 95% CI.

**Mirror Method.** As described above, the right-brain *t*-statistic map is flipped and then subtracted from the left-brain *t*-statistic map, and then 1000 samples, each with 5% of voxels, are sampled without replacement from the difference map. The mean difference score is the LI, and the 2.5th and 97.5th percentiles from the 1,000 estimates give a 95% CI around the LI. The MNI152\_T1\_2mm\_brain template is not perfectly symmetric: 0.53% of voxels on the right have no homolog on the left, and 0.93% of voxels on the left have no homolog on the right. These voxels are ignored when the difference score is computed.

To make it easier to visualise the mirror method data on the same scale as the LI Toolbox data and the  $LI_{fTCD}$ , the differences were divided by the absolute maximum difference in the sample, 2.7, so that the scale maximum was 1.0. This is a linear transform so does not affect correlations. Note, however, that these values are not conventional LIs, as they are simple differences, rather than proportions. We refer to these as scaled mirror LIs.

### **Regions of Interest**

Initial comparisons of methods were done using a combined mask of the frontal, temporal, and parietal lobes as an approximation of the MCA territory. This mask was selected for comparability to the fTCD method, where the LI is based on comparisons of blood flow in left and right MCAs.

The left and right mean activations and LI values were then calculated separately for frontal, temporal, parietal, and cerebellar ROIs, as defined by masks in the LI Toolbox, which are based on a population-based atlas (Hammers et al., 2003; Wilke & Lidzba, 2007). Note that we anticipate that cerebellar lateralisation will be opposite to that seen for other ROIs (Jansen et al., 2006; Seghier et al., 2011).

In all analyses a region 5 mm either side of the midline was excluded from consideration, again using a template provided in the LI Toolbox. This was done for compatibility with the LI Toolbox method, which defaults to masking out midline structures ( $\pm 5$  mm). The LI Toolbox and the mirror method assume  $X = 0$  corresponds to the longitudinal fissure along its length, which may not be true. Also, with typical voxel dimensions and smoothing, signals from the two medial surfaces of the cerebrum are likely to be combined in voxels at this location. The midline of the cerebellum is occupied by the vermis, which is an unpaired structure and so excluded from these calculations.

### **Data Analysis**

We first present averaged  $t$ -statistic maps showing fMRI patterns of activation for the three tasks, averaged for those who were selected as left-lateralised, bilateral, or right-lateralised (based on fTCD word generation). We also present  $t$ -statistic maps used in the mirror method, that is, maps of the L–R  $t$  values obtained by subtraction of the flipped right hemisphere values from the homologous regions of the left hemisphere.

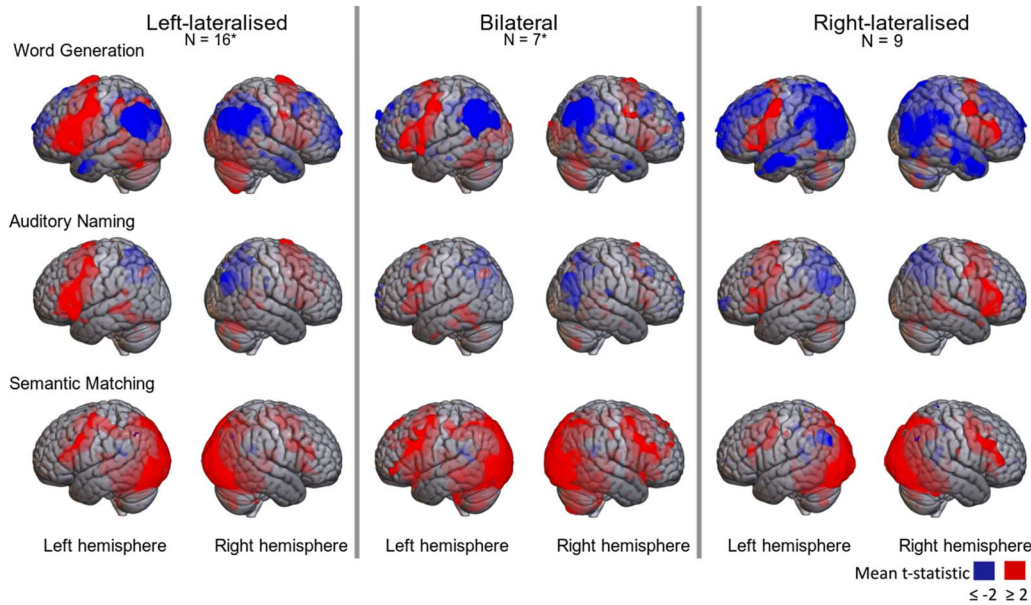
The next set of analyses was conducted to look at the agreement between LI estimates from fMRI and those from fTCD for word generation and semantic matching. The similarity between  $LI_{fMRI}$  and  $LI_{fTCD}$  was assessed using scatterplots and bivariate correlations. As we anticipated that the LI values may not fit a normal distribution, we used Spearman's rank correlations instead of Pearson's correlations. Bootstrapping was used to calculate the 95% CIs around the Spearman correlations (Hervé, 2022).

Finally, we considered agreement between the three tasks and four ROIs for the two methods of computing  $LI_{fMRI}$ .

## **RESULTS**

### **Patterns of Activation**

Patterns of activation from  $t$ -statistic maps are shown in Figure 6 for averaged data from laterality subgroups (left, bilateral, and right) for the three tasks, imaged with MRICroGL (NITRC, n.d.). Note that transparent coloured areas correspond to subcortical activations. These maps clearly show left-sided lateralisation in frontal regions for the left-lateralised group, which is most pronounced for word generation and auditory naming. A more mixed picture is seen for the bilateral and right-lateralised groups. It is clear from these maps that positive activation in language regions with production tasks is accompanied by deactivation (negative  $t$ -values) in the parietal region.

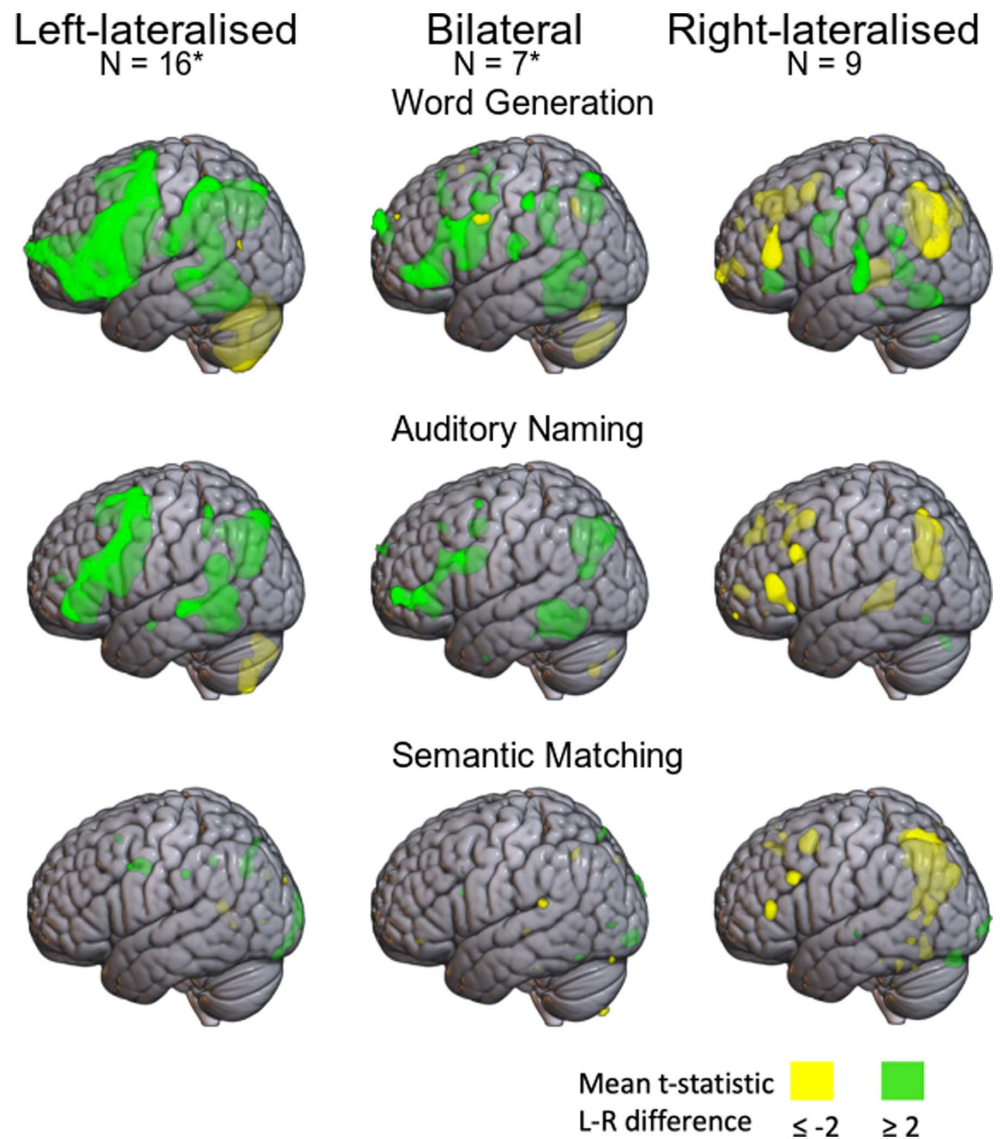


**Figure 6.** Heatmaps of  $t$  statistic maps for the three tasks superimposed on SPM152 brain; averages for group defined by functional transcranial Doppler ultrasound lateralality on word generation. \*Missing data reduces  $N$  by one for auditory naming bilateral group, and for semantic matching, left-lateralised group.

Figure 7 shows the subtracted  $t$ -statistic maps used in the mirror method, with lateralisation group based on fTCD word generation. Here, green indicates left-lateralisation, and yellow right-lateralisation. The left-lateralised group shows both frontal and posterior left-lateralisation (green) on word generation and auditory naming; note that the posterior left-lateralisation reflects the fact that negative  $t$ -values are seen in both left and right parietal regions in these tasks, but the negativity is more extreme in the right hemisphere, so the difference score favours the left side. The bilateral group also appears left-lateralised on these tasks, but with lateralisation confined to smaller regions. The right-lateralised group shows right-lateralisation (yellow) in frontal and posterior regions, but again lateralisation is less extensive than for the typically lateralised group. Note, however, that the averages for the bilateral and right-lateralised groups are based on smaller samples. These maps also emphasise the lack of lateralisation on semantic matching in the left-lateralised and bilateral groups.

#### Correspondence of LI in the MCA

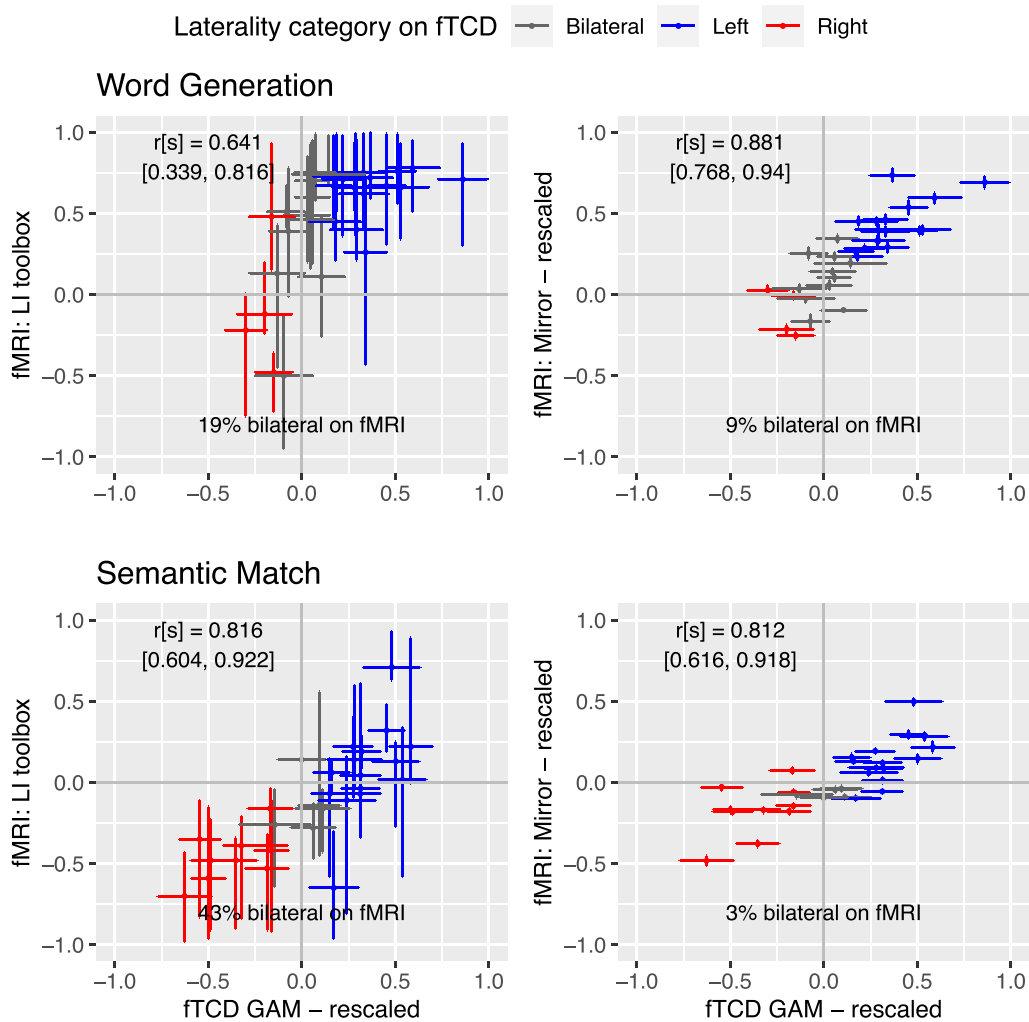
Figure 8 shows the relationship between fTCD lateralality ( $x$  axis) and fMRI lateralality ( $y$  axis) for the original LI Toolbox bootstrap method in the left panels and the new mirror method in the right panels. The upper panels show results for word generation, and the lower panels for semantic matching. Points are colour-coded according to categorical lateralality classification on  $LI_{fTCD}$  for that task: grey points are those where the 95% CI crosses zero, and are hence coded bilateral. Points in red are right-lateralised, and those in blue are left-lateralised. Note that the high proportion of bilateral/right-lateralised cases reflects the fact that the sample was selected to include 50% atypically lateralised individuals. The plot also shows the Spearman correlation between fTCD and fMRI lateralality, with the 95% CI. The percentage of individuals identified as having bilateral language on fMRI is also shown, again using the criterion of whether the 95% CI of the estimate crosses zero. Not shown in the plots is the correlation



**Figure 7.** Heatmaps of difference  $t$  statistic maps used in the mirror method for the three tasks superimposed on SPM152 brain; averages for left-lateralised, bilateral, and right-lateralised groups (defined by functional transcranial Doppler ultrasound word generation). Green denotes left-lateralised voxels, and yellow denotes right-lateralised voxels.

between laterality estimates from the two fMRI methods: the Spearman correlation [95% CI] between mirror and toolbox methods was 0.759 [0.526, 0.871] for word generation and 0.858 [0.685, 0.939] for semantic matching.

Two points are immediately evident from the plots. First, with few exceptions, fMRI and fTCD categorisations agree fairly well, that is, those categorised as right-lateralised on fTCD tend to have LI values below zero on fMRI, and those categorised as left-lateralised on fTCD tend to have LI values above zero on fMRI. Second, the 95% CIs for individual LI estimates are considerably larger for the bootstrapped LI Toolbox estimates than for the mirror method: this has the consequence that a far higher proportion of toolbox estimates categorise the individual as having bilateral language.



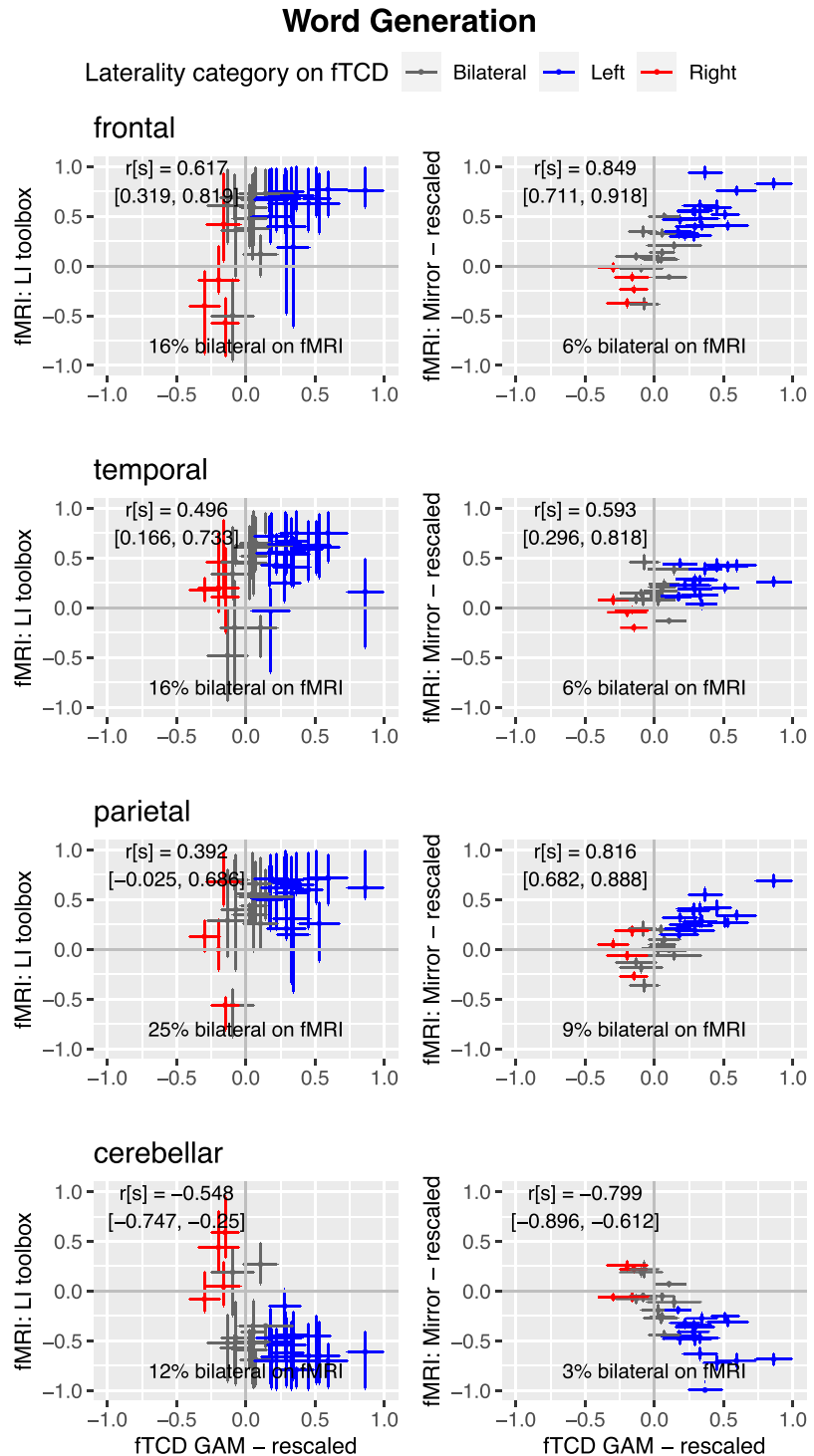
**Figure 8.** Scatterplots of relationship between functional transcranial Doppler ultrasound (fTCD) and two methods of fMRI laterality for middle cerebral artery region. Spearman correlations with 95% CIs are shown as annotations. GAM = generalised additive model.

We will discuss possible explanations of these results below, but first we turn to look at agreement between LIs for the four ROIs outlined above. (See [Supplementary Material 2](#) for details of the mean voxel activations on left and right for each region that are the basis for the LI calculations.)

Figure 9 shows correspondence of each fMRI laterality method with fTCD LI for each ROI for word generation. In general, the correlation is stronger for the mirror  $LI_{fMRI}$  than for toolbox  $LI_{fMRI}$ , with the absolute value of the mirror  $LI_{fMRI}$  correlation being outside the 95% CI of that for the toolbox  $LI_{fMRI}$  correlation for all but the temporal ROI. The difference is most striking for the parietal region. The correlations with fTCD laterality are positive, except for the cerebellar region, where a negative association is seen.

Figure 10 shows correspondence of each fMRI laterality method with fTCD LI for each ROI for semantic matching. The pattern of correlations is more similar for the two methods for this task, and it is noteworthy that for the cerebellum, the CIs for the correlations with  $LI_{fTCD}$  include zero for both measures.

For both tasks, as was seen with the MCA ROI, the proportion of individuals categorised with bilateral language is much smaller with the mirror method than with the toolbox method, reflecting the large CIs associated with the toolbox estimates. In other words, people who appear to have bilateral language with the toolbox method appear reliably lateralised to left or right with the mirror method.

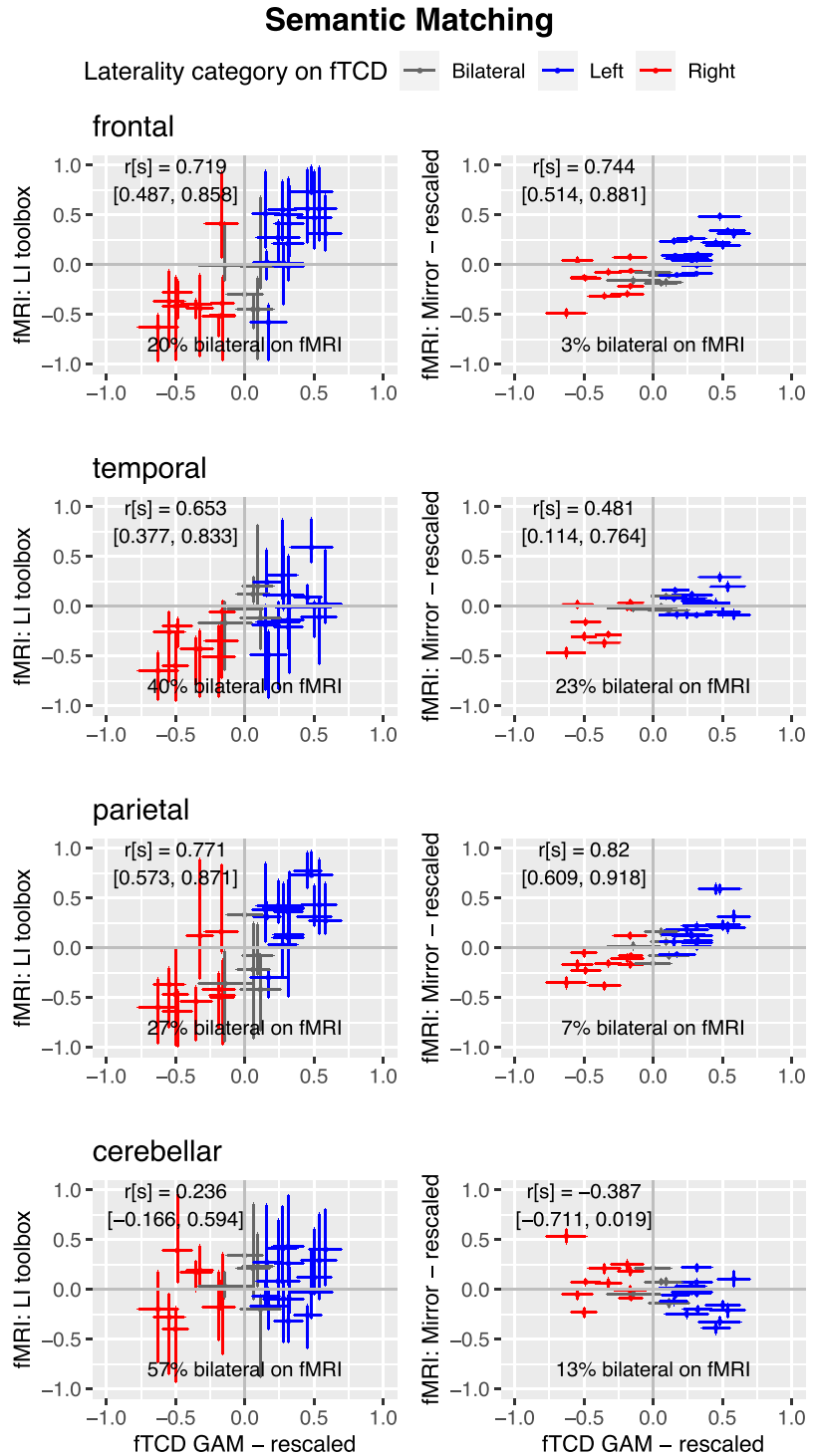


**Figure 9.** Scatterplots of relationship for word generation between fTCD and two methods of fMRI laterality for four regions of interest.

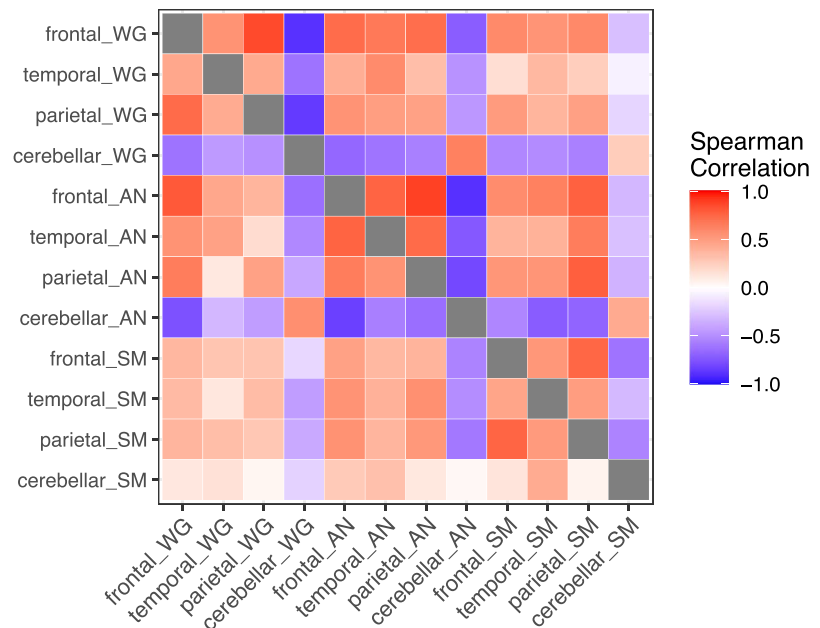


**Consistency of Laterality Across Tasks and ROIs**

Spearman correlations between LIs from the three tasks and four ROIs are shown in Figure 11. Considering first the within-task correlations for the mirror method, one can see that for all



**Figure 10.** Scatterplots of relationship for semantic matching between fTCD and two methods of fMRI laterality for four regions of interest.



**Figure 11.** Heatmap of Spearman correlations between LIs for three tasks and four regions of interest. Mirror method is above the diagonal, and toolbox method below the diagonal. Tasks denoted by WG (word generation), AN (auditory naming), and SM (semantic matching).

three tasks, there is a positive correlation between frontal and parietal LIs, and a negative correlation between frontal/parietal and cerebellar LIs. Correlations with temporal regions are generally similar in direction to frontal correlations, but weaker. In addition, LIs for a given ROI are generally positive across the three tasks. Similar trends are seen for the toolbox LIs, but correlations tend to be a bit smaller in magnitude.

## DISCUSSION

Estimates of  $LI_{fMRI}$  using the mirror method generally agree well with  $LI_{fTCD}$ , whereas the toolbox LI has somewhat weaker correlations. It is important to note that the  $LI_{fTCD}$  cannot be regarded as a gold standard, as it does not have any sensitivity to localisation of activation. One possible explanation for the better agreement with mirror  $LI_{fMRI}$  is that the data processing is more similar to that in fTCD, involving simple subtraction of measures of activation on the left and right.

There is, however, another reason why the toolbox LI may give weaker agreement with  $LI_{fTCD}$ : Like virtually all methods of computing LI from fMRI, the toolbox  $LI_{fMRI}$  ignores task-related decreases in activation. This is justified by a simple assumption, stated by Wilke and Lidzba (2007) as: “‘interesting’ data will be of above-average intensity.”

### Importance of Task-Related Deactivation

Figure 6 suggests that language laterality may depend on deactivation as well as activation. Indeed, this is already suggested by scrutiny of the mean task-related blood flow in the fTCD data for the word generation task (Figure 1). We first see an increase in blood flow to both hemispheres as the participant starts to perform the task, but then blood flow declines in both hemispheres, with the right hemisphere values dropping below the baseline level.

Our fMRI results have some similarities with those of Seghier et al. (2011), who considered how a global LI related to activation of individual voxels in left and right hemispheres. Using a written semantic matching task, they found negative correlations between global LI and activations in some right hemisphere regions—that is, the stronger the overall left-lateralisation, the lower the right-sided parameter values. This was not discussed in terms of deactivation, but could reflect a similar phenomenon as we report here. Task-related deactivation has been described for language tasks by Jansen et al. (2006), who noted that it poses problems for calculation of a conventional LI, because it can lead to values greater than 1. For instance, if mean activation is 2 for a left ROI and is  $-1$  for a right ROI, then the computed LI is  $3/1 = 3$ . If left and right ROIs have similar absolute activations opposite in polarity, the LI approaches infinity. Quite simply, the conventional LI is a proportional measure that is not appropriate for negative values. This means that we have to either ignore deactivation or abandon the use of the conventional LI.

We suggest that deactivation appears to be an important feature in the brain's lateralised response. The brain maps in Figure 6 show that deactivation is particularly striking in the parietal lobe. The quantitative analyses of average left and right activation in different ROIs in Supplementary Material 2 confirm that, even for those with typical language lateralisation, we see average decrease of activation below zero in both left and right hemispheres, with the deactivation being more marked on the right. Furthermore, as can be seen in Figure 11, there are positive correlations between LI in the frontal and parietal ROIs, suggesting a common mechanism determines the extent of lateralisation. There is also a negative correlation with the frontal/parietal LIs and that obtained from the cerebellar ROI; this is evident in all three tasks when the mirror method is used.

If, as seems to be the case, functional lateralisation for language is manifest in deactivation of some regions, then we will fail to detect this phenomenon by relying on an LI that focuses only on voxels with suprathreshold activation.

The mechanism underlying such phenomena is not understood, though we know that the circle of Willis allows for compensatory changes in blood flow between posterior and anterior cerebral blood vessels, and between the two sides of the brain (Payne, 2017). It is possible that task-evoked neural activity that results in increased flow of oxygenated blood to the left frontal lobe will lead to a corresponding decrease in blood flow in posterior regions and in the right hemisphere relative to the left. Indeed, individual differences in cerebral lateralisation might be related to individual differences in the vascular system—in a significant minority of people the circle of Willis is incomplete (Kızılgöz et al., 2022; Payne, 2017); our data raise the intriguing possibility that this might have consequences for measures of functional asymmetry that rely on haemodynamic changes as a proxy for neuronal activity.

### **Categorisation of Bilateral Language**

Another difference between the two methods for computing  $LI_{fMRI}$  is in the size of CIs around LI estimates, which are larger for the toolbox LI. It follows that if bilateral language is categorised using the CI around the mean estimate, then a high proportion of cases will be classed as bilateral using the toolbox method. The CIs around the mirror LI estimates are surprisingly small, given that they were based on relatively small samples of voxels in the difference map within a region. If we look only at averaged maps, such as in Figure 6, we may get the impression that those with atypical lateralisation have similar activation on left and right sides of the brain during language tasks. The mirror LI estimates suggest that averaged maps may obscure the fact that most individuals have a significant bias to left- or right-sided language, even for a weakly lateralised task such as semantic matching.

Development of the LI Toolbox was a major breakthrough in laterality measurement from fMRI, providing a serious attempt to tackle the inconsistent and arbitrary methods that were being used to quantify individual differences. The method allows one to compute a fast and reliable measure of  $LI_{fMRI}$  that has been shown to be robust to outliers, and is supported by open and thoroughly documented code. This makes it straightforward to explore how different settings of parameters can influence the LI estimates. The wide CIs obtained with toolbox estimates do not reflect instability of the measures; rather, they reflect the fact that when a large set of possible thresholds is considered, the range of  $LI_{fMRI}$  estimates compatible with the observed data is large. This forces us back to the question of what is the best criterion for selecting a threshold, and we run into a difficult choice: A weighted mean is useful for giving most weight to the strongest evidence, but at the same time it means we rely most heavily on the least reliable estimates, based on a small number of voxels. In addition, if we take either the trimmed or weighted mean value from the distribution of LIs, this will be influenced by the range of  $t$  values used to compute thresholds. Ultimately, while it is good to have an agreed convention of how to compute the  $LI_{fMRI}$ , there remains some arbitrariness in how this is done.

Other studies have compared different approaches to estimation of laterality from fMRI data (see, e.g., Abbott et al., 2010; Brumer et al., 2020; Fesl et al., 2010), but, as far as we are aware, they all use the classic LI formula,  $(L - R) / (L + R)$ , varying in how ROI or thresholds are determined, and whether L and R are number of activated voxels, or summed activation on left and right; in general, therefore, they do not take deactivation into account.

### **Limitations**

The mirror method is made possible by the fact that the MNI152\_T1\_2mm\_brain template, which was used to linearly transform all  $t$  maps, is nearly symmetrical, and so there are corresponding voxels on left and right for most brain regions. This does not, however, mean that the regions on left and right are equivalent: we know already that in some regions there are not only functional differences between the two sides, but also anatomical differences, in cerebral vasculature (Dubois & Adolphs, 2016), in gyral and sulcal patterns (e.g., Marie et al., 2015), in torque (e.g., Kong et al., 2021), and in white matter tracts (Forkel et al., 2022) and distributions of regional grey matter (e.g., Carrion-Castillo et al., 2020; Zhao et al., 2022). The approach adopted here of transforming to a standard brain and dropping voxels with no mirror counterpart ignores individual differences in lateralised morphology. There might be interest in future in taking such differences into account, for instance, by conducting the analysis without transforming to a standard brain, and instead using nonlinear transforms to align the two sides according to specific anatomic features. Such an approach is beyond the scope of the current analysis, but could be useful for indicating whether specific anatomical variations are related to functional lateralisation.

We have focused largely on estimating parameters and have presented few statistical tests here because we lacked a priori hypotheses and with 32 participants had limited statistical power to detect differences between methods, ROIs, and tasks. This study may be regarded as hypothesis-generating rather than hypothesis-testing, and may be used as the basis for making predictions in future analyses of new datasets. We present some specific predictions below.

### **Future Directions**

In the course of comparing left- and right-sided activation associated with language generation, we noted that differences between hemispheres are not always driven by positive

activation; in parietal lobes in particular, left-sided language processing tended to reflect reduced activity in both hemispheres, which is more pronounced on the right side. An unexpected benefit of our novel method, the mirror method, is that it allows us to capture this source of asymmetry without needing to specify any thresholds. This could be useful in future when comparing extent of lateralisation across different tasks or different ROIs, and might cast light on the underlying mechanisms of functional lateralisation.

This study has generated predictions which could be tested in existing datasets as well as in novel studies:

*Prediction 1:* The pattern of activation and deactivation of left and right hemispheres in different regions seen in Figure 6 will replicate, with tasks involving language generation showing strong left hemisphere bias in all regions, but with a gradient from overall activation in frontal regions to overall deactivation in parietal regions.

*Prediction 2:* Cerebellar activation will show a complementary pattern of lateralisation to frontal activation for tasks involving language generation, but not for receptive language tasks.

*Prediction 3:* Most individuals who are categorised as having bilateral language on conventional fMRI LIs will show significant bias to left or right when assessed using the mirror method. In other words, most cases of bilateral language are probably cases of weak lateralisation which is hard to detect because of measurement error, rather than equal involvement of both hemispheres.

*Prediction 4:* Functional cerebral asymmetry may be related to anatomical differences in the circle of Willis (Kızılgöz et al., 2022; Nyasa et al., 2021; Payne, 2017).

#### ACKNOWLEDGMENTS

We are most grateful to Lisa Bruckert for making data from her doctoral thesis available to us for this study, and to Marko Wilke for helpful discussions about the LI Toolbox. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

#### FUNDING INFORMATION

Dorothy V. M. Bishop, H2020 European Research Council (<https://dx.doi.org/10.13039/100010663>), Award ID: 694189.

#### AUTHOR CONTRIBUTIONS

**Dorothy V. M. Bishop:** Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Methodology: Equal; Project administration: Equal; Resources: Supporting; Software: Lead; Supervision: Equal; Writing – original draft: Lead; Writing – review & editing: Lead. **Zoe V. J. Woodhead:** Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Supporting; Project administration: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting. **Kate E. Watkins:** Conceptualization: Equal; Data curation: Supporting; Funding acquisition: Supporting; Investigation: Supporting; Methodology: Supporting; Resources: Lead; Supervision: Equal; Writing – review & editing: Supporting.

## DATA AND CODE AVAILABILITY STATEMENT

Locations of data and analysis scripts are specified on Open Science Framework: <https://osf.io/f6uyt/>.

## REFERENCES

- Abbott, D. F., Waites, A. B., Lillywhite, L. M., & Jackson, G. D. (2010). fMRI assessment of language lateralization: An objective approach. *NeuroImage*, *50*(4), 1446–1455. <https://doi.org/10.1016/j.neuroimage.2010.01.059>, PubMed: 20097290
- Analysis Group. (n.d.). FEAT [Software]. FMRIB Software Library. <https://www.fmrib.ox.ac.uk/fsl/fslwiki/FEAT>
- Baciu, M., Juphard, A., Cousin, E., & Le Bas, J. F. (2005). Evaluating fMRI methods for assessing hemispheric language dominance in healthy subjects. *European Journal of Radiology*, *55*(2), 209–218. <https://doi.org/10.1016/j.ejrad.2004.11.004>, PubMed: 16036149
- Badcock, N. A., Bishop, D. V. M., Hardiman, M. J., Barry, J. G., & Watkins, K. E. (2012). Co-localisation of abnormal brain structure and function in specific language impairment. *Brain and Language*, *120*(3), 310–320. <https://doi.org/10.1016/j.bandl.2011.10.006>, PubMed: 22137677
- Berker, E. A., Berker, A. H., & Smith, A. (1986). Translation of Broca's 1865 report. Localization of speech in the third left frontal convolution. *Archives of Neurology*, *43*(10), 1065–1072. <https://doi.org/10.1001/archneur.1986.00520100069017>, PubMed: 3530216
- Bishop, D. V. M. (2013). Cerebral asymmetry and language development: Cause, correlate, or consequence? *Science*, *340*(6138), Article 1230531. <https://doi.org/10.1126/science.1230531>, PubMed: 23766329
- Bishop, D. V. M., Holt, G., Whitehouse, A. J. O., & Groen, M. (2014). No population bias to left-hemisphere language in 4-year-olds with language impairment. *PeerJ*, *2*, Article e507. <https://doi.org/10.7717/peerj.507>, PubMed: 25165624
- Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T. A., Gaillard, W. D., Malow, B., & Theodore, W. H. (1998). Regional cerebral blood flow during auditory responsive naming: Evidence for cross-modal neural activation. *NeuroReport*, *9*(10), 2409–2413. <https://doi.org/10.1097/00001756-199807130-00047>, PubMed: 9694237
- Bradshaw, A. R., Bishop, D. V. M., & Woodhead, Z. V. J. (2017). Methodological considerations in assessment of language lateralisation with fMRI: A systematic review. *PeerJ*, *5*, Article e3557. <https://doi.org/10.7717/peerj.3557>, PubMed: 28713656
- Bruckert, L. (2016). *Is language laterality related to language abilities?* [Doctoral dissertation]. University of Oxford. <https://ora.ox.ac.uk/objects/uuid:05e80d0d-8d0b-4cb2-8f94-22763603fab5>
- Bruckert, L., Thompson, P. A., Watkins, K. E., Bishop, D. V. M., & Woodhead, Z. V. J. (2021). Investigating the effects of handedness on the consistency of lateralization for speech production and semantic processing tasks using functional transcranial Doppler sonography. *Laterality*, *26*(6), 680–705. <https://doi.org/10.1080/1357650X.2021.1898416>, PubMed: 33715589
- Brumer, I., De Vita, E., Ashmore, J., Jarosz, J., & Borri, M. (2020). Implementation of clinically relevant and robust fMRI-based language lateralization: Choosing the laterality index calculation method. *PLOS One*, *15*(3), Article e0230129. <https://doi.org/10.1371/journal.pone.0230129>, PubMed: 32163517
- Carrion-Castillo, A., Pepe, A., Kong, X.-Z., Fisher, S. E., Mazoyer, B., Tzourio-Mazoyer, N., Crivello, F., & Francks, C. (2020). Genetic effects on planum temporale asymmetry and their limited relevance to neurodevelopmental disorders, intelligence or educational attainment. *Cortex*, *124*, 137–153. <https://doi.org/10.1016/j.cortex.2019.11.006>, PubMed: 31887566
- Compumedics DWL. (n.d.). *Doppler-BoxX* (Receiver box), *QL* (Version 3.2; Software), probes, headset [Apparatus]. <https://www.dwl.de/en/>
- Cousin, E., Peyrin, C., Pichat, C., Lamalle, L., Le Bas, J.-F., & Baciu, M. (2007). Functional MRI approach for assessing hemispheric predominance of regions activated by a phonological and a semantic task. *European Journal of Radiology*, *63*(2), 274–285. <https://doi.org/10.1016/j.ejrad.2007.01.030>, PubMed: 17339089
- Deppe, M., Knecht, S., Henningsen, H., & Ringelstein, E. B. (1997). AVERAGE: A Windows program for automated analysis of event related cerebral blood flow. *Journal of Neuroscience Methods*, *75*(2), 147–154. [https://doi.org/10.1016/S0165-0270\(97\)00067-8](https://doi.org/10.1016/S0165-0270(97)00067-8), PubMed: 9288646
- Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in Cognitive Sciences*, *20*(6), 425–443. <https://doi.org/10.1016/j.tics.2016.03.014>, PubMed: 27138646
- Fesl, G., Bruhns, P., Rau, S., Wiesmann, M., Ilmberger, J., Kegel, G., & Brueckmann, H. (2010). Sensitivity and reliability of language laterality assessment with a free reversed association task: A fMRI study. *European Radiology*, *20*(3), 683–695. <https://doi.org/10.1007/s00330-009-1602-4>, PubMed: 19789883
- Forkel, S. J., Friedrich, P., Thiebaut de Schotten, M., & Howells, H. (2022). White matter variability, cognition, and disorders: A systematic review. *Brain Structure and Function*, *227*(2), 529–544. <https://doi.org/10.1007/s00429-021-02382-w>, PubMed: 34731328
- Hammers, A., Allom, R., Koeppe, M. J., Free, S. L., Myers, R., Lemieux, L., Mitchell, T. N., Brooks, D. J., & Duncan, J. S. (2003). Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*, *19*(4), 224–247. <https://doi.org/10.1002/hbm.10123>, PubMed: 12874777
- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015. <https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x>, PubMed: 11095242
- Hervé, M. (2022). *RVAideMemoire: Testing and plotting procedures for biostatistics* [Software]. <https://CRAN.R-project.org/package=RVAideMemoire>
- Howard, D., & Patterson, K. (1992). *The pyramids and palm trees test: A test for semantic access from words and pictures*. Thames Valley Test Publishers.
- Jansen, A., Menke, R., Sommer, J., Förster, A. F., Bruchmann, S., Hempleman, J., Weber, B., & Knecht, S. (2006). The assessment of hemispheric lateralization in functional MRI—Robustness and reproducibility. *NeuroImage*, *33*(1), 204–217. <https://doi.org/10.1016/j.neuroimage.2006.06.019>, PubMed: 16904913
- Jenkinson, M. (2003). Fast, automated, N-dimensional phase-unwrapping algorithm. *Magnetic Resonance in Medicine*, *49*(1), 193–197. <https://doi.org/10.1002/mrm.10354>, PubMed: 12509838

- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825–841. <https://doi.org/10.1006/nimg.2002.1132>, PubMed: 12377157
- Johnstone, L. T., Karlsson, E. M., & Carey, D. P. (2020). The validity and reliability of quantifying hemispheric specialisation using fMRI: Evidence from left and right handers on three different cerebral asymmetries. *Neuropsychologia*, 138, Article 107331. <https://doi.org/10.1016/j.neuropsychologia.2020.107331>, PubMed: 31917204
- Kızılgöz, V., Kantarcı, M., & Kahraman, Ş. (2022). Evaluation of the Willis variants using magnetic resonance angiography. *Scientific Reports*, 12, Article 17611. <https://doi.org/10.1038/s41598-022-21833-w>, PubMed: 36266391
- Kong, X.-Z., Postema, M., Schijven, D., Carrión-Castillo, A. C., Pepe, A., Crivello, F., Joliot, M., Mazoyer, B., Fisher, S. E., & Francks, C. (2021). Large-scale phenomic and genomic analysis of brain asymmetrical skew. *Cerebral Cortex*, 31(9), 4151–4168. <https://doi.org/10.1093/cercor/bhab075>, PubMed: 33836062
- Marie, D., Jobard, G., Crivello, F., Perchey, G., Petit, L., Mellet, E., Joliot, M., Zago, L., Mazoyer, B., & Tzourio-Mazoyer, N. (2015). Descriptive anatomy of Heschl's gyri in 430 healthy volunteers, including 198 left-handers. *Brain Structure and Function*, 220(2), 729–743. <https://doi.org/10.1007/s00429-013-0680-x>, PubMed: 24310352
- Muschelli, J., & Fortin, J.-P. (n.d.). *MNI Template* [Software]. <https://neuroconductor.org/help/MNI Template/>
- Neurobehavioral Systems. (2024). *Presentation* (Version 20.0) [Software]. <https://www.neurobs.com/>
- [NITRC] NeuroImaging Tools & Resources Collaboratory. (n.d.). *MRICroGL* (Version 14.4.1) [Software]. <https://www.nitrc.org/projects/mricrogl/>
- Nyasa, C., Mwakikunga, A., Tembo, L., Dzamalala, C., & Ihunwo, A. O. (2021). Distribution of variations in anatomy of the circle of Willis: Results of a cadaveric study of the Malawian population and review of literature. *Pan African Medical Journal*, 38, Article 11. <https://doi.org/10.11604/pamj.2021.38.11.27126>, PubMed: 34567338
- Parker, A. J., Woodhead, Z. V. J., Carey, D. P., Groen, M. A., Gutierrez-Sigut, E., Hodgson, J., Hudson, J., Karlsson, E. M., MacSweeney, M., Payne, H., Simpson, N., Thompson, P. A., Watkins, K. E., Egan, C., Grant, J. H., Harte, S., Hudson, B. T., Sablik, M., Badcock, N. A., & Bishop, D. V. M. (2022). Inconsistent language lateralisation: Testing the dissociable language laterality hypothesis using behaviour and lateralised cerebral blood flow. *Cortex*, 154, 105–134. <https://doi.org/10.1016/j.cortex.2022.05.013>, PubMed: 35777191
- Payne, S. (2017). *Cerebral blood flow and metabolism: A quantitative approach*. World Scientific. <https://doi.org/10.1142/10463>
- RStudio Team. (2015). *RStudio: Integrated development for R* (Version 0.98.1074) [Software]. <https://www.rstudio.com/>
- Seghier, M. L. (2008). Laterality index in functional MRI: Methodological issues. *Magnetic Resonance Imaging*, 26(5), 594–601. <https://doi.org/10.1016/j.mri.2007.10.010>, PubMed: 18158224
- Seghier, M. L., Kherif, F., Josse, G., & Price, C. J. (2011). Regional and hemispheric determinants of language laterality: Implications for preoperative fMRI. *Human Brain Mapping*, 32(10), 1602–1614. <https://doi.org/10.1002/hbm.21130>, PubMed: 20814960
- Sensimetrics. (2024). *Model S14* earphones [Apparatus]. <https://shop.sens.com/products/s14>
- Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3), 143–155. <https://doi.org/10.1002/hbm.10062>, PubMed: 12391568
- Thompson, P. A., Watkins, K. E., Woodhead, Z. V. J., & Bishop, D. V. M. (2023). Generalized models for quantifying laterality using functional transcranial Doppler ultrasound. *Human Brain Mapping*, 44(1), 35–48. <https://doi.org/10.1002/hbm.26138>, PubMed: 36377321
- Vingerhoets, G., Verhelst, H., Gerrits, R., Badcock, N. A., Bishop, D. V. M., Carey, D. P., Flindall, J., Grimshaw, G., Harris, L. J., Hausmann, M., Hirnstein, M., Jäncke, L., Joliot, M., Specht, K., Westerhausen, R., & LIC Consortium. (2023). Laterality indices consensus initiative (LICI): A Delphi expert survey report on recommendations to record, assess, and report asymmetry in human behavioural and brain research. *Laterality*, 28(2–3), 122–191. <https://doi.org/10.1080/1357650X.2023.2199963>, PubMed: 37211653
- Wada, J., & Rasmussen, T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance. *Journal of Neurosurgery*, 17(2), 266–282. <https://doi.org/10.3171/jns.1960.17.2.0266>
- Watkins, K. E., Paus, T., Lerch, J. P., Zijdenbos, A., Collins, D. L., Neelin, P., Taylor, J., Worsley, K. J., & Evans, A. C. (2001). Structural asymmetries in the human brain: A voxel-based statistical analysis of 142 MRI scans. *Cerebral Cortex*, 11(9), 868–877. <https://doi.org/10.1093/cercor/11.9.868>, PubMed: 11532891
- Wilke, M., & Lidzba, K. (2007). LI-tool: A new toolbox to assess lateralization in functional MR-data. *Journal of Neuroscience Methods*, 163(1), 128–136. <https://doi.org/10.1016/j.jneumeth.2007.01.026>, PubMed: 17386945
- Wilke, M., & Schmithorst, V. J. (2006). A combined bootstrap/histogram analysis approach for computing a lateralization index from neuroimaging data. *NeuroImage*, 33(2), 522–530. <https://doi.org/10.1016/j.neuroimage.2006.07.010>, PubMed: 16938470
- Woodhead, Z. V. J., Bradshaw, A. R., Wilson, A. C., Thompson, P. A., & Bishop, D. V. M. (2019). Testing the unitary theory of language lateralization using functional transcranial Doppler sonography in adults. *Royal Society Open Science*, 6(3), Article 181801. <https://doi.org/10.1098/rsos.181801>, PubMed: 31032035
- Woodhead, Z. V. J., Rutherford, H. A., & Bishop, D. V. M. (2020). *Measurement of language laterality using functional transcranial Doppler ultrasound: A comparison of different tasks* (Version 3). Wellcome Open Research. <https://wellcomeopenresearch.org/articles/3-104/v3>
- Zhao, L., Matloff, W., Shi, Y., Cabeen, R. P., & Toga, A. W. (2022). Mapping complex brain torque components and their genetic architecture and phenomic associations in 24,112 individuals. *Biological Psychiatry*, 91(8), 753–768. <https://doi.org/10.1016/j.biopsych.2021.11.002>, PubMed: 35027165