THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Heritability of language laterality assessed by functional transcranial Doppler ultrasound

OPEN ACCESS

Check for updates

RESEARCH ARTICLE

REVISED

Heritability of language laterality assessed by functional transcranial Doppler ultrasound: a twin study [version 3; peer review: 3 approved]

Previously titled: Negligible heritability of language laterality assessed by functional transcranial Doppler ultrasound: a twin study

Dorothy V.M. Bishop [iD][1], Timothy C. Bates [iD][2]

[1]Department of Experimental Psychology, University of Oxford, Oxford, Oxon, OX2 6GG, UK
[2]Psychology Department, University of Edinburgh, Edinburgh, EH8 9JZ, UK

Abstract

Background: Prior studies have estimated heritability of around 0.25 for the trait of handedness, with studies of structural brain asymmetry giving estimates in a similar or lower range. Little is known about heritability of functional language lateralization. This report describes heritability estimates using functional language laterality and handedness phenotypes in a twin sample previously reported by Wilson and Bishop (2018).

Methods: The total sample consisted of 194 twin pairs (49% monozygotic) aged from 6 to 11 years. A language laterality index was obtained for 141 twin pairs, who completed a protocol where relative blood flow through left and right middle cerebral arteries was measured using functional transcranial Doppler ultrasound (fTCD) while the child described animation sequences. Handedness data was available from the Edinburgh Handedness Inventory (EHI) and Quantification of Hand Preference (QHP) for all 194 pairs. Heritability was assessed using conventional structural equation modeling, assuming no effect of shared environment (AE model).

Results: For the two handedness measures, heritability estimates (95% CI) were consistent with prior research: .25 (.03 - .34) and .18 (0 – .31) respectively for the EHI and QHP. For the language laterality index, however, the twin-cotwin correlations were close to zero for both MZ and DZ twins, and the heritability estimate was zero (0 - .15).

Conclusions: A single study cannot rule out a genetic effect on language lateralisation. It is possible that the low twin-cotwin correlations were affected by noisy data: although the split-half reliability of the fTCD-based laterality index was high (0.85), we did not have information on test-retest reliability in children, which is likely to

Open Peer Review

Reviewer Status ✔ ✔ ✔

	Invited Reviewers		
	1	**2**	**3**
version 3 (revision) 16 Sep 2020			✔ report
version 2 (revision) 11 Mar 2020	✔ report	✔ report	? report
version 1 28 Oct 2019	? report	? report	? report

1. **Clyde Francks** [iD], Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands Radboud University, Nijmegen, The Netherlands

2. **Guy Vingerhoets** [iD], Ghent University, Ghent, Belgium

3. **Chris McManus** [iD], University College London, London, UK University College London, London, UK

be lower. We cannot reject the hypothesis that there is low but nonzero heritability for this trait, but our data suggest that individual variation in language lateralisation is predominantly due to stochastic variation in neurodevelopment.

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
laterality, language, genetics, twins, handedness, language lateralisation, functional transcranial Doppler ultrasound

Introduction

Lateralisation of language and motor function in humans display two notable features: First, there is a pronounced population bias - to the left hemisphere for lateralisation of language, and to the right hand (controlled by the left hemisphere) for handedness; second there is individual variation in the direction and extent of lateralisation; a minority of individuals show a reversal of the typical pattern, and others show little or no asymmetry. The frequency of these atypical patterns will depend on how they are measured and defined: Vingerhoets (2019) estimated 6.5% of people have right language dominance and 10–15% have bilateral representation of language.

Other primates show some indications of lateralised functions, but humans are distinctive in their strong population bias to right-handedness. It is often assumed that evolution of manual laterality is related to development of a complex and lateralised language faculty in humans, but the biological origins of the population bias are not well understood for either trait. There is an association between handedness and language lateralisation, evident from a range of methods, including the impact of focal lesions, presurgical testing for language dominance in epilepsy, and imaging methods in

healthy populations, but it is complex: whereas around 83–88% of right-handers have left-hemisphere language, this is true for around 64–68% of non-right-handers (see Carey & Johnstone, 2014, for a comprehensive meta-analysis including a range of methodologies for assessing language laterality).

Twins provide a useful natural experiment for estimating the contribution of genetic variation to individual differences in a trait. The twin method compares the similarity of identical (monozygotic or MZ) twins versus non-identical (dizygotic or DZ) twins to derive estimates of the relative contributions of genetic variants, environment shared by the twins, and other twin-specific influences (including chance) on individual variation in a trait. This is best understood intuitively by imagining a variety of hypothetical situations. In the first, the trait is solely determined by random chance: in that case, two members of a twin pair (A and B) would be no more similar than two unrelated people, and in a sample of twins, the correlation between twin A and twin B would be zero, regardless of zygosity. In a second fictitious scenario, the trait is solely determined by an environmental factor common to both twins, such as home environment. In that case, there would be perfect correlation between the traits for twin A and twin B, regardless of zygosity. In the third scenario, the trait is determined solely by genes: because MZ twins are genetically identical, the correlation between two members of a twin pair will be 1, but for DZ twins, who on average have 50% of their segregating genes in common, the correlation will be 0.5.

In practice, our goal is not to attribute variation in the observed trait (phenotype) to one cause or the other, but rather to estimate their relative contributions. The usual approach to twin analysis is to specify that the total variance in a trait, v is equal to $a^2 + c^2 + e^2$, where a is additive genetic variance, c is shared (common) environment and e is random, nonshared environment. Similarity between pairs of MZ twins is the sum of genetic and shared environmental influences, $a^2 + c^2$, and similarity between pairs of DZ twins is $0.5\ a^2 + c^2$, so we can estimate heritability (a^2) as twice the difference in correlation between MZ and DZ twins, and then obtain the values for c^2 and e^2 by simple algebra (Sham, 1998). In contemporary twin research, this logic is implemented in a model-fitting approach, which makes it possible to test assumptions of the model and obtain standard errors of path estimates (Eaves *et al.*, 1978).

Twin studies of manual laterality have shown that individual differences in handedness are largely determined by chance, with genetic variation playing a relatively minor role. A meta-analysis of 35 twin studies estimated heritability of around 0.23, with shared environment effect of zero (Medland *et al.*, 2006).

Studies of language laterality are far less numerous, because of the difficulty and expense of studying large numbers of twins. Geschwind *et al.* (2002) wrote a paper entitled 'Heritability of lobar brain volumes in twins supports genetic models of cerebral laterality and handedness', but close inspection of the results reveals that the authors did not provide any

evidence of heritability of structural brain asymmetry. Participants were 72 MZ and 67 DZ male twin pairs aged around 72 years. Conventional twin analysis was conducted and showed high heritability for volumes of the four lobes of the brain on the left and right. 86% of the MZ pairs and 88% of the DZ pairs were concordant for handedness. Further analyses were conducted to compare lobar volumes and asymmetries in those with consistent vs inconsistent handedness, but key data on MZ and DZ concordance for brain asymmetry were not presented. The general finding of high heritability for regional brain volumes is consistent with subsequent twin studies, but to throw light on genetics of cerebral asymmetry, we need data on MZ and DZ twin concordances for a laterality index that indicates the relative size of the two hemispheres. Such data were provided by Eyler *et al.* (2014), who studied 130 MZ and 92 DZ adult twin pairs and concluded: "Our findings suggest that genetic factors do not play a significant role in determining individual variation in the degree of regional cortical size asymmetries measured with MRI, although they may do so for volume of some subcortical structures." (p. 1110).

Jahanshad *et al.* (2010) studied structural brain connectivity in 60 MZ and 45 same-sex DZ right-handed twin pairs using diffusion tensor imaging. They stated: "We expected genetic factors to play a substantial role in the lateralization of the fiber anisotropy in language association regions of the temporal lobe, including the arcuate fasciculus", but in practice heritability estimates were modest at best. They started by looking at twin concordance in a voxel-wise analysis before moving to look at laterality in 12 regions of interest, concluding that genetic factors accounted for 33% of the variance in asymmetry for the inferior fronto-occipital fasciculus (part of the ventral language pathway), 37% for the anterior thalamic radiation, and 20% for the forceps major and uncinate fasciculus. Exclusion of left-handers from the sample may have led to inflated estimates of heritability, because most left-handed twins have a right-handed cotwin, and this discordance could be reflected in discrepant structural asymmetry for members of a twin pair. Neither this study nor that by Eyler *et al.* (2014) found any effect of shared environment on laterality: most variance was explained by the E (non-shared environment) term, reflecting a lack of correlation between both MZ and DZ twins.

Two recent studies by the ENIGMA consortium looked at brain asymmetry for subcortical and cortical structures respectively (Guadalupe *et al.*, 2017; Kong *et al.*, 2018). The first study included analyses of 1170 individuals from 71 genetically informative pedigrees. Subcortical asymmetries, though small in magnitude, were significantly heritable for four of the seven regions, with genetic factors accounting for between .15 and .27 of variation (Guadalupe *et al.*, 2017). Cortical brain asymmetries were more substantial, and in an increased pedigree sample, significant heritability was found for six of 34 measures of cortical thickness and four measures of area, after Bonferroni correction. These analyses were repeated in another genetically informative sample that included 143 MZ pairs and 85 DZ pairs; significant heritability for one of the thickness measures (parahippocampal gyrus) was replicated in this new sample, as was one of the area measures (superior temporal), with heritability estimates ranging from .15 to .23.

It is unclear how these structural asymmetries relate to functional language laterality; there was no evidence of any relationship with handedness in very large meta-analyses done by Guadalupe *et al.* (2017) or Kong *et al.* (2018).

We are aware of only three previous studies that assessed functional language lateralisation in a genetically informative design. Bryden (1975) used two different measures in both parents and two siblings from 49 families. Although he found one statistically significant association between mother and child, he drew attention to inconsistent findings – not only were the correlations between siblings negative, but one of the highest correlations was between mother and father. As he wryly noted, "This correlation would suggest non-random mating for laterality, a characteristic that one would hardly expect to be of significance in selecting one's spouse. "(p. 206). He noted that the measures had only moderate reliability (split half of .61 and .66) but concluded: "one should at least consider seriously the hypothesis that speech lateralization is primarily determined by environmental factors" (p.209).

Ocklenburg *et al.* (2016) used a behavioural measure of language lateralisation, relative ear advantage on dichotic listening, to estimate heritability from parent-offspring relationships in 103 families. Correlations between offspring and mid-parent were close to zero for a laterality index based on free listening, but significant heritability estimates of 0.28 - 0.36 were obtained when participants were instructed to direct attention to one ear. This pattern of results is complicated to interpret, as it could reflect a heritable impact on the ability to direct attention. The authors concluded that the findings "implicate a major contribution of non-genetic influences to individual language lateralization." Somers *et al.* (2015) assessed cerebral lateralisation using functional transcranial Doppler ultrasound in a multi-generational pedigree sample from a single community that had been geographically isolated for generations and so had low genetic heterogeneity. A potential advantage of the study was the use of a pedigree-based method of analysis, which gives higher power than a method reliant just on twin pairs. The selected sample was enriched for left handedness (309 people from 37 families). The heritability of handedness was estimated from pedigree data as 0.24, and the heritability of atypical language lateralisation (coded as a binary variable) was 0.31. The authors noted that heritability may have been overestimated because of over-sampling of families with several left-handed members; selecting only families with at least two left-handers per generation could artificially inflate within-family similarity for laterality. Nevertheless, the heritability of handedness was similar to that obtained from other samples without such ascertainment bias. In both the Ocklenburg *et al.* study and the Somers *et al.* study, heritability was estimated from family relationships, ignoring any effect of shared environmental influences. This seems a reasonable assumption, given that none of the twin studies of laterality reviewed above has found an effect of shared environment.

The importance of phenotype definition
One challenge for researchers studying cerebral lateralisation is how to conceptualise the phenotype. For cerebral lateralisation,

it is possible to obtain a quantitative index reflecting the extent to which activation is more left- than right-sided; in fMRI a laterality index is commonly computed, where 1 is fully left, 0 is equal, and -1 is fully right. For handedness, a similar index may be computed, based on number of right-handed items endorsed on an inventory, relative skill of the two hands, or extent to which preference is maintained across the midline. Depending on how the index is computed, the distribution of scores may be strongly skewed to one side, or even bimodal.

For handedness, the non-normal distribution of preference scores has led to genetic models that propose that handedness is a mixture distribution formed by combining two underlying genotypes: one with a bias to right-handedness, and one with no bias (Annett, 1985; McManus, 1985). However, a failure to find reliable genetic association of handedness with common variants led to an alternative view, which is that atypical lateralisation, of either hand or brain, is caused by any one of a large number of rare genetic variants that add noise to neurodevelopment (Armour et al., 2014). According to this view, there are numerous inherited causal mutations which would be expected to differ from family to family. However, within families, these mutations would be the same for MZ twins and their cotwins, whereas they would be identical on average in 50% of DZ twin pairs. Thus, twin models should be sensitive to such a heritable "neurodevelopmental noise" trait.

The genetic model of laterality that one adopts will affect the optimal analysis. If there are heritable influences on the whole continuum of lateralisation, then the standard method of twin analysis may be the best approach, although an ordinal approach may be needed for non-normal data. If, however, there are distinct genetic influences leading to a skewed laterality distribution, where there is a mixture of two underlying phenotypes, then an 'extremes analysis' may be more appropriate (Bishop, 2005). This approach, pioneered by DeFries & Fulker (1988), involves identifying extreme cases (probands) from a twin sample. Insofar as genetic factors are involved, it is expected that the scores of their cotwins will fall below the population mean, with this effect being stronger for MZ twins, who have all genetic variants in common, compared with DZ twins, who share only 50% of genetic variants on average.

The current study
As far as we are aware, to date there has not been a twin study that uses a direct, functional measure of cerebral lateralisation for language. We report genetic analysis from a study of 141 twin pairs, showing that, consistent with previous studies of handedness and brain asymmetry, chance (or environmental factors not shared between twins) plays the major role in determining individual differences in language lateralisation.

Our sample consists of twin children recruited for a study of the genetic bases of developmental language disorder (DLD), for whom language lateralisation was assessed using functional transcranial Doppler ultrasound (fTCD). Data from these children have previously been reported in the context of an

analysis focusing on relationships between cerebral lateralisation and language functioning (Wilson & Bishop, 2018). That analysis found no difference in language laterality between children with language disorders and those with typical language development, despite the internal consistency of the laterality index obtained with this measure being good (split-half reliability for odd and even trials = 0.84). Furthermore, comparison with a previous study using the same methods confirmed that language lateralisation in twins did not differ from that observed in single-born children. This sample provides a useful opportunity to fill a gap in the literature with an analysis comparing MZ and DZ twins in order to estimate the relative contribution of genetic and non-genetic variation to individual differences in language laterality.

Methods
Participants
For a detailed account of selection of participants, see Wilson & Bishop (2018). In brief, we recruited 194 pairs of twins aged 6 years 0 months to 11 years 11 months, using a sampling approach with the aim of including around 75% twin pairs where one or both had parental report of language or literacy difficulties. Our previous analysis found no association between language laterality and language disorder, and so all children were treated together here. Using a broad definition of language problems, including any mention of history of speech-and-language therapy or communication difficulties, out of 96 MZ twin pairs, 41 (43%) were concordant for language problems, 24 (25%) were discordant for language problems, and the remaining 31 (32%) had neither twin with language problems. Of the 98 DZ twin pairs, 21 (21%) were concordant for language problems, 44 (45%) were discordant for language problems, and the remaining 33 (34%) had neither twin with language problems.

Handedness assessments were completed for all children, and language laterality assessment (see below) was available for 141 pairs. The breakdown of the sample by zygosity and gender is shown in Table 1.

Table 1. N twin pairs by zygosity and sex.

Twin type	All	With fTCD
MZ female	49	30
MZ male	47	35
DZ female	32	25
DZ male	29	21
DZ male/female	37	30
Total	194	141

MZ = monozygotic; DZ = dizygotic; fTCD = functional transcranial Doppler ultrasound.

Zygosity determination

(The following paragraph is copied from Newbury *et al.*, 2018). Oragene kits (OG-500, DNA Genotek Inc, Ontario Canada) were used to collect saliva for DNA analysis from children with SCTs and their parents and available twin pairs. DNA extraction was performed using an ethanol precipitation protocol as detailed in the standard protocol (DNA genotek). All extracted DNA was genotyped on the Infinium 'Global Screening Array-24 (v1)', which includes 692824 SNPs including rare and common variations. Data were processed in the Illumina BeadStudio/GenomeStudio software (v. 2.03) and all SNPs with a GenTrain (quality) score of < 0.5 were excluded at this stage. All genotypes were further filtered using PLINK software v1.07 (Purcell *et al.*, 2007); as recommended by Anderson *et al.* (2010), samples with a genotype success rate below 95% or a heterozygosity rate ±2 SD from the mean were removed, as were SNPs with a Hardy-Weinberg equilibrium $P < 0.000001$ or a minor allele frequency of less than 1%. Identity data within families and twin-pairs were used to exclude samples with unexpected gender or relationships. SNPs that showed an inheritance error rate > 1% or skewed missing rates between genotype plates were also excluded.

DNA was available for 191 twin pairs who were compared across 250,875 SNPs. All gave unambiguous zygosity signals on Identity by State (IBS), i.e., the proportion of SNPs for which any given twin pair share genotypes: this was either close to 1.0 (MZ) or close to 0.5 (DZ). For twins with missing or inadequate DNA samples we relied on parental report of zygosity.

Laterality assessment

1. Handedness

Hand preference was assessed using a hand preference battery based on items from the Edinburgh Handedness Inventory (EHI) (Oldfield, 1971), modified to exclude one item deemed unsuitable for children (striking a match). With adults, the EHI is administered as a questionnaire, but in our study children were asked to demonstrate each of ten actions: writing, drawing, throwing a ball, using scissors, using a toothbrush, cutting with a knife, using a spoon, using a broom (upper hand), taking the lid off a box, and dealing cards. One point was awarded for exclusive right hand use, zero points for left hand use, and half a point if both hands were used, giving a score ranging from zero to ten.

Strength of hand preference was assessed using the Quantification of Hand Preference (QHP) task (Bishop *et al.*, 1996), which measures the tendency to continue to use the preferred hand when cards are picked up from different spatial locations. Three cards are set out in each of seven positions extending at 30 degree intervals from the left to the right of the child's midline. The child is not told that handedness is being assessed, and treats the task as a picture-name matching game, where they have to pick up the named card and put it in a centrally-placed box. The same quasi-random order of positions was used for all children, starting with a card at the midline and continuing until the child had picked up and placed three cards at each of seven locations, to give a total of 21 trials. For each card, two points were awarded for right-handed use, zero points for left-handed use, and one point if the card was transferred from one hand to another in the course of placing it in the box. Test-retest reliability of the QHP in adults has been shown to be good when there are five items in each position (Doyen & Carlier, 2002), but it should be noted that a more recent study with 6- to 7-year-old children using 3 items per position found test-retest reliability of only 0.35 (Pritchard *et al.*, 2019); results from this test should therefore be interpreted cautiously.

2. Language laterality

Language laterality was assessed using functional transcranial Doppler ultrasound (fTCD) recorded while the child described short episodes from a story presented as an animation. The equipment consisted of Doppler-Box™ X digital (Smart Medical) with QL software. A DiaMon® monitoring headset was used with two 2 MHz hand-held probes (2.9m length). A video demonstration of this procedure using an earlier version of the equipment and headset is available from Bishop *et al.* (2010). Transcranial Doppler ultrasound is used in medical contexts to assess the integrity of the cerebral blood vessels. For assessing cerebral lateralisation, left and right ultrasound probes are attached to a headset and positioned so as to detect lateralised changes in blood flow in the middle cerebral arteries.

On each trial, the child silently views a 12 s clip from a cartoon that included sounds but no speech. A response cue appears when the video clip finishes to indicate the start of a 10 s period during which the child is asked to describe what happened in the cartoon. A second cue then indicates that the child should stop talking and relax. This paradigm has previously been found to have good validity and internal consistency (Bishop *et al.* 2009). In adults, we recently demonstrated test-retest reliability of 0.84 for a Sentence Generation task that was similar to the task used here, but with static pictures rather than video sequences as stimuli (Woodhead *et al.*, 2019).

A maximum of 30 trials was administered, depending on the child's tolerance of the procedure. The child's verbal responses were recorded and subsequently transcribed, and the examiner noted behaviour during the procedure. Trials were excluded if the child either spoke during a silent period, or failed to talk during the 'talk' period: these need to be omitted because they invalidate the trial, which involves comparing the period when the child talks with a baseline period when no talking occurs.

The analysis of the animation task data consists of a standard sequence of processing steps, following original work by Deppe *et al.* (1997). We used a custom script written in R (R core team, 2019) for data processing. This included an initial step of identifying trials where there was very brief signal dropout (affecting one datapoint) and interpolating the mean value in such cases. Trials with more prolonged signal dropout were discarded. After these preliminary steps, heart cycle integration was applied to remove the heartbeat, followed by signal normalisation, artefact rejection, epoching and baseline correction. The averaged left and right velocity plots were subtracted to give a difference waveform.

Following Wilson & Bishop (2018) we excluded data from 11 twin pairs where one of the twins had fewer than 12 accepted trials, as the LI is likely to be unreliable when based on such a small amount of data. In addition, data were excluded for one twin pair where one child's laterality index was more than 5 SD from the mean.

In our previous report of data from this sample (Wilson & Bishop, 2018), we used the conventional approach for obtaining a laterality index based on the peak difference (maximum or minimum) in a period of interest, predefined as 4 to 14 seconds after the cue to speak. This method involves finding the largest maximum or minimum in the difference wave, and measuring the size of the difference for a 2 s period around that peak. Our subsequent studies with adults suggested that this method is not ideal, because there are cases where the difference wave shifts from positive to negative, or vice versa, within the period of interest, and quite minor differences in size of positive and negative peaks can determine whether laterality is coded as left or right (Woodhead et al., 2018). Accordingly, in more recent studies, we have calculated a laterality index (LI) as the mean amplitude of blood flow velocity difference in the whole period of interest (4 to 14 s). In the current dataset, the correlation between this mean-based LI and the traditional peak-based LI is very strong: Pearson correlation = 0.911, DF = 280, but the distributions differ. The split-half reliability, based on correlation of LIs from odd and even trials is closely similar to that from the original method, r = 0.85. The original method, however, gives a non-normal distribution of laterality indices, with a point of rarity around zero, which appears to be a spurious artefact of the method of computation. As requested by reviewers, results obtained with the original peak method are included in our analysis for completeness, and the two methods are compared in a scatterplot in Appendix 1 (https://osf.io/tfyk3/).

The standard error of the LI for each individual was computed from the LI obtained across individual trials. This allowed us to consider whether the child's LI was significantly different from zero, i.e. whether the 95% confidence interval spanned zero. Where this was the case, laterality was categorised as left or right, and where the LI was not significantly different from zero, the laterality was coded as bilateral. Note that coding of bilateral laterality can result if data are merely noisy. For comparison with Somers et al. (2015), we also categorised individuals on the basis of the peak-based laterality index into 'typical' and 'atypical' laterality, with the latter group including all those who were not significantly left-lateralised (i.e., the confidence interval of their laterality index spanned, or fell below, zero).

Following a suggestion by Francks (2019), we also conducted genetic analysis of the mean left- and right-sided blood flow measures.

Procedure

Sections of this paragraph are copied from Wilson & Bishop, 2018. Ethical approval was obtained for the study in 2011 from the Berkshire NHS Research Ethics Committee (reference 11/SC/0096), and data collection started in August of that year, finishing in October 2016. Information sheets, consent forms and ethics approval documents are available on Open Science Framework. Where families had expressed interest in the study, they were interviewed by telephone to assess whether the children were likely to meet inclusion criteria, and an appointment was made to see the twins at home or at school, depending on parental preference. Families were widely dispersed around the UK, including Northern Ireland, Scotland, Wales and England, so testing was scheduled where possible to minimise travel. During the course of recruitment, which lasted for a period of five years, a total of eight research assistants as well as the senior author were involved in assessing children. In some cases, two testers worked together, each seeing one twin, and in others a single tester saw both children sequentially. The assessment was conducted in a single session lasting between 2–3 hours per child, with breaks where needed.

Data analysis

1. AE modeling

As noted above, the usual approach to twin analysis involves decomposing variance of a phenotype into components attributed to additive genetic (a), common (shared) environment (c) and nonshared environment (e). This decomposition is typically implemented using structural equation modeling with maximum likelihood estimates (Rijsdijk & Sham, 2002), which make it possible to test whether the data meet underlying assumptions, to compare fit of different models, and to obtain standard errors of estimates. Large samples are needed to accurately estimate additive genetic (a^2), shared environmental (c^2) influences, both of which lead to positive covariance between two members of both MZ and DZ pairs: they are distinguished by the fact that genetic influence leads to greater covariance for MZ than for DZ twins. In the context of laterality, however, prior studies have found shared environmental influences to be negligible in adequately powered twin studies, and it is safe to ignore the c^2 term (Medland et al., 2006; Medland et al., 2009). This simplifies the analysis, making it possible to detect genetic effects with smaller samples, as any positive correlation between twins and their cotwins can be interpreted as a genetic effect. An AE model was fitted to the raw data using two R packages (version 3.6, R Core Team, 2019): OpenMx package (verson 2.13.2) (Neale et al., 2016) with the umx package (version 3.0.0) used for the non-normal handedness data (Bates et al., 2019). Both univariate and multivariate models were evaluated. Scripts used to pre-process Doppler files are provided as extended data (Bishop, 2019).

2. Power analysis

A power analysis was conducted to estimate the power to detect heritability of 0.15 or more, using the power.ACE.test function in umx (see Figure 1). This showed that for the larger sample of twins for whom handedness data were available, there is 80% power to detect heritability of 0.25 in an AE model with alpha of 0.05, and for the smaller sample with language laterality data, power of 67%. In this smaller subset with language laterality data, 80% power is obtained for heritability of
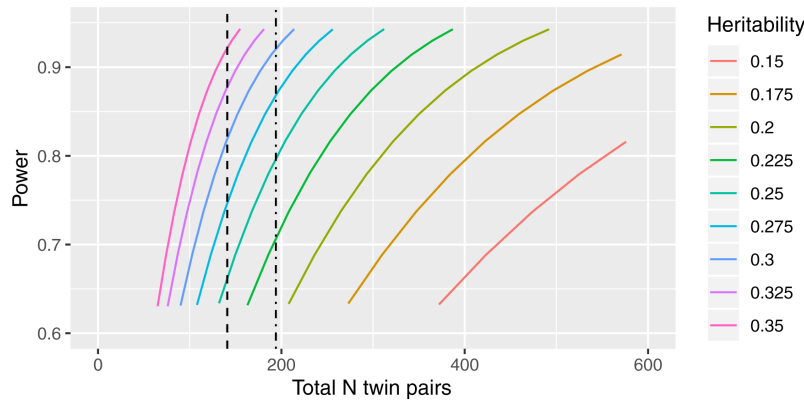
Figure 1. Power in relation to sample size for different levels of heritability in an AE model. The two vertical lines correspond to the sample size for handedness measures (right-most line) and the laterality index (left-most line).

0.3, and over 95% power for heritability of 0.35. Thus, even the low heritability handedness estimates reviewed in the Introduction should be detectable with this sample. For language laterality, there is around a 2 in 3 chance of detecting a true but small effect of around 0.25, and strong power to detect the higher heritability reported in the DTI study by Jahanshad *et al*. (2010). However, neither sample is adequately powered to detect heritability levels below .2

Results

Figure 2 shows the distributions of scores obtained on the two handedness measures and the language laterality index from fTCD, and Figure 3 shows scatterplots depicting the association of the laterality indices between two members of a twin pair (see underlying data (Bishop, 2019)).

The density plots reveal that data from the handedness measures are highly non-normal, following the usual J-shaped distribution for handedness measures, with the majority of cases bunched up at the right handed end of the scale.

Figure 3 shows that the correlations between two members of a twin pair are low for all measures.

Contrary to expectation, no association was found between handedness and the language laterality index: with cases divided into those with laterality indices above and below zero, 76% of the left-handers and 78% of the right-handers were left-lateralised for language.

The OpenMx package was used to run an AE model with the data from the two handedness tasks and the language laterality task. It was anticipated that the model would not fit with the data because (a) the handedness data were highly non-normal, and (b) the correlations for the language laterality index were close to zero. However, a good fit was obtained for all three measures. Heritability estimates for the two handedness measures were compatible with those obtained in previous studies, with values of a^2 of 0.25 for the Edinburgh Handedness Inventory and 0.18 for the Quantification of Hand Preference

task (see Table 2). The fit of a model including a genetic term was substantially better than for one excluding it for both tasks (p-values less than 0.01 for both measures). For the language laterality index, the estimated value of a^2 was zero, and a model with no genetic term gave as good a fit as one including it. The same was true for both the laterality index based on the traditional peak method, and for the binary category of typical/atypical laterality that was comparable to that used by Somers *et al*. (2015).

Because of concerns that the non-normality of the handedness data could distort heritability estimates, analyses of the EHI and QHP were repeated using the umx package (Bates, 2018) to run an ordinal version of ACE analysis. This gave slightly different estimates of heritability than the standard analysis using the AE model, with values of a^2 of 0.24 and 0.22 for EHI and QHP respectively, and c^2 close to zero.

We also ran a multivariate model that included laterality indices from EHI, QHP and fTCD, to see how far heritable influences are shared between measures (Table 3). This revealed that some of the genetic influence on QHP was shared with EHP, and that there was no shared genetic influence with the language laterality LI from fTCD.

As suggested by Francks (2019), we also did a twin analysis of the left and right blood flow volumes. The pattern of twin-cotwin correlations for these measures was quite different from that seen for the difference scores, with modest but significant correlations between twins and cotwins, but no evidence of a zygosity-specific effect. For left-sided blood flow, the twin-cotwin correlation (95% CI) was 0.40 (0.24 to 0.55) for MZ twins and 0.33 (0.18 to 0.46) for DZ twins; for right-sided blood flow, the correlation was 0.24 (0.07 to 0.39) for MZ twins and 0.32 (0.16 to 0.45) for DZ twins. This suggests that shared environment, rather than genetic similarity, drives twin-cotwin similarity. Repeating the analysis with age and sex residualised did not affect results. A CE model was fitted to each measure, and gave as good a fit as an ACE model, with significant (p < .001) c^2 estimates of 0.34 for left-sided
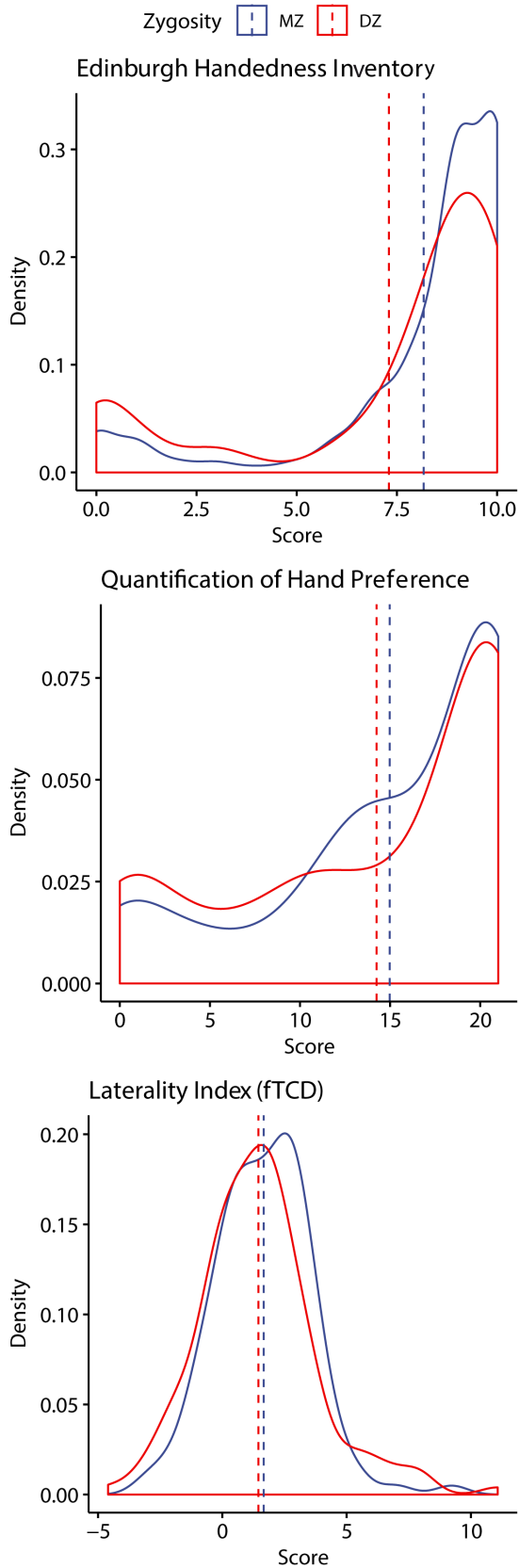
Figure 2. Vertical dotted line shows mean for MZ (blue) and DZ (red) twins. Density plots for handedness and language laterality indices.
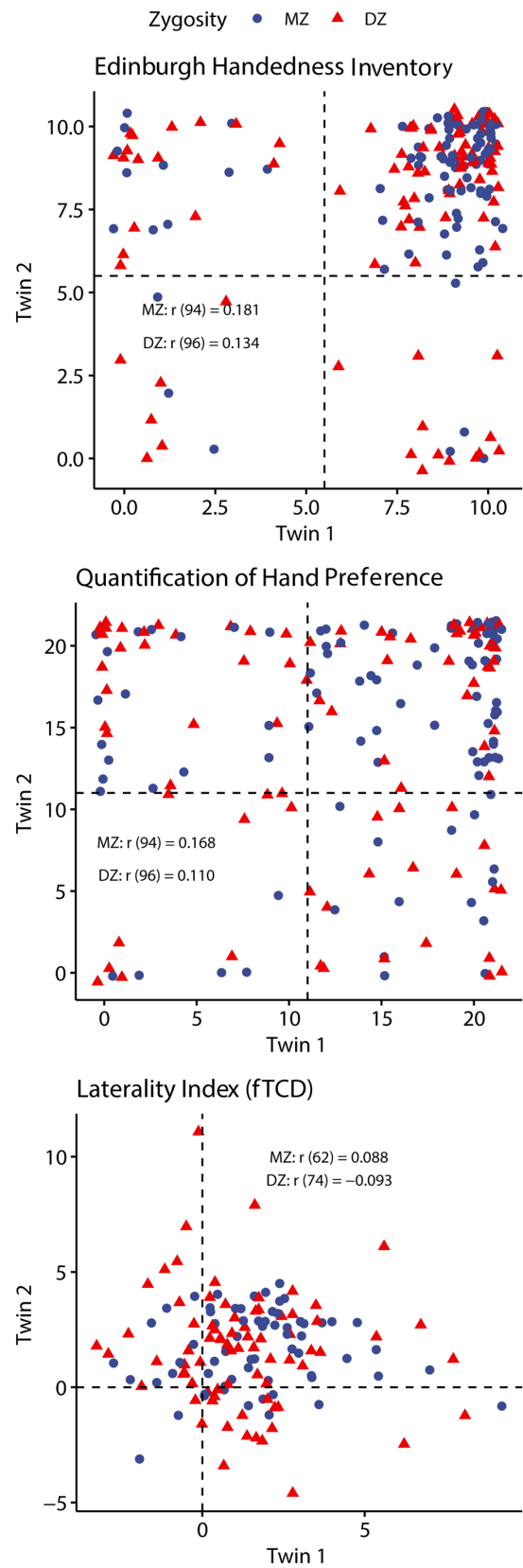
Figure 3. Scatterplots by zygosity for handedness and language laterality indices. The data from handedness measures are jittered; Spearman correlation coefficients are shown.

Table 2. Heritability estimates from AE model for different measures. 95% CI in brackets. MZ = monozygotic; DZ = dizygotic.

Measure	MZ*	DZ*	rMZ	rDZ	a²	chisq	p**
Edinburgh Handedness Inventory	96	98	0.18 (0.04 to 0.32)	0.13 (-0.01 to 0.28)	0.25 (0.03 to 0.34)	11.7	< .001
Quantification of Hand Preference	96	98	0.17 (0.01 to 0.3)	0.11 (-0.05 to 0.25)	0.18 (0 to 0.31)	6.8	.009
Laterality index (fTCD)	65	76	0.09 (-0.08 to 0.26)	-0.09 (-0.25 to 0.06)	0 (0 to 0.15)	0	1
Laterality index (peak FTCD)	65	76	0.11 (-0.07 to 0.27)	-0.07 (-0.23 to 0.09)	0.02 (0 to 0.20)	0	.823
Binary typical (1)/atypical (0)	65	76	0.02 (-0.16 to 0.19)	0.02 (-0.13 to 0.19)	0 (0 to 0.25)	0	1

* N pairs
** Significance of change of fit if a² term dropped

Table 3. Multivariate AE model estimates, unsquared standardized path estimates with SEs.

		Shared by all 3 measures		Independent of EHP		Specific to LI	
		Estimate	SE	Estimate	SE	Estimate	SE
A	EHP	0.458*	0.098				
A	QHP	0.243	0.124	0.343*	0.093		
A	LI	0.123	0.113	-0.152	0.123	0.000	0.213
E	EHP	0.889*	0.051				
E	QHP	0.401*	0.068	0.814*	0.043		
E	LI	-0.134	0.066	0.039	0.064	0.971*	0.039

Note for table 3: Asterisks denote paths that are statistically significant at .05 level.

flow and 0.27 for right-sided flow. The left- and right-sided flow measures were closely linked, and intercorrelated with r = 0.87.

The basic logic of the DeFries & Fulker (1988) method for analysing heritability of extreme scores is that if we select probands with extreme scores, then the scores of cotwins should regress more to the population mean for DZ twins than for MZ twins. The plausibility of such a model can be readily tested by selecting twins with an extreme score and then using a t-test to compare co-twin scores for MZ and DZ twins. Results of this analysis are shown for all three phenotypes in Table 4, which shows no reliable difference between cotwins for MZ and DZ probands. These data must be interpreted with extreme caution because of the small sample sizes, but they do not lend any support to the idea that atypical laterality is caused by a qualitatively different genetic process than normal range variation, either for handedness or for language laterality.

Discussion

The twin analysis of handedness data from this study gave results that were consistent with those from previous meta-analyses, with around 20% of variance accounted for by genetic factors. The language laterality index, however, gave results that were unexpected in two respects. First twin-twin correlations were close to zero for both MZ and DZ twins, and appeared therefore to be determined entirely by chance; second, there was no difference in rates of left-sided laterality between left-handers (76%) and right-handers (78%), with the latter figure being lower than usually found using other indicators of cerebral lateralisation (Carey & Johnstone, 2014).

This raises the question as to the validity of the laterality index obtained using functional transcranial Doppler ultrasound. If chance is the principal determinant of the LI, is this just because the measure is unreliable within individuals? We do not have test-retest data on children, but it seems unlikely poor

Table 4. Extreme probands, t-tests comparing MZ/DZ co-twins.
MZ = monozygotic; DZ = dizygotic; QHP = Quantification of Hand Preference,
LI = laterality index.

	Edinburgh Handedness	QHP	Language LI
N MZ	21	40	22
N DZ	40	58	39
MZ proband mean (SD)	1.43 (1.66)	3.61 (3.86)	-1.06 (2.89)
DZ proband mean (SD)	1.15 (1.42)	3.89 (4.22)	-1.24 (3.29)
MZ co-twin mean (SD)	6.86 (3.48)	13.46 (7.14)	1.34 (2.49)
DZ co-twin mean (SD)	6.58 (3.63)	14.22 (7.19)	2.14 (2.57)
t	0.30	-0.52	-1.19
df	42.3	84.4	44.9
p	0.768	0.606	0.239

reliability is the whole explanation, for three reasons. First, the split half reliability of the LI in this sample is around 0.85, which indicates reasonable consistency from trial to trial across the testing session. With adults, we have explicitly considered test-retest reliability of the LI obtained with fTCD, and found that, while it varies from task to task, it is generally good, with test-retest correlation of 0.84 for the task that is most similar to the animation description task (Woodhead et al., 2019). Second, as shown in Figure 3, the children studied here showed a robust bias to the left hemisphere at the group level. Third, the current result is broadly consistent with the handful of studies that have looked at structural or functional brain lateralisation. Although these have revealed some heritable laterality indices, these are typically small in magnitude. It is also compatible with a recent study sequencing the genomes of 33 subjects with right-hemisphere language dominance and 34 typical left-dominant subjects and finding no associated mutations distinguishing the individuals with atypical language laterality (Carrion-Castillo et al., 2019).

When one considers that the levels of blood flow to left and right hemisphere show no indication of genetic influence, it is not surprising that the laterality index, based on the difference between these measures, is not heritable. The evidence for a shared environment effect on the blood flow measures is unexpected; one possibility is that the intrauterine environment could be implicated in influencing development of cerebral vasculature.

A further question is why language laterality shows zero heritability, whereas handedness shows small but significant heritability. This difference in findings may prove to be an uninteresting artefact of the smaller sample size for the language laterality measure than for handedness, combined with perhaps lower reliability of the measure, leading to reduced power to detect a true effect. As the difference in heritability estimates between measures was not large, we cannot dismiss the possibility that the true level of heritability

for language laterality is similar to that for handedness - slight but not totally absent.

One does need to be cautious about assuming lack of genetic effect on the basis of a small sample. In the past, the first author concluded that handedness was not heritable, on the basis of small-scale twin studies that found no evidence of genetic influence, but subsequent meta-analyses have shown consistent but low heritability. It has nevertheless been remarkably difficult to find any genetic variants consistently linked with variations in handedness, and many promising findings appear to have been false positives (de Kovel & Francks, 2019). It subsequently became clear, however, that there is a genetic effect, but it is small and only clearly detectable in large samples. In a sample of over one million people, Cuellar-Partida et al. (2020) identified 41 common genetic variants that were associated with left-handedness, and 7 associated with ambidexterity, each with a very small effect. While very low levels of heritability pose methodological problems, the search for genetic variants continues, not so much with the goal of explaining large amounts of variance in handedness, but rather with the goal of illuminating the biological pathways involved in determining asymmetry. This can be feasible, provided there are suitable phenotypic measures available in very large samples (de Kovel & Francks, 2019; Wiberg et al., 2019)

It is possible that the same will prove to be the case for language lateralisation, especially given prior findings of significant heritability in structural measures of subcortical brain regions (Eyler et al., 2014; Guadalupe et al., 2017), cortical area and thickness (Kong et al., 2018) and language-related fibre tracts (Jahanshad et al., 2010), plus the large family study of Somers et al. (2015) that used a binary phenotype, and the mixed findings on dichotic listening by Ocklenburg et al. (2016). A further point to note is that different language tasks show different degrees of lateralisation (Woodhead et al., 2019), and we simply do not know which may be the most heritable. As Ocklenburg et al. (2016) noted: "there seems

to be a considerable phenotype-dependent variability regarding the heritability of language lateralization." (p. 37). A research priority should be the development of optimal methods for deriving a valid and reliable language laterality index from brain measures, as without these, progress will be limited.

The current small study is not sufficient to prove zero heritability for lateralised brain function. On the other hand, it is striking how difficult it has been to replicate previous studies of genetic associations with laterality, and the flexibility with which the phenotype can be defined does increase the likelihood that some findings may be type I errors (Bishop, 1990).

Our data are compatible with a more radical model in which language laterality is the consequence of a general population left-sided brain bias for language which is genetic but not heritable, i.e., does not show any individual variation. If a genetic biasing factor applies to the whole population, without there being any variation, then heritability will be zero. The postulated population bias mechanism would have to be at least somewhat probabilistic, with some individuals showing atypical lateralisation just by chance. Such a model is consistent with the view of neurodevelopment proposed by Mitchell (2018). He noted that it is customary to interpret the 'e' term of an ACE model as reflecting some systematic environmental influence that is not shared by the two members of a twin pair, literally 'non-shared environment'. He argues that this neglects the likely role of stochastic influences on neurodevelopment (and in many traits), and notes that evidence for such 'developmental noise' comes from the numerous instances where there is phenotypic variability despite genetic identity. This is seen not only in MZ twins with the same genetic sequence but different phenotypic outcomes, but also in the two sides of the face, which are seldom totally symmetric, despite having the same DNA.

It would be rash to draw strong conclusions from a single study with a null result, especially when the expected level of heritability is low. It is possible that with larger samples and different measures we will be able to confirm that language lateralisation is a heritable phenotype. These results, however, encourage us to at least consider the provocative possibility that language lateralisation may be a phenotype that breaks Turkheimer's (2000) first law of behaviour genetics: 'All human behavioural traits are heritable'.

Data availability

Underlying data
Open Science Framework: Double entry data. https://doi.org/10.17605/OSF.IO/5h82q (Bishop, 2019)

This project contains the following underlying data:
- doubleentry_data_dictionary.xlsx (Excel spreadsheet with data dictionary for TwinLatOSF.csv)

- Twins_Doppler_processed_NewLI.xlsx (CSV file containing handedness and language laterality data)

- heritability lat writeup_forpaper.rmd (R markdown script to create this paper with figures and analyses).

- Appendix 1: Scatterplot showing relationship between two methods of deriving laterality index (peak vs mean measures).

Extended data
Open Science Framework: Preprocessing for mean LIs - twins. https://doi.org/10.17605/OSF.IO/CPKHB (Bishop, 2019)

This registered project contains the following extended data:
- R_doppler_v2_NEW_LI_2019_DB.R (R script for preprocessing of Doppler files)

- Individual trial LIs.csv (CSV file containing individual language laterality file data)

- Doppler_raw.zip (Zip file containing individual raw Doppler readings)

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

References

Anderson CA, Pettersson FH, Clarke GM, et al.: **Data quality control in genetic case-control association studies.** *Nat Protoc.* 2010; **5**(9): 1564–1573.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Annett M: **Left, right, hand and brain: the Right Shift Theory**. Hillsdale, NJ: Erlbaum. 1985.
Reference Source

Armour JA, Davison A, McManus IC: **Genome-wide association study of**

handedness excludes simple genetic models. *Heredity (Edinb).* 2014; **112**(3): 221–225.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Bates TC: **umx: A helper package for Structural Equation Modeling in OpenMx**. 2018.
Reference Source

Bates TC, Maes H, Neale MC: **umx: Twin and Path-Based Structural Equation**

Modeling in. *Twin Res Hum Genet.* 2019; **22**(1): 27–41.
PubMed Abstract | **Publisher Full Text**

Bishop DV: **DeFries-Fulker analysis of twin data with skewed distributions: cautions and recommendations from a study of children's use of verb inflections.** *Behav Genet.* 2005; **35**(4): 479–90.
PubMed Abstract | **Publisher Full Text**

Bishop DV: **How to increase your chances of obtaining a significant association between handedness and disorder.** *J Clin Exp Neuropsychol.* 1990; **12**(5): 812–6.
PubMed Abstract | **Publisher Full Text**

Bishop DVM: **"Preprocessing for Mean Lis - Twins."**. *OSF.* 2019.
http://www.doi.org/10.17605/OSF.IO/CPKHB

Bishop DV, Badcock NA, Holt G: **Assessment of cerebral lateralization in children using functional transcranial Doppler ultrasound (fTCD).** *J Vis Exp.* 2010; (43): 2161.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Bishop DV, Ross VA, Daniels MS, *et al.*: **The measurement of hand preference: a validation study comparing three groups of right-handers.** *Br J Psychol.* 1996; **87**(pt 2): 269–285.
PubMed Abstract | **Publisher Full Text**

Bishop DV, Watt H, Papadatou-Pastou M: **An efficient and reliable method for measuring cerebral lateralization during speech with functional transcranial Doppler ultrasound.** *Neuropsychologia.* 2009; **47**(2): 587–590.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Bryden MP: **Speech lateralization in families: a preliminary study using dichotic listening.** *Brain Lang.* 1975; **2**(2): 201–211.
PubMed Abstract | **Publisher Full Text**

Carey DP, Johnstone LT: **Quantifying cerebral asymmetries for language in dextrals and adextrals with random-effects meta analysis.** *Front Psychol.* 2014; **5**: 1128.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Carrion-Castillo A, Van der Haegen L, Tzourio-Mazoyer N, *et al.*: **Genome sequencing for rightward hemispheric language dominance.** *Genes Brain Behav.* 2019; **18**(5): e12572.
PubMed Abstract | **Publisher Full Text**

Cuellar-Partida GC, Tung JY, Eriksson N, *et al.*: **Genome-wide association study identifies 48 common genetic variants associated with handedness.** *Nat Human Behav.* 2020; in press

DeFries JC, Fulker DW: **Multiple regression analysis of twin data: etiology of deviant scores versus individual differences.** *Acta Genet Med Gemellol (Roma).* 1988; **37**(3–4): 205–216.
PubMed Abstract | **Publisher Full Text**

Deppe M, Knecht S, Henningsen H, *et al.*: **AVERAGE: a Windows program for automated analysis of event related cerebral blood flow.** *J Neurosci Methods* 1997; **75**(2): 147–154.
PubMed Abstract | **Publisher Full Text**

de Kovel CGF, Francks C: **The molecular genetics of hand preference revisited.** *Sci Rep.* 2019; **9**(1): 5986.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Doyen A, Carlier MF: **Measuring handedness: a validation study of Bishop's reaching card test.** *Laterality.* 2002; **7**(2): 115–130.
PubMed Abstract | **Publisher Full Text**

Eaves LJ, Last KA, Young PA, *et al.*: **Model-fitting approaches to the analysis of human behaviour.** *Heredity (Edinb).* 1978; **41**(3): 249–320.
PubMed Abstract | **Publisher Full Text**

Eyler LT, Vuoksimaa E, Panizzon MS, *et al.*: **Conceptual and data-based investigation of genetic influences and brain asymmetry: a twin study of multiple structural phenotypes.** *J Cogn Neurosci.* 2014; **26**(5): 1100–1117.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Francks C: **Peer Review Report For: Negligible heritability of language laterality assessed by functional transcranial Doppler ultrasound: a twin study [version 1; peer review: 3 approved with reservations].** *Wellcome Open Res.* 2019; **4**: 161.
Publisher Full Text

Geschwind DH, Miller BL, DeCarli C, *et al.*: **Heritability of lobar brain volumes in twins supports genetic models of cerebral laterality and handedness.** *Proc Natl Acad Sci U S A.* 2002; **99**(5): 3176–3181.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Guadalupe T, Mathias SR, vanErp TGM, *et al.*: **Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex.** *Brain Imaging Behav.* 2017; **11**(5): 1497–1514.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jahanshad N, Lee AD, Barysheva M, *et al.*: **Genetic influences on brain asymmetry: a DTI study of 374 twins and siblings.** *NeuroImage.* 2010; **52**(2):

455–469.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kong XZ, Mathias SR, Guadalupe T, *et al.*: **Mapping cortical brain asymmetry in 17,141 healthy individuals worldwide via the ENIGMA Consortium.** *Proc Natl Acad Sci U S A.* 2018; **115**(22): E5154–E5163.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

McManus IC: **Handedness, language dominance and aphasia: a genetic model.** *Psychol Med Monogr Suppl.* 1985; **8**: 1–40.
PubMed Abstract | **Publisher Full Text**

Medland SE, Duffy DL, Wright MJ, *et al.*: **Genetic influences on handedness: data from 25,732 Australian and Dutch twin families.** *Neuropsychologia.* 2009; **47**(2): 330–337.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Medland SE, Duffy DL, Wright MJ, *et al.*: **Handedness in twins: joint analysis of data from 35 samples.** *Twin Res Hum Genet.* 2006; **9**(1): 46–53.
PubMed Abstract | **Publisher Full Text**

Mitchell KJ: **Innate: How the wiring of our brains shapes who we are**. Princeton and Oxford: Princeton University Press. 2018.
Publisher Full Text

Neale MC, Hunter MD, Pritikin JN, *et al.*: **OpenMx 2.0: Extended Structural Equation and Statistical Modeling.** *Psychometrika.* 2016; **81**(2): 535–549.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Newbury DF, Simpson NH, Thompson PA, *et al.*: **Stage 2 Registered Report: Variation in neurodevelopmental outcomes in children with sex chromosome trisomies: testing the double hit hypothesis [version 2; peer review: 2 approved].** *Wellcome Open Res.* 2018; **3**: 85.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ocklenburg S, Ströckens F, Bless JJ, *et al.*: **Investigating heritability of laterality and cognitive control in speech perception.** *Brain Cogn.* 2016; **109**: 34–9.
PubMed Abstract | **Publisher Full Text**

Oldfield RC: **The assessment and analysis of handedness: the Edinburgh inventory.** *Br J Psychol.* 1971; **66**: 53–9.

Pritchard VE, Malone SA, Burgoyne K, *et al.*: **Stage 1 Registered Report: The relationship between handedness and language ability in children [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2019; **4**: 30.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Purcell S, Neale B, Todd-Brown K, *et al.*: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet.* 2007; **81**(3): 559–575.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

R Core Team: **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2019.
Reference Source

Rijsdijk FV, Sham PC: **Analytic approaches to twin data using structural equation models.** *Brief Bioinform.* 2002; **3**(2): 119–33.
PubMed Abstract | **Publisher Full Text**

Sham P: **Statistics in Human Genetics**. London: Arnold. 1998.

Somers M, Ophoff RA, Aukes MF, *et al.*: **Linkage analysis in a Dutch population isolate shows no major gene for left-handedness or atypical language lateralization.** *J Neurosci.* 2015; **35**(23): 8730–8736.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Turkheimer E: **Three laws of behavior genetics and what they mean.** *Curr Dir Psychol Sci.* 2000; **9**(5): 160–164.
Publisher Full Text

Vingerhoets G: **Phenotypes in hemispheric functional segregation? Perspectives and challenges.** *Phys Life Rev.* 2019; **30**: 1–18.
PubMed Abstract | **Publisher Full Text**

Wiberg A, Ng M, Al Omran Y, *et al.*: **Handedness, language areas and neuropsychiatric diseases: insights from brain imaging and genetics.** *Brain.* 2019; **142**(10): 2938–2947.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Wilson AC, Bishop DVM: **Resounding failure to replicate links between developmental language disorder and cerebral lateralisation.** *PeerJ.* 2018; **6**: e4217.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Woodhead ZVJ, Bradshaw AR, Wilson AC, *et al.*: **Testing the unitary theory of language lateralization using functional transcranial Doppler sonography in adults.** *R Soc Open Sci.* 2019; **6**(3): 181801.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Woodhead ZVJ, Rutherford H, Bishop DVM: **Measurement of language laterality using functional transcranial Doppler ultrasound: a comparison of different tasks [version 2; peer review: 2 approved].** *Wellcome Open Res.* 2018; **3**: 104.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 3

Reviewer Report 18 September 2020

https://doi.org/10.21956/wellcomeopenres.17905.r40488

✔ **Chris McManus** (iD)
  [1] Research Department for Medical Education, University College London, London, UK
  [2] Research Department of Clinical, Educational and Health Psychology, University College London, London, UK

Thanks to the authors for those minor but useful revisions. I am now very happy to recommend indexing of this interesting, unusual and important study.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuropsychology of cerebral lateralisation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 11 May 2020

https://doi.org/10.21956/wellcomeopenres.17276.r38147

? **Chris McManus** (iD)
  [1] Research Department for Medical Education, University College London, London, UK
  [2] Research Department of Clinical, Educational and Health Psychology, University College London,

London, UK

Firstly my thanks to the authors for their helpful re-analyses and thoughtful comments on the revision. It was good to see that estimates I made of the upper confidence interval for the heritability were broadly correct, even if done by a different method. My apologies also for the delay in getting back – the revision arrived on my desk just at the same time as a vast amount of work arising in medical schools from the Covid-19 outbreak, and I have only just found a sufficient gap to be able to return to this very interesting and undoubtedly important study.
Inevitably any revision, particularly with new data analyses, results in further questions being raised, and it is inevitable that some of those will require some sort of reply.

The replies to the current revision by the author can broadly be divided into two groups. Those to do with the details of the data and the analyses, and those to do more generally with philosophy of science, and issues arising from the replication crisis. I will deal with the latter first.

Philosophical issues. As the authors say, many of the criticisms, "get to the nub of issues that have become prominent in discussions of the so-called 'reproducibility crisis'". Type II errors have always been problematic, particularly with low sample sizes (and this it has to be said is a very low sample size compared with many evaluating heritability).

Acceptable studies in the previous literature. Discursive reviews had many problems, one of which was to review lots of studies, reject most as being inadequate in some form or another, and then conclude there were but a few decent bits of evidence, all of which happened to support one's own theory. Meta-analysis in large part got rid of that, particularly if study quality measures were included, and the file drawer methods, funnel plots, etc, helped to deal with the problems of the literature being biased. Now however a new issue arises, as "in the absence of pre-registered studies, we also need to adopt a cautious stance to the prior literature". There are of course no other pre-registered studies in the literature except this one, which does a good job of cleaning the Augean stables, but sadly leaves almost no babies in their bath water (to mix metaphors). It does make life easier, of course, but medicine went through this a decade or two ago when some decided that the only treatments that should be used were those supported by randomised controlled trials. That fell apart when it was realised that there were no RCTs for penicillin, that there were no RCTs for treating common conditions such as ear wax, and that a decent RCT carried out for patients aged 50-69 could not be used if one happened to be aged 49 or 70... Easy but strict rules make stupid decisions, and in general judgement is required. At this point I am tempted to say that there are also no pre-registered studies testing out what Newton or Darwin talked about... You get the idea.

P-hacking is another issue, but the solution is not to reject any choice of outcome measure, but instead perhaps to try all possible ways of dividing up outcome measures – variants of this in data science go under the heading "targeted learning" – and very complicated they can be!

Enough said, but I personally am not going to throw away the entire previous literature as yet, and in particular I will tend to stick with what Broca and his successors found, rather than saying that they can safely be ignored. I will also continue to believe that it is at least plausible that the underlying genetic basis for lateralisations is held in common.

Use of language as a philosophical issue. This is also a philosophical/methodological issue

concerning the extent to which language biases the interpretation of otherwise relatively objective numbers. Thus to say in the discussion that, "the possibility that the true level of heritability for language laterality is similar to that for handedness – slight but not totally absent." "Slight" here is mainly being used to dismiss a result, rather than to reflect the reality of the presence of an association. Whether the heritability of handedness is indeed "slight" or "small" (as earlier in the paragraph) depends on the interpretation of a heritability of 0.25 in the very large meta-analysis of Medland. Yes, 0.75 of the variance indicates what in the case of language, was called, "a major contribution of non-genetic influences to individual language lateralization" (Oklenburg *et al.*), but that something else is not necessarily environmental factors. There is an elision here from not genetic to implying that it must therefore be environmental (and hence important). Measurement error is also a part of it, but also is what can be called 'deep chance' – quantum variation, or the chance processes by which random processes flip to right or left. They are part of the environment in some strict sense but not in any real sense. [Do we usually say that when a coin comes down heads or tails that it is due to 'environmental variation' – and if so, casino owners might get very worried]. My guess is that the majority of variation in handedness is due to randomness (and hence the C in the DC model), with only perhaps 5 or 10% at most of the variation due to real environmental variation. A full partitioning of the heritability would show that more clearly. Non-genetic is not environmental.

The last line of the paper is nice – I always like researchers at least to consider "provocative possibili[ies]", and hence that Turkheimer may have been wrong --but I will be putting my money on this one eventually being shown to be yet another example of the First Law being correct. But that use of "slight" was also mainly provocative I suspect…

The data and the analyses. Lots of improvements here.

Appendix 1 is useful, with its plot of peakLI versus mean LI in fTCD. So useful that useful that I couldn't help feeling that that it needs to be properly in the main paper if you strongly want to argue that fTCD and Peak fTCD are different. In fact they don't look very different in Appendix 1. There is a notch, but the interesting thing is the individuals whose direction would be mis-classified in the two systems. There are 16 individuals whose meanLI is positive but whose peakLI is negative, which is perhaps not surprising. More unexpected is there being only 2 individuals who move from negative to positive. Imagine that the true distribution is normal with a mean of 2.5 and SD of 2.5, with the cut point at zero, so that about 16% have an LI less than zero. Imagine that only those within .5 units (0.2 SDs) of the cut point will be affected by the peak algorithm, and half of those above the cut point move below, and vice-versa for those below the cut point. If N=200 then overall about 32 have negative LIs (16%). Of those likely to move, 9 are below the cut point and 11 above the cut point, with half of each group moving. Overall therefore about 5% move, with the same proportion in each direction. So why in the data do 16 go one way and only 2 the other? I have my prejudices but I can help wondering whether there is actually a bimodal distribution underlying this? All of that means that I think you should put appendix 1 into the main text and let people see it in clear sight. Electronic journals have few space limitations, and the point being made is an important one, particularly if you wish to claim that meanLI is better than peakLI.

Table 2 is particularly useful with its inclusions of confidence intervals. It is worth setting these out more clearly:

EHI   .03 to 0.34
QHP   0 to 0.31
LImean   0 to 0.15
LIpeak   0 to 0.20
Binary   0 to 0.25
Somers   0.08 to 0.51

[Somers is a rough estimate from the values that they give. You are welcome to get a better estimate, but I think the point will still be made.]

Looking at these I find it very difficult to assert that we can say there is heritability for handedness (although of course the Summers meta-analysis makes that very clear) but that there is not heritability for language lateralisation, as the estimates clearly all overlap substantially, or for that matter that the Somers estimate is in any major way different from the present one.

Ultimately I fear that the main conclusion is that, large though the present study is in terms of studies of the inheritance of language lateralisation, neither this nor the Somers study is compelling evidence that there is or is not heritability for language lateralisation. Taking it one step further, there is also no compelling evidence from these data that the heritability of handedness is substantially different from the heritability of language. Of course we do know that the Summers meta-analysis says something very different for handedness (but Summers was not pre-registered so perhaps it should be thrown out as well [forgive!]). I'm sorry, but that is surely an honest assessment of what there is here. Bring in some Bayesian priors then it could be swung in other directions, but the heritability estimates are pretty clear as giving no solid answers either way.

The abstract. The previous paragraph makes it clear that these confidence intervals are important and interesting. They should therefore be in the abstract, as that is what most people will read of this study. So put them in there.

The multivariate analyses. I think the second author was right here, and that it should have been included, for to do so is to treat the readers as adults who can cope with complexity. Overkill might have been a good description if all the hypotheses were already dead and stiff, but they clearly are not. Looking at the table for the multivariate analysis, there is immediately a clear demonstration of EHI and QHP showing a co-inheritance of about 0.6 (and I suspect exactly the same inheritance as the upper limit seems to include 1). I can't say that it surprises me that EHI and QHP are mainly measuring the same thing, so that their covariation is mainly genetic, but it is of interest to those of us in laterality (and reassures us that actually the boring old EHI, whatever its problems, is probably doing as well as the elegant and much more theoretically justifiable QHP).

The core of the multivariate analyses though is what LI is doing. Yes, the estimate is zero, but once again with an SE of .213, the upper limit seems to be about 0.4, which is robust, and throws . Whether there is co-inheritance of language and handedness is a key question (and one or two genetic models suggest there may be), and the EHI/LI covariance of .123 (with its SE of .113) might suggest that, although it is difficult to be sure. Either way, I am sure that these numbers are better in the public domain, as a part of open science, rather than as a throw-away remark to me lurking in a reply to comments. I think the adults would would appreciate that.

A minor point. The authors might like to reconsider their comments in the discussion about the de Kovel and Francks 2019 study. In particular, the more recent study by Gabriel Cuellar Partida, Medland, *et al*, in BioRxiv, http://dx.doi.org/10.1101/831321[1], found 41 genetic loci related to handedness, with a narrow sense heritability of 12%. There is also a reassuring similarity to the 2013 ANYAS paper by McManus, Davison and Armour which, "estimate[d] at least 40 loci are involved in determining handedness" doi: 10.1111/nyas.12102[2].

Summary
This review has been rather discursive, and is partly a review of a response to a review, so let me say what I think is needed for this paper to be acceptable:
- The abstract should contain standard errors and confidence intervals.
- The paper should contain the table summarising the multivariate analyses.
- The paper should contain the figure from Appendix 1 showing the relationship between the conventional LI peak measure and the new LI mean measure.
- A reference to Cuellar Partida and Medland should probably be included.
- The authors don't necessarily need to change much else, but they may wish to. Personally I would have included a figure showing all of the heritabilities, with their confidence intervals, which would make clear the uncertainties surrounding all of the estimates.

References
1. Partida G, Tung J, Eriksson N, Albrecht E, et al.: Genome-wide association study identifies 48 common genetic variants associated with handedness. *bioRxiv*. 2019. Publisher Full Text
2. McManus IC, Davison A, Armour JA: Multilocus genetic models of handedness closely resemble single-locus models in explaining family data and are compatible with genome-wide association studies.*Ann N Y Acad Sci*. 2013; **1288**: 48-58 PubMed Abstract | Publisher Full Text

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuropsychology of cerebral lateralisation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 01 Sep 2020

Dorothy Bishop, University of Oxford, Oxford, UK

We thank Chris McManus for his thorough reading and evaluation of the paper which has helped us improve the paper further.

Re the point about non-genetic not being the same as environmental, we are in total agreement. This is discussed in the final section, where Mitchell's emphasis on stochastic processes is alluded to. The last sentence of the abstract makes the same point. McManus, and Annett, in their genetic models, were ahead of the field in attributing a substantial part

of variance in heritability to chance.

Specific recommendations for revision

Re the specific recommendations in the report by McManus: we have attended to these, with the following exceptions:

We have not incorporated the scatterplot comparing mean and peak methods, which is in an Appendix. We agree that the best way to measure LI is far from certain, and there is a realistic possibility that there is underlying bimodality. Adopting a mean measure seems the optimal approach in the face of uncertainty, because if you take a continuous measure when the underlying reality is bimodal, then nothing is lost: indeed the comments by McManus about movement in each direction suggest one way to explore whether a continuous distribution fails to fit expectations. But if you assume bimodality when the underlying reality is continuous, the peak method induces a spurious bimodality, by creating an artefactual dip in the distribution close to zero. The reason for omitting the plot in the main paper is because (a) this paper is about heritability, rather than measurement questions, and to digress into a discussion of that issue would distract from the main point, and (b) it really doesn't make any difference to the kinds of quantitative analysis we report here – the difference between measurement methods becomes much more important if the focus is on categorical/quantitative aspects of laterality. The data are available to those interested in this issue, but are a distraction in the context of this paper.

We have not added a further figure to show the heritability estimates with CIs, as Table 2 provides this information.

Competing Interests: No competing interests were disclosed.

Reviewer Report 14 April 2020

✔   **Guy Vingerhoets** 🆔

Ghent University, Ghent, Belgium

The authors have responded clearly to all my comments. I'm happy to approve the manuscript, pending correction of some minor typos in the power analysis section (percentages). In the same section I also find the sentence "For language laterality, there is around *a 23 in 34 chance* of detecting a true but small effect of around 0.25,..." somewhat difficult to grasp. Perhaps better expressed as a percentage?

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuropsychology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 17 March 2020

https://doi.org/10.21956/wellcomeopenres.17276.r38148

✓ **Clyde Francks** ⓘD

[1] Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Many thanks to the authors for responding positively and thoughtfully to my earlier comments. I have only some minor points at this stage:

1. In the results section there are confidence intervals reported for the MZ and DZ correlation coefficients, for the separate analyses of L and R blood flows. However, for the main analysis of the laterality index, I think these correlations are only shown in Figure 3, without confidence intervals. The correlations and their confidence intervals for the laterality index could be included in the text. The finding of a shared environmental component to L and R blood flows suggests that a CE model could be tested for the laterality index, although the MZ and DZ correlations for the laterality index already indicate that C will not be significant.

2. There are typos in the section on power analysis (805%, 6773%). The heritabilities in the power section have three or four figures after the decimal point.

3. Please see the Discussion section of this paper about handedness, which also addressed randomness in early development (in the context of low heritability and weak associations with early life factors): de Kovel *et al*. (2019)[1].

References
1. de Kovel C, Carrión-Castillo A, Francks C: A large-scale population study of early life factors influencing left-handedness. *Scientific Reports*. 2019; **9** (1). Publisher Full Text

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Human neurogenetics.

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard.

- -

Version 1

Reviewer Report 26 November 2019

https://doi.org/10.21956/wellcomeopenres.16993.r36875

? **Chris McManus** (iD)
1 Research Department for Medical Education, University College London, London, UK
2 Research Department of Clinical, Educational and Health Psychology, University College London, London, UK

This is a complicated and difficult paper to review, not least because the central finding, of zero heritability for language dominance, is unexpected given both theory and other studies in the literature which have looked at the inheritance of cerebral asymmetries, including language and handedness, and their inter-relation. It has to be said immediately that the opening word of the title, "Negligible", is perhaps unnecessarily provocative given the potential problems of the present study.

Expectations. The abstract begins by saying that "it is widely assumed that individual differences in language lateralisation have a strong genetic basis". [Without going further into the issue, referring to effects as 'strong' or 'small' or 'low at best", are terms that not well defined, and suggest more of rhetoric than scientific argument; and I note in passing that I do it myself in this review]. Few studies of the inheritance of lateralisation argue for strong effects since the deep randomness of fluctuating asymmetry tends to preclude strong effects. The use of "assumed" suggests a lack of substance to any genetic basis for lateralisation, but that is surely not the case. Handedness is undoubtedly under genetic control in part, and most data suggests that handedness and language lateralisation are correlated (although again that is often described as 'small', although the tetrachoric correlations are of the order of 0.7). Various genetic models also suggest that handedness and language dominance may well be under control of the same genetic systems. Even if the single gene models of Annett and McManus are wrong, the multilocus model of McManus *et al.* (Ann NY Acad Sci, 2013)[1], the companion piece to the cited Armour *et al.* paper[2], fits well to twin and family data. Although the calculations are not in the paper, the same logic as for the single gene DC model also explains the expected genetic link between handedness and language dominance. The present study also describes the Ocklenburg *et al.* heritability for language dominance of 0.28-0.36[3] , and the Somers *et al.* heritability of 0.31[4]. All of that is more than "a wide assumption" but a strong *a priori* in Bayesian terms. Is it therefore the case that the present zero heritability for language lateralisation is a rare case that breaks Turkheimer's first law of behaviour genetics[5]? It is a strong claim, and strong evidence is needed, with it being clear that there are not other methodological issues which make it inconsistent with the rest of the

literature.

The present study with its reasonably large number of twin pairs, may be unique in the published literature, to my knowledge. However although language lateralisation is difficult and expensive to measure using fMRI, Badzakowa-Trakjov *et al* (2010; not referenced in the current paper)[6] included data on 34 MZ pairs and 11 DZ pairs (with data being available from the authors); they did not however calculate heritabilities. The Human Connectome Project with its 132 MZ and 101 DZ pairs, for which data are available for download, has information on handedness and probably has measures related to language dominance (although I haven't dug into that vast set of information), and again heritabilities will be calculable. Finally, the rapidly-growing UK Biobank has large amounts of data including brain scanning on about 18,000 individuals at present, which should rise to 100,000 soon; and twin pairs have been identified, so that eventually there could be perhaps 300 MZ and 600 MZ pairs. Some mention is perhaps worth making of other data sources.

Heritabilities. A weakness of the present study is the absence of information on confidence intervals of estimates, particularly of heritability.

A starting point is the heritability of handedness, which is graphed in Figures 2a and 2b, with heritabilities in Table 2, of .190 for EHI and .170 for QHP. These are consistent with existing data and seem to provide reassurance that there is sufficient power in the present study for identifying heritability. Not being provided with confidence intervals, I looked at the raw data in more detail. Considering just EHI, the raw correlations from Figure 2 are .181 and .134 for MZ and DZ twins. Bootstrapping the data (R=10,000) to get a confidence interval (stratified analysis in R using boot()) gives 2.5^{th} and 97.5^{th} percentiles for the correlations of MZ and DZ twins of -.049 to .426 and -.081 to .339, which are wide and of the order expected given the Ns (96MZ and 98 DZ pairs). Calculating heritability (a^2) from the bootstrapped data with umxACE() gives a 95% confidence interval of zero to 0.506, with a median of 0.232, consistent with the estimate in the paper. Overall, 21.5% of bootstrapped heritability estimates were effectively zero (<0.0001). None of that is unexpected given the overall sample size for twins, and the fact that heritability of handedness in twins is only really robust across very large samples in meta-analyses. It does however raise difficult questions for the robustness of the estimate of heritability in language dominance in the present study, with its somewhat smaller Ns.

The key data for estimating heritability of fTCD LI in the present study are the MZ and DZ correlations in Figure 2 which are .088 and -0.093. The negative DZ correlation immediately suggests that any standard modern calculation of heritability is likely to be exceedingly low. It should also be noted that Ns are smaller than for handedness, with 65MZ and 76 DZ pairs. The bootstrapped MZ correlation has a 95% range of -.096 to 0.487 (median=.176) , and the DZ correlation has a 95% range of -.184 to .264 (median = .024). Given that 12% of MZ correlations and 42% of DZ correlations are less than zero, it is unsurprising that 85% of heritability estimates were 0, although some were positive, the 97.5^{th} percentile being of 0.117, and the 99.5^{th} and 99.95^{th} percentiles being 0.195 and 0.310. The 95% confidence interval for the fTCD is therefore about 0 to 0.117. Whether that includes "negligible" is debatable.

In summary, although the number of twin pairs is large in conventional terms, the statistical power of the study is relatively low, as can be seen by the 95% confidence interval for the handedness data from 0 to .506 (in a possible range of 0 to 1). For language dominance the confidence interval is from 0 to 0.11, with most estimates being zero, which is a result of many of

the correlations being negative, or DZ correlations being greater than MZ, and again that reflects the sample size (heroic though it may be in practical terms).

More sophisticated genetic models. Genetic models of handedness and language dominance generally assume – and are effective in so doing – that a single genetic system causes both handedness and language dominance, albeit they are not correlated perfectly due to random variation (which is indistinguishable from error variance but is actually due to deep chance). It would seem sensible therefore to fit a bivariate genetic model to the handedness and language dominance data. I tried it on the raw data, and there was little evidence of shared genetic variance between the handedness and language dominance phenotypes, but that is hardly surprising given the zero heritability for language dominance. Bootstrapping is more complex, and I haven't tried it but it should be done.

Language dominance and handedness. Most work on language lateralisation has found that atypical lateralisation is correlated with left-handedness (and for instance the Badzakowa-Trakjov *et al.* paper found a correlation of .357 (p<.001) between word generation and handedness[6]). Many other studies find a similar correlation (and the present paper quotes Vingerhoets (2019)[7] with estimates of 6.5% of right handers and 10-15% of left-handers having atypical language laterality, although the latter estimate may be on the low side). The uncited meta-analysis by Carey and Johnstone (2014)[8] contains a systematic review using different methods of assessing language dominance and finds risk ratios of 1.36 for Wada studies, 1.22 for dichotic/tachistoscopic studies, 1.21 for fMRI/TDS/ECT studies. Considering just the eight fTCD studies, of 516 right-handers, 6.4% showed atypical dominance, compared with 41.2% of 369 left-handers.

The question immediately arises as to the correlation of language dominance and handedness in the present study, which is not, I think, reported. The correlation between fTCD LI and the EHI is -.0642, p=.2841. Using writing hand and fTCD LI dichotomised around zero as Left or Right, 21.4% of 238 right-handers and 23.8% of left-handers have atypical language lateralisation which is clearly not significant. Also of interest is that 21.7% of all participants seem to have atypical language lateralisation which is higher than is usual and raises questions about the measure of language lateralisation and the sample.

Calculating the laterality index for fTCD. Language lateralisation indices in most previous studies using fTCD have followed Deppe *et al.* and assessed the maximum difference between right and left flow[9]. That might well have problems since values of zero are inevitably made very unlikely, and results in a dip around zero. The present study uses a new algorithm based on calculating the mean difference between flow in the right and left arteries during the event window. That seems sensible, but it is not at all clear whether the results are equivalent. A previous study (Woodhead, Rutherford and Bishop, 2018)[10] compared the Deppe method with the mean method and reported a correlation of laterality indices of 0.97, although data were not shown and the participants were almost entirely right-handed. Whether means were different was unclear, and a scattergram would have been useful. An advantage is claimed to be that the "bimodality of the laterality index distribution is not seen when the means-based method is used", although that is not self-evidently good when there are strong *a priori* expectations that laterality indices may well be bimodal. The present paper doesn't include I think the Deppe method indices in the data files and therefore no further exploration could be carried out.

Taken overall, the lack of an association of language lateralisation with handedness – which was

found in most other previous studies – coupled with a high proportion of atypical language lateralisation overall, and a new method of data calculation does raise worries that the present results are in part artefactual. It would be reassuring to know that precisely the same results were obtained when the data were processed with DopOSCCI.

The sample. Most studies using fTCD have used undergraduate participants or typically developing children. Although it is not mentioned in the abstract, the present participants are from the Wilson and Bishop (2018)[11] study where the study, "us[ed] a sampling approach with the aim of including around 75% twin pairs where one or both had parental report of language or literacy difficulties". The present paper reports that 43% of MZ pairs were concordant for language difficulties compared with 21% of DZ pairs (p.5). The implication is not only that many of the sample have language or literacy problems, with 55% of MZ twins and 44% of DZ twins having problems, but that also it is probable that there is an inherited component. Fitting an AE model gives estimates of heritability of .279 (CI = .130 to .416). Overall it seems clear that this population is probably far from representative of the general population. The results should probably be interpreted carefully.

Summary. Interesting though this study is, there are multiple reasons to treat it with great care, and in particular the headline conclusion of "negligible heritability of language laterality" may be somewhat overstated. There are potential problems with:
- the relatively small sample size for a twin study;

- the wide confidence intervals for heritability of handedness which generally does show heritability in twins but here has a lower confidence interval of zero;

- a 95% bootstrapped confidence interval for the heritability of language lateralisation which has an upper limit of 0.11;

- the lack of bivariate modelling of handedness and language dominance with a single underlying gene;

- the lack of an association of handedness and language lateralisation, although most studies find such an association, for which there are theoretical expectations;

- the use of a new algorithm for calculating the laterality index which also appears to give a high rate of atypical dominance, and the lack of calculations using the traditional algorithm;

- the sample not being representative of the population but being selected for a high rate of language and literacy difficulties which themselves appear to be heritable.

References

1. McManus IC, Davison A, Armour JA: Multilocus genetic models of handedness closely resemble single-locus models in explaining family data and are compatible with genome-wide association studies.*Ann N Y Acad Sci*. 2013; **1288**: 48-58 PubMed Abstract | Publisher Full Text

2. Armour JA, Davison A, McManus IC: Genome-wide association study of handedness excludes simple genetic models.*Heredity (Edinb)*. 2014; **112** (3): 221-5 PubMed Abstract | Publisher Full Text

3. Ocklenburg S, Ströckens F, Bless JJ, Hugdahl K, et al.: Investigating heritability of laterality and

cognitive control in speech perception.*Brain Cogn*. **109**: 34-39 PubMed Abstract | Publisher Full Text

4. Somers M, Ophoff RA, Aukes MF, Cantor RM, et al.: Linkage analysis in a Dutch population isolate shows no major gene for left-handedness or atypical language lateralization.*J Neurosci*. 2015; **35** (23): 8730-6 PubMed Abstract | Publisher Full Text

5. Turkheimer E: Three Laws of Behavior Genetics and What They Mean. *Current Directions in Psychological Science*. 2000; **9** (5): 160-164 Publisher Full Text

6. Badzakova-Trajkov G, Häberling IS, Roberts RP, Corballis MC: Cerebral asymmetries: complementary and independent processes.*PLoS One*. 2010; **5** (3): e9682 PubMed Abstract | Publisher Full Text

7. Vingerhoets G: Phenotypes in hemispheric functional segregation? Perspectives and challenges. *Phys Life Rev*. 2019. PubMed Abstract | Publisher Full Text

8. Carey DP, Johnstone LT: Quantifying cerebral asymmetries for language in dextrals and adextrals with random-effects meta analysis.*Front Psychol*. 2014; **5**: 1128 PubMed Abstract | Publisher Full Text

9. Deppe M, Knecht S, Henningsen H, Ringelstein E: AVERAGE: a Windows® program for automated analysis of event related cerebral blood flow. *Journal of Neuroscience Methods*. 1997; **75** (2): 147-154 Publisher Full Text

10. Woodhead ZVJ, Rutherford HA, Bishop DVM: Measurement of language laterality using functional transcranial Doppler ultrasound: a comparison of different tasks.*Wellcome Open Res*. 2018; **3**: 104 PubMed Abstract | Publisher Full Text

11. Wilson AC, Bishop DVM: Resounding failure to replicate links between developmental language disorder and cerebral lateralisation.*PeerJ*. 2018; **6**: e4217 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: The authors had inadvertently omitted the data and code files for this paper, and I contacted the authors to point this out. This would not have affected the impartiality of the review.

Reviewer Expertise: Neuropsychology of cerebral lateralisation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 22 Feb 2020

Dorothy Bishop, University of Oxford, Oxford, UK

Response to reviewers

Our thanks to all three reviewers for their thoughtful comments on this article. There was strong concordance in their critiques, and so we will respond to the general points that were made before going on to cover specific points by individual reviewers.

General comments
This paper, and the criticisms, get to the nub of issues that have become prominent in discussions of the so-called 'reproducibility crisis'. The fundamental question is how do we determine whether a null result, like the one reported here, is a type II error. We agree with the reviewers that it is not enough just to present a null result and conclude that the true effect is zero. A null result could arise because the study is underpowered to detect a true but small effect, or because the methods lack reliability or validity. In the Discussion we considered these possibilities and concluded that we could not draw a strong conclusion about our null finding, though we could specify a likely upper bound for heritability. However, the use of the word 'negligible' in the title attracted criticism from all reviewers, and we accept that this is too value-laden and have deleted it.

Another point is that when evaluating research findings we should never rely on evidence from a single study. Again, we noted in the Discussion how Bishop's earlier findings of insignificant heritability of handedness had been overturned by subsequent, larger studies, which emphasises the need for caution. The suggestions by reviewers to try other analyses to see if the results look different are consistent with a Bayesian approach that demands especially strong evidence to overturn a strong prior belief.

There is, however, a need for caution here. Prior expectations come from at least two sources. First, there is the issue of needing a plausible mechanism, and we agree that a genetic basis for individual differences in a neurobiological phenotype is plausible. The main source of priors, however, will be previous literature. All three reviewers have cited additional sources for genetic influences on laterality. But there is a question of just how much confidence one should place in prior literature, given that there are many inconsistent findings, plus three systematic biases that distort results: publication bias, p-hacking and citation bias. It is always difficult to discuss these biases, because it looks as if one is singling out other researchers for criticism, and impugning their integrity. Nevertheless, their prevalence is not in doubt (Greenwald, 1975; Sterling, Rosenbaum, & Weinkam, 1995; Fanelli, 2012), and there is circumstantial evidence that they have influenced the field of laterality. Consider, for instance two twin studies, one on handedness by Davis and Annett (1994), and one on structural brain asymmetry by Geschwind et al (2002). Both studies are interpreted by their authors as supporting genetic models of cerebral lateralisation, but

neither reported heritability estimates for laterality (or the zygosity-based correlations that underpin these), despite having the data available for doing this. One conclusion is that the heritability estimates were not convincing, and so went unreported.

The role of publication bias in distorting beliefs about solidness of findings was nicely documented in a simulation by Nissen et al (2016). Where this effect is compounded by p-hacking, then we can end up with up with solid beliefs based purely on the fact that we are sampling biased evidence. On top of that we all have a tendency to confirmation bias, which means that even when null findings are published, we tend to disregard them, which further biases the evidence (Bishop, 2020).

The field of laterality research is at particular risk of bias because there are so many different ways of conceptualising the phenotype. This point was made with regard to handedness by Bishop (1990), who noted that if you do not prespecify in advance how you plan to convert a handedness scale into groups, then you raise the chance of finding a 'significant' result to well above 5%. This is equally true for other types of laterality, where there is no agreement about accepted measurement practices. And where laterality measures are part of a larger battery of tests, then it is likely that results on heritability will usually be published only if significant. The reviewers note the need for cautious interpretation of our results, and we agree, but in the absence of pre-registered studies, we also need to adopt a cautious stance to the prior literature, as there is a substantial risk of type I error. Large sample sizes can save us from type II errors but they are no defence against type I errors if p-hacking is possible. Findings that have been replicated using the same methods can be given much more weight than one-off studies.

Interpretation of evidence in this field is further complicated by the fact that there are many different forms of laterality – as well as handedness we have both structural and functional brain laterality. Once we move from handedness, little is known about the reliability of these different measures of phenotype, but it is clear that they are not interchangeable, and the relationships between them are not clearly understood.

These points are amplified below when dealing with specific points raised by reviewers; the final paragraph of the Discussion has been amended to make it clear that we are not claiming that we have definitely proven a null result, but rather that very low, or even absent heritability of functional language lateralisation should at least be treated as a realistic contender, rather than dismissed as implausible.

Responses to specific comments by reviewers

1. Clyde Francks

1.1. *Request for confidence interval for heritability estimate and goodness of fit statistics*

These have now been provided in Table 2.

1.2. *Additional papers*
Thanks for drawing our attention to these papers that include genetic analysis of structural

brain asymmetries. We should not have overlooked these papers which are indeed familiar to the first author, but in mitigation neither paper mention the words genetic or heritable in the title, and only Guadalupe et al mentioned 'heritability' in the keywords, so it is easy to miss these when conducting a systematic search for relevant papers. They are now included in the account of structural asymmetries.

The recent GWA studies of handedness are now mentioned in the Discussion, and we note that while these show very low SNP-based heritability, this does not preclude successful gene mapping – though the sample sizes required make it unlikely this will be feasible using measures of language function based on fMRI or fTCD. Unfortunately, UK Biobank did not include language function activation methods in fMRI.

1.3 *Alternative pipeline for Doppler method*
CF notes: "*Somers et al. (2015)[5] found a heritability of 31% for atypical language lateralisation (coded as a binary variable), using functional transcranial Doppler ultrasound in a multi-generational pedigree sample. If the same data processing pipeline from that study would be applied in the current study, might the results no longer be discrepant between the two studies?*" Somers used the traditional peak-based approach to analysing the fTCD data. Although for some analyses they used a continuous laterality index, heritability was reported only for a binary category of typical vs atypical. Correspondence with Dr Somers confirmed that this method was not suited to continuous data.

Results from analysis of peaks is now added to Table 2 for completeness, but please note that one reason for abandoning the peak approach is that it forces the data into a bimodal distribution. We moved away from this approach when we found that there were some individuals with L-R difference waves that hovered around zero: depending on whether the peak difference was greater for L or R, the peak would be taken at that point, and for those with very slight differences between the sides, this could seem fairly arbitrary. See Woodhead, Rutherford and Bishop, 2018, for discussion of this point.

1.4 *Heritability of L and R*
CF notes: "*In general, measuring asymmetry necessarily involves calculating some kind of difference score, which means that error variance in both left and right measures can affect the asymmetry measure. This may partly explain the low heritabilities of asymmetry indexes, and the authors could acknowledge this. In brain structural analysis (of e.g. grey matter volumes, surface areas, thicknesses), we typically see higher heritabilities for L and R separately than for the asymmetry index (L-R)/(L+R). Might it be informative to calculate heritabilities for L and R, in the context of functional transcranial Doppler sonography? If L and R are not themselves heritable, there may be a problem with the approach.*"

This is an interesting suggestion, which we have adopted. The correlation between the L and R flow measures *within* individuals is very high (close to .9). As now described in the paper, the raw L and R mean flow measures showed significant twin-twin correlations, but there was no effect of zygosity, suggesting that the similarity between twins was partly due to shared environment, rather than genetic influences. This is an intriguing result, but it is hard to interpret. This is not an age effect: age was not correlated with the flow measures, and residualising scores on age and sex did not make any difference. There have in the past

been environmental explanations proposed for handedness, including shared in utero environment, which we have now alluded to. CF implies that a lack of heritability of the individual L and R flow measures may indicate a problem with the approach, but another possibility is that this trait is simply not under genetic control. We are not aware of any previous literature on heritability of cerebral blood flow.

1.5. *Support for assumption of strong heritability*
Abstract: The authors state that 'it is widely assumed that individual differences in language lateralisation have a strong genetic basis'. Are there references to support this? I am not sure that many researchers working on the genetics of laterality have this impression (this one does not).

This comment has been removed, as it is based solely on informal impressions – mainly surprised reactions when these results are discussed.

1.6 *Rewording*
The Discussion can be revised in parts, to reflect the recent literature indicated above.
The authors wrote that 'Our data are compatible with a more radical model in which language laterality is the consequence of a general population left-sided brain bias for language which does not show any individual variation.' For clarity, it would help to make explicit that genetic variation is meant here. I agree that the author's data are compatible with this model, but the limited sample size means that the data are also compatible with a heritability up to whatever level is encompassed by the confidence interval around the best estimate.

We have reworded as requested

2. **Guy Vingerhoets**

2.1 *Methods clarification*
The only detail that puzzles me is the period of interest of the fTCD procedure. In one paragraph a start and stop cue indicating a 10s period is mentioned during which the child describes the 12s cartoon. In another paragraph the period of interest is defined as 4 to 14s after the cue to speak, but this method was abandoned for the 'whole period of interest'. Would that be the 10s period between both cues then?

This has now been written more clearly to explain how the peak method works, and how it differs from the mean method.

2.2 *Validity of fTCD laterality*
fTCD- derived LI's offer only crude estimations of asymmetry as they will pick up velocity changes due to co-activation of many other mental components that take place in the MCA territory and that may not be related to language. The child has just seen a 12s cartoon and now must describe it. Attention, memory, visual imagery, movement(?), receptive and productive language areas all come into play and most of these functions will be associated with activation of the lateral cortex. Besides not being very region-specific, fTCD does not allow for the use of control-tasks that can correct for task-unspecific activation.

This is a point that is often raised by those who work with fMRI, and the Oxford group has

given it much consideration – for instance, using tasks based on Mazoyer et al's sentence generation and control tasks, Woodhead et al (2018) computed laterality indices based on the difference in lateralised activation for sentence vs list generation. We showed this made little overall difference, because list generation was not lateralised. The central point is that while, of course, there are numerous nonlinguistic factors involved in performing the animation description task, any nonlateralised activity is subtracted out by our analytic procedure, when we take the difference score. One can see in the waveforms associated with different tasks periods when blood flow increases and decreases – e.g. the blood flow increases just before starting to talk, then falls away. But so long as these are symmetrical effects they do not affect the laterality index, which is based on the difference waveform.

2.3 *Relative size of effect for fTCD and fMRI*
We are currently doing some direct comparisons of laterality indices from fMRI and fTCD with adults, and are finding good levels of agreement of laterality indices for equivalent tasks, but we would caution against attempting direct comparisons of magnitude of effects, because of the very different ways in which brain activation is measured. With fMRI a general linear model is used to measure how the brain activation relates to experimental design variables, and then t-statistics of the voxels within the left and right hemisphere ROIs are computed and thresholded, and either the count (extent) or the sum (magnitude) of supra-threshold t-values is calculated in each hemisphere. With fTCD, the blood flow signal is epoched, normalised and baseline corrected, so that both left and right sensors have an average signal of zero in a resting time period at the start of each trial. The intensities of the left and right signals are then directly compared within a period of interest when the participant is performing the task.

3. Chris McManus

3.1 *Plausibility of laterality based on prior literature*
CM *"The use of "assumed" suggests a lack of substance to any genetic basis for lateralisation, but that is surely not the case."*

All reviewers challenged this statement, so it has been removed. We would stress that we are not arguing that all laterality is non-heritable, and we agree that as regards handedness, there is enough evidence to be confident in heritability of around .25, and growing evidence of genetic influences on structural brain asymmetry in some regions (as now reviewed more fully).
However, this is not the same as functional language laterality, where there is a paucity of evidence. As far as we are aware, the sum total of prior evidence is contained in 3 studies:
a) Ocklenburg, Bryden, and Somers. CM mentions the Ocklenburg study that we cited. This found zero heritability for the standard dichotic listening task, but heritability of .28 to .36 for a task condition with directed attention. The authors of that study concluded that their results: "implicate a major contribution of non-genetic influences to individual language lateralization."
b) Bryden (now mentioned in our revision), used two measures in both parents and two siblings from 49 families. Although he found one statistically significant association between mother and child, he drew attention to inconsistent findings – not only were the correlations between siblings negative, but one of the highest correlations was between

mother and father. As he wryly noted, "This correlation would suggest non-random mating for laterality, a characteristic that one would hardly expect to be of significance in selecting one's spouse. " (p. 206). He concluded: " ..the present study has failed to find any particularly compelling evidence for a genetic basis for speech lateralization. While the problems associated with the use of an indirect measure of only moderate reliability may have doomed this study from the start, it does suggest that one should at least consider seriously the hypothesis that speech lateralization is primarily determined by environmental factors" (p.209). The split half reliability of the measures was .61 and .66.

c) The strongest evidence for heritability of language laterality on a functional brain measure is from the Somers et al study mentioned above using fTCD with a multi-generational pedigree sample from an isolated community. The heritability of atypical language lateralisation (coded as a binary variable) was 0.31. This sample was not at all representative of the general population: it was deliberately selected to have relatively low genetic heterogeneity and to include families with several left-handed members. This might affect generalisability of findings, but the most serious limitation of the sample was that the selection method was biased toward phenotypic similarity of those who were in the sample; i.e. insofar as handedness is related to language laterality, then selecting only families with at least two left-handers per generation could artificially inflate within-family similarity for laterality. On the other hand, a potential advantage of the study was the use of a pedigree-based method of analysis, which gives higher power than a method reliant just on twin pairs.

Taken together, the evidence is far from compelling, with moderate reliabilities, modest samples sizes, and heritability estimates compatible both with zero and with modest values such as .3 or so.

3.2 *Other studies*

CM notes "Badzakowa-Trakjov et al (2010; not referenced in the current paper)[6] included data on 34 MZ pairs and 11 DZ pairs (with data being available from the authors); they did not however calculate heritabilities."

This was not cited precisely because it is hard to derive any information about heritability from the study, because the sample was not only small, but also was selected to be biased to include discordant pairs (half the pairs studied had discordant handedness). Following the prompt from CM, we downloaded the data (available from PLOS One). The correlation between 34 MZ twin pairs on Word Generation was -.11 and for 11 DZ pairs it was .32. This is hardly encouraging for a genetic theory of language lateralisation, but, for the reasons noted above, it would be rash to draw much of a conclusion from this.

The Human Connectome Project with its 132 MZ and 101 DZ pairs, for which data are available for download, has information on handedness and probably has measures related to language dominance.

As CF points out above, there are some analyses of structural asymmetries. As far as we know there are no data on functional asymmetry

Biobank: again, there are analyses of handedness, but, although there are MRI data on a

subset of individuals, the functional MRI in Biobank did not include a language task. We believe there are moves afoot to derive a laterality measure from resting state fMRI, but it is unclear how this would relate to laterality on something like a word or sentence generation task.

3.3 *Need for confidence intervals*
Thanks for the computations of confidence intervals around heritability estimates – these were helpful and in line with our bootstrapped computations, which are now given.

3.4 *Multivariate models*
The second author in fact had suggested we include these, but the first author thought this would be overkill, given that, as the reviewer points out, it would involve testing whether zero heritability in one trait is shared with low heritability in another. These results are given here for completeness, but not incorporated in the main paper. Path estimates are shown for the AE model, as all C estimates were zero.

Unsquared standardized estimates of paths from multivariate model.
Asterisks denote paths that are statistically significant at .05 level.

	Term	Measure	Estimate.EQL	SE.EQL	Estimate.QL	SE.QL	Estimate.L	SE.L
1	A	EHI	0.458*	0.098				
2	A	QHP	0.243	0.124	0.343*	0.093		
3	A	LI	0.123	0.113	-0.152	0.123	0.000	0.213
4	E	EHI	0.889*	0.051				
5	E	QHP	0.401*	0.068	0.814*	0.043		
6	E	LI	-0.134	0.066	0.039	0.064	0.971	0.039

N.B. EQL loadings for paths to all three measures: QL paths independent of EHI; L paths specific to language LI.

Information Criteria:

	df Penalty	Parameters Penalty	Sample-Size Adjusted
AIC:	5562.0751	8922.075	8923.893
BIC:	-489.8496	8976.597	8929.033

3.5 *Language dominance and handedness*
Relevant data on this are now added.
We note also that the lack of association is inconsistent with previous literature which tends to find a significant association between handedness and laterality measures.

3.6 *Numbers with atypical dominance*
CM notes "A further query about validity of the data is raised by the relatively high proportion with atypical lateralisation."

With fTCD, the proportions with atypical lateralisation are entirely dependent on the language measure used, as shown by Woodhead et al (2018). This leads us to conclude that language lateralisation is not a fixed binary property of the brain, but varies in degree according to task demands. Some discussion of this point is now added.

3.7 *Use of the Deppe original method*
Whether means were different was unclear, and a scattergram would have been useful. An advantage is claimed to be that the "bimodality of the laterality index distribution is not seen when the means-based method is used", although that is not self-evidently good when there are strong a priori expectations that laterality indices may well be bimodal.

A scattergram has now been added in the Appendix. We disagree regarding the 'strong a priori expectations': there are some theories that treat handedness as bimodal, but others that do not, and, for language laterality, there is no particular reason to assume bimodality. Data on LIs obtained using the 'peak' method are now added.

CM states "It would be reassuring to know that precisely the same results were obtained when the data were processed with DopOSCCI."
Please see Woodhead et al (2018) where we note: " the R script had been developed in our group to fulfil the need for a reproducible and efficient method for processing large numbers of datasets, without using commercial (Matlab) software that required a licence (see Wilson & Bishop, 2018). As with DopOSCCI the analytic pipeline closely followed procedures developed by Deppe et al. (2004), with one additional option: the possibility of identifying brief periods of signal spiking or dropout and interpolating over these, to avoid rejecting trials. Wilson and Bishop (2018) compared results from DopOSCCI and the R script and found only small differences in the LIs computed by the two methods". Given that we have spent a great deal of time developing scripts that allow us to process the data efficiently in R, and checking that it gives results highly consistent with the prior DopOSCCI approach, we do not think it reasonable to be asked to revert to the prior method. Everyone makes errors of course, and we cannot rule out that there may be a bug somewhere, but we feel it is unreasonable to expect us to keep analysing our data using different methods. We do agree that there is an element of arbitrariness in the data processing pipeline for fTCD – results will vary depending on the sequence of operations, treatment of outliers, the period of interest, and whether peaks or means are used, but in our experience these differences have only minor impact on the final laterality index. We are interested to evaluate other approaches, but the method presented here was judged to be optimal and is the best we can do at present. Our scripts are available, together with the raw data, so others are welcome to try different approaches.

3.8 *Sample*
Overall it seems clear that this population is probably far from representative of the general population.
While this is true, in our prior paper, we showed that there was no difference in cerebral lateralisation for children with and without language problems, and similar results to those obtained by Groen et al (2012) with singleborn children. Prior studies have found no evidence for any genetic link between language disorders and lateralisation (Bishop, 2001; 2005).

Overall
The criticisms offered by the reviewers are generally fair, and we are happy to moderate the way in which the current results are described to clarify the limited conclusions that can be

drawn. For the reasons stated at the outset, it is important that null results are published so that future meta-analyses are not biased in favour of positive findings. We hope that there will be more studies on this topic so that in future it might be possible to incorporate these data in a meta-analysis, to give a more precise estimate of heritability of language laterality.

References

Bishop, D. V. M. (1990). How to increase your chances of obtaining a significant association between handedness and disorder. *Journal of Clinical and Experimental Ñeuropsychology, 12*, 786-790.
Bishop, D. V. M. (2001). Individual differences in handedness and specific language impairment: evidence against a genetic link. *Behavior Genetics, 31*, 339-351.
Bishop, D. V. M. (2005). Handedness and specific language impairment: a study of 6-year-old twins. *Developmental Psychobiology, 46*, 362-369.
Bishop, D. V. M. (2020). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research. The 45th Sir Frederic Bartlett Lecture *Quarterly Journal of Experimental Psychology, 73*(1), 1-19. doi:10.1177/1747021819886519
Bryden, M. P. (1975). Speech lateralization in families: A preliminary study using dichotic listening. *Brain and Language, 2*, 201-211.
Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*(3), 891-904. doi:10.1007/s11192-011-0494-7
Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*(1), 1-20.
Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *eLife, 5*, e21451. doi:10.7554/eLife21451.001
Sterling, T. D., Rosenbaum, W. D., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician, 49*(1), 108-112.

Competing Interests: None

Reviewer Report 12 November 2019

? **Guy Vingerhoets** iD

Ghent University, Ghent, Belgium

By comparing monozygotic and dizygotic twin pairs on measures of hand and language laterality the authors aim to evaluate the genetic influence of these asymmetries. Consistent with previous reports a modest heritability effect of slightly over .20 was found for handedness, but for

language laterality twin-cotwin correlations were close to zero. The authors conclude that heritability of language lateralization is low at best.

This is a well written and methodologically sound and detailed study investigating the genetic effect on hand and language laterality. In the introduction the authors explain the rationale for their approach and review the available data on genetic variation of laterality. I would suggest to slightly rearrange this section as after shortly introducing genetic variation on handedness and language, they move on to describe the genetics of structural asymmetries for two paragraphs, to return to behavioral laterality in the next paragraph. Maybe it is better to treat structural and functional asymmetry more separately, with the latter being more relevant to the present study than the former. In general, the findings of these studies reveal only moderate effects which seems in contrast with the claim in the first sentence of the abstract.

I'm not an expert in heritability research, but the approach of the authors seems methodologically sound. The number of twins included is high, and a power analysis addresses the feasibility to detect an effect. The sample is clearly described, and laterality assessment of handedness and language are explained in detail. The only detail that puzzles me is the period of interest of the fTCD procedure. In one paragraph a start and stop cue indicating a 10s period is mentioned during which the child describes the 12s cartoon. In another paragraph the period of interest is defined as 4 to 14s after the cue to speak, but this method was abandoned for the 'whole period of interest'. Would that be the 10s period between both cues then?

Results are illustrated with distribution plots and scatterplots and the results of the AE model are presented in the classical and ordinal version.

In the discussion, the authors mention the consistent finding regarding handedness (genetic factors account for about 20% of the variance), and quickly move on to interpret the close to zero correlation for twin-cotwin language laterality findings.

They start by questioning the validity of the fTCD derived language laterality index and raise three arguments in favor of their measure: (1) split-half reliability is high, (2) fTCD laterality indices (LI) reflect the expected population bias, and (3) the low heritability is consistent with other reports on structural and functional lateralization. They further argue that their findings cannot dismiss the possibility that there is a small but real genetic effect on language lateralization and that the sample may be too small to detect it. In other words, while their data suggest no role for genetic factors on language laterality, the authors leave open the possibility of a small (the title says negligible, which I find a somewhat subjective term) effect. The argument used for this interpretation is grounded on (potentially insufficient) sample size rather than (potentially invalid) measurement. While this might be the case, I would argue that fTCD measurement has several drawbacks the authors may wish to consider. Although not perfect, the reliability of fTCD is overall reasonable (Stroobant & Vingerhoets, 2001[1]; Vingerhoets & Stroobant, 2002[2]). But reliability is not validity. fTCD measures blood flow velocity in the basal part of the middle cerebral arteries. This artery supplies blood to most of the lateral surface of the brain or roughly 80% of each hemisphere. As a result, fTCD- derived LI's offer only crude estimations of asymmetry as they will pick up velocity changes due to co-activation of many other mental components that take place in the MCA territory and that may not be related to language. The child has just seen a 12s cartoon and now must describe it. Attention, memory, visual imagery, movement(?), receptive and productive language areas all come into play and most of these functions will be associated with

activation of the lateral cortex. Besides not being very region-specific, fTCD does not allow for the use of control-tasks that can correct for task-unspecific activation. My point is that fTCD-derived LI's of language tasks may be sufficient to (reliably) reflect the population bias of left hemisphere language dominance, but that they may not be sufficient to provide valid markers of lateralization strength. Therefore, its use as a binary variable may be more successful than when used as a continuous variable. Note the fTCD index plotted in Figure 1: the mean LI lies between .20 to .30 in favor of the left hemisphere. This is not very lateralized which may be due to the joint activation of other mental functions that dilute the asymmetry of the language component. Compare this mean LI with the fMRI-derived mean value of about .65 in favor of the left hemisphere using a sentence production task based on cartoon drawings (Mazoyer *et al*. 2004[3]). Although the fMRI LI's were based on whole hemisphere data, the application of a control task was able to filter out much of the non-relevant general mental activation, a procedure not possible with fTCD. Finally, the fMRI results were obtained in adults, not in 6 to 12-year-old children whose language lateralization may be more variable, in particular when they have language or literacy difficulties.

This consideration should not do short to the excellent work presented in this paper. It is simply not feasible to place all these children in an MRI scanner. The results presented provide important information on the effect of genetic factors on language laterality in children. I broadly agree with the conclusion that while we cannot completely dismiss the idea that genes have no role in language laterality, its effect is likely to be modest. In order for that message to come across, the word 'negligible' might need some fine tuning.

References
1. Stroobant N, Vingerhoets G: Test-retest reliability of functional transcranial Doppler ultrasonography. *Ultrasound in Medicine & Biology*. 2001; **27** (4): 509-514 Publisher Full Text
2. Vingerhoets G, Stroobant N: Reliability and validity of day-to-day blood flow velocity reactivity in a single subject: an fTCD study. *Ultrasound in Medicine & Biology*. 2002; **28** (2): 197-202 Publisher Full Text
3. Mazoyer B, Zago L, Jobard G, Crivello F, et al.: Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness.*PLoS One*. 2014; **9** (6): e101165 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuropsychology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 22 Feb 2020

Dorothy Bishop, University of Oxford, Oxford, UK

Response to reviewers

Our thanks to all three reviewers for their thoughtful comments on this article. There was strong concordance in their critiques, and so we will respond to the general points that were made before going on to cover specific points by individual reviewers.

General comments

This paper, and the criticisms, get to the nub of issues that have become prominent in discussions of the so-called 'reproducibility crisis'. The fundamental question is how do we determine whether a null result, like the one reported here, is a type II error. We agree with the reviewers that it is not enough just to present a null result and conclude that the true effect is zero. A null result could arise because the study is underpowered to detect a true but small effect, or because the methods lack reliability or validity. In the Discussion we considered these possibilities and concluded that we could not draw a strong conclusion about our null finding, though we could specify a likely upper bound for heritability. However, the use of the word 'negligible' in the title attracted criticism from all reviewers, and we accept that this is too value-laden and have deleted it.

Another point is that when evaluating research findings we should never rely on evidence from a single study. Again, we noted in the Discussion how Bishop's earlier findings of insignificant heritability of handedness had been overturned by subsequent, larger studies, which emphasises the need for caution. The suggestions by reviewers to try other analyses to see if the results look different are consistent with a Bayesian approach that demands especially strong evidence to overturn a strong prior belief.

There is, however, a need for caution here. Prior expectations come from at least two sources. First, there is the issue of needing a plausible mechanism, and we agree that a genetic basis for individual differences in a neurobiological phenotype is plausible. The main source of priors, however, will be previous literature. All three reviewers have cited additional sources for genetic influences on laterality. But there is a question of just how much confidence one should place in prior literature, given that there are many inconsistent findings, plus three systematic biases that distort results: publication bias, p-hacking and citation bias. It is always difficult to discuss these biases, because it looks as if one is singling out other researchers for criticism, and impugning their integrity. Nevertheless,

their prevalence is not in doubt (Greenwald, 1975; Sterling, Rosenbaum, & Weinkam, 1995; Fanelli, 2012), and there is circumstantial evidence that they have influenced the field of laterality. Consider, for instance two twin studies, one on handedness by Davis and Annett (1994), and one on structural brain asymmetry by Geschwind et al (2002). Both studies are interpreted by their authors as supporting genetic models of cerebral lateralisation, but neither reported heritability estimates for laterality (or the zygosity-based correlations that underpin these), despite having the data available for doing this. One conclusion is that the heritability estimates were not convincing, and so went unreported.

The role of publication bias in distorting beliefs about solidness of findings was nicely documented in a simulation by Nissen et al (2016). Where this effect is compounded by p-hacking, then we can end up with up with solid beliefs based purely on the fact that we are sampling biased evidence. On top of that we all have a tendency to confirmation bias, which means that even when null findings are published, we tend to disregard them, which further biases the evidence (Bishop, 2020).

The field of laterality research is at particular risk of bias because there are so many different ways of conceptualising the phenotype. This point was made with regard to handedness by Bishop (1990), who noted that if you do not prespecify in advance how you plan to convert a handedness scale into groups, then you raise the chance of finding a 'significant' result to well above 5%. This is equally true for other types of laterality, where there is no agreement about accepted measurement practices. And where laterality measures are part of a larger battery of tests, then it is likely that results on heritability will usually be published only if significant. The reviewers note the need for cautious interpretation of our results, and we agree, but in the absence of pre-registered studies, we also need to adopt a cautious stance to the prior literature, as there is a substantial risk of type I error. Large sample sizes can save us from type II errors but they are no defence against type I errors if p-hacking is possible. Findings that have been replicated using the same methods can be given much more weight than one-off studies.

Interpretation of evidence in this field is further complicated by the fact that there are many different forms of laterality – as well as handedness we have both structural and functional brain laterality. Once we move from handedness, little is known about the reliability of these different measures of phenotype, but it is clear that they are not interchangeable, and the relationships between them are not clearly understood.

These points are amplified below when dealing with specific points raised by reviewers; the final paragraph of the Discussion has been amended to make it clear that we are not claiming that we have definitely proven a null result, but rather that very low, or even absent heritability of functional language lateralisation should at least be treated as a realistic contender, rather than dismissed as implausible.

Responses to specific comments by reviewers

1. Clyde Francks

1.1. *Request for confidence interval for heritability estimate and goodness of fit statistics*

These have now been provided in Table 2.

1.2. *Additional papers*
Thanks for drawing our attention to these papers that include genetic analysis of structural brain asymmetries. We should not have overlooked these papers which are indeed familiar to the first author, but in mitigation neither paper mention the words genetic or heritable in the title, and only Guadalupe et al mentioned 'heritability' in the keywords, so it is easy to miss these when conducting a systematic search for relevant papers. They are now included in the account of structural asymmetries.

The recent GWA studies of handedness are now mentioned in the Discussion, and we note that while these show very low SNP-based heritability, this does not preclude successful gene mapping – though the sample sizes required make it unlikely this will be feasible using measures of language function based on fMRI or fTCD. Unfortunately, UK Biobank did not include language function activation methods in fMRI.

1.3 *Alternative pipeline for Doppler method*
CF notes: "Somers et al. (2015)[5] found a heritability of 31% for atypical language lateralisation (coded as a binary variable), using functional transcranial Doppler ultrasound in a multi-generational pedigree sample. If the same data processing pipeline from that study would be applied in the current study, might the results no longer be discrepant between the two studies?" Somers used the traditional peak-based approach to analysing the fTCD data. Although for some analyses they used a continuous laterality index, heritability was reported only for a binary category of typical vs atypical. Correspondence with Dr Somers confirmed that this method was not suited to continuous data.

Results from analysis of peaks is now added to Table 2 for completeness, but please note that one reason for abandoning the peak approach is that it forces the data into a bimodal distribution. We moved away from this approach when we found that there were some individuals with L-R difference waves that hovered around zero: depending on whether the peak difference was greater for L or R, the peak would be taken at that point, and for those with very slight differences between the sides, this could seem fairly arbitrary. See Woodhead, Rutherford and Bishop, 2018, for discussion of this point.

1.4 *Heritability of L and R*
CF notes: "In general, measuring asymmetry necessarily involves calculating some kind of difference score, which means that error variance in both left and right measures can affect the asymmetry measure. This may partly explain the low heritabilities of asymmetry indexes, and the authors could acknowledge this. In brain structural analysis (of e.g. grey matter volumes, surface areas, thicknesses), we typically see higher heritabilities for L and R separately than for the asymmetry index (L-R)/(L+R). Might it be informative to calculate heritabilities for L and R, in the context of functional transcranial Doppler sonography? If L and R are not themselves heritable, there may be a problem with the approach."

This is an interesting suggestion, which we have adopted. The correlation between the L and R flow measures *within* individuals is very high (close to .9). As now described in the

paper, the raw L and R mean flow measures showed significant twin-twin correlations, but there was no effect of zygosity, suggesting that the similarity between twins was partly due to shared environment, rather than genetic influences. This is an intriguing result, but it is hard to interpret. This is not an age effect: age was not correlated with the flow measures, and residualising scores on age and sex did not make any difference. There have in the past been environmental explanations proposed for handedness, including shared in utero environment, which we have now alluded to. CF implies that a lack of heritability of the individual L and R flow measures may indicate a problem with the approach, but another possibility is that this trait is simply not under genetic control. We are not aware of any previous literature on heritability of cerebral blood flow.

1.5. *Support for assumption of strong heritability*
Abstract: The authors state that 'it is widely assumed that individual differences in language lateralisation have a strong genetic basis'. Are there references to support this? I am not sure that many researchers working on the genetics of laterality have this impression (this one does not).

This comment has been removed, as it is based solely on informal impressions – mainly surprised reactions when these results are discussed.

1.6 *Rewording*
The Discussion can be revised in parts, to reflect the recent literature indicated above.
The authors wrote that 'Our data are compatible with a more radical model in which language laterality is the consequence of a general population left-sided brain bias for language which does not show any individual variation.' For clarity, it would help to make explicit that genetic variation is meant here. I agree that the author's data are compatible with this model, but the limited sample size means that the data are also compatible with a heritability up to whatever level is encompassed by the confidence interval around the best estimate.

We have reworded as requested

2. **Guy Vingerhoets**

2.1 *Methods clarification*
The only detail that puzzles me is the period of interest of the fTCD procedure. In one paragraph a start and stop cue indicating a 10s period is mentioned during which the child describes the 12s cartoon. In another paragraph the period of interest is defined as 4 to 14s after the cue to speak, but this method was abandoned for the 'whole period of interest'. Would that be the 10s period between both cues then?

This has now been written more clearly to explain how the peak method works, and how it differs from the mean method.

2.2 *Validity of fTCD laterality*
fTCD- derived LI's offer only crude estimations of asymmetry as they will pick up velocity changes due to co-activation of many other mental components that take place in the MCA territory and that may not be related to language. The child has just seen a 12s cartoon and now must describe it. Attention, memory, visual imagery, movement(?), receptive and productive language

areas all come into play and most of these functions will be associated with activation of the lateral cortex. Besides not being very region-specific, fTCD does not allow for the use of control-tasks that can correct for task-unspecific activation.

This is a point that is often raised by those who work with fMRI, and the Oxford group has given it much consideration – for instance, using tasks based on Mazoyer et al's sentence generation and control tasks, Woodhead et al (2018) computed laterality indices based on the difference in lateralised activation for sentence vs list generation. We showed this made little overall difference, because list generation was not lateralised. The central point is that while, of course, there are numerous nonlinguistic factors involved in performing the animation description task, any nonlateralised activity is subtracted out by our analytic procedure, when we take the difference score. One can see in the waveforms associated with different tasks periods when blood flow increases and decreases – e.g. the blood flow increases just before starting to talk, then falls away. But so long as these are symmetrical effects they do not affect the laterality index, which is based on the difference waveform.

2.3 *Relative size of effect for fTCD and fMRI*
We are currently doing some direct comparisons of laterality indices from fMRI and fTCD with adults, and are finding good levels of agreement of laterality indices for equivalent tasks, but we would caution against attempting direct comparisons of magnitude of effects, because of the very different ways in which brain activation is measured. With fMRI a general linear model is used to measure how the brain activation relates to experimental design variables, and then t-statistics of the voxels within the left and right hemisphere ROIs are computed and thresholded, and either the count (extent) or the sum (magnitude) of supra-threshold t-values is calculated in each hemisphere. With fTCD, the blood flow signal is epoched, normalised and baseline corrected, so that both left and right sensors have an average signal of zero in a resting time period at the start of each trial. The intensities of the left and right signals are then directly compared within a period of interest when the participant is performing the task.

3. Chris McManus

3.1 *Plausibility of laterality based on prior literature*
CM *"The use of "assumed" suggests a lack of substance to any genetic basis for lateralisation, but that is surely not the case."*

All reviewers challenged this statement, so it has been removed. We would stress that we are not arguing that all laterality is non-heritable, and we agree that as regards handedness, there is enough evidence to be confident in heritability of around .25, and growing evidence of genetic influences on structural brain asymmetry in some regions (as now reviewed more fully).
However, this is not the same as functional language laterality, where there is a paucity of evidence. As far as we are aware, the sum total of prior evidence is contained in 3 studies:
a) Ocklenburg, Bryden, and Somers. CM mentions the Ocklenburg study that we cited. This found zero heritability for the standard dichotic listening task, but heritability of .28 to .36 for a task condition with directed attention. The authors of that study concluded that their results: "implicate a major contribution of non-genetic influences to individual language

lateralization."

b) Bryden (now mentioned in our revision), used two measures in both parents and two siblings from 49 families. Although he found one statistically significant association between mother and child, he drew attention to inconsistent findings – not only were the correlations between siblings negative, but one of the highest correlations was between mother and father. As he wryly noted, "This correlation would suggest non-random mating for laterality, a characteristic that one would hardly expect to be of significance in selecting one's spouse. " (p. 206). He concluded: " ..the present study has failed to find any particularly compelling evidence for a genetic basis for speech lateralization. While the problems associated with the use of an indirect measure of only moderate reliability may have doomed this study from the start, it does suggest that one should at least consider seriously the hypothesis that speech lateralization is primarily determined by environmental factors" (p.209). The split half reliability of the measures was .61 and .66.

c) The strongest evidence for heritability of language laterality on a functional brain measure is from the Somers et al study mentioned above using fTCD with a multi-generational pedigree sample from an isolated community. The heritability of atypical language lateralisation (coded as a binary variable) was 0.31. This sample was not at all representative of the general population: it was deliberately selected to have relatively low genetic heterogeneity and to include families with several left-handed members. This might affect generalisability of findings, but the most serious limitation of the sample was that the selection method was biased toward phenotypic similarity of those who were in the sample; i.e. insofar as handedness is related to language laterality, then selecting only families with at least two left-handers per generation could artificially inflate within-family similarity for laterality. On the other hand, a potential advantage of the study was the use of a pedigree-based method of analysis, which gives higher power than a method reliant just on twin pairs.

Taken together, the evidence is far from compelling, with moderate reliabilities, modest samples sizes, and heritability estimates compatible both with zero and with modest values such as .3 or so.

3.2 *Other studies*

CM notes "Badzakowa-Trakjov et al (2010; not referenced in the current paper)[6] included data on 34 MZ pairs and 11 DZ pairs (with data being available from the authors); they did not however calculate heritabilities."

This was not cited precisely because it is hard to derive any information about heritability from the study, because the sample was not only small, but also was selected to be biased to include discordant pairs (half the pairs studied had discordant handedness). Following the prompt from CM, we downloaded the data (available from PLOS One). The correlation between 34 MZ twin pairs on Word Generation was -.11 and for 11 DZ pairs it was .32. This is hardly encouraging for a genetic theory of language lateralisation, but, for the reasons noted above, it would be rash to draw much of a conclusion from this.

The Human Connectome Project with its 132 MZ and 101 DZ pairs, for which data are available for download, has information on handedness and probably has measures related to language dominance.

As CF points out above, there are some analyses of structural asymmetries. As far as we know there are no data on functional asymmetry

Biobank: again, there are analyses of handedness, but, although there are MRI data on a subset of individuals, the functional MRI in Biobank did not include a language task. We believe there are moves afoot to derive a laterality measure from resting state fMRI, but it is unclear how this would relate to laterality on something like a word or sentence generation task.

3.3 *Need for confidence intervals*
Thanks for the computations of confidence intervals around heritability estimates – these were helpful and in line with our bootstrapped computations, which are now given.

3.4 *Multivariate models*
The second author in fact had suggested we include these, but the first author thought this would be overkill, given that, as the reviewer points out, it would involve testing whether zero heritability in one trait is shared with low heritability in another. These results are given here for completeness, but not incorporated in the main paper. Path estimates are shown for the AE model, as all C estimates were zero.

Unsquared standardized estimates of paths from multivariate model.
Asterisks denote paths that are statistically significant at .05 level.

	Term	Measure	Estimate.EQL	SE.EQL	Estimate.QL	SE.QL	Estimate.L	SE.L
1	A	EHI	0.458*	0.098				
2	A	QHP	0.243	0.124	0.343*	0.093		
3	A	LI	0.123	0.113	-0.152	0.123	0.000	0.213
4	E	EHI	0.889*	0.051				
5	E	QHP	0.401*	0.068	0.814*	0.043		
6	E	LI	-0.134	0.066	0.039	0.064	0.971	0.039

N.B. EQL loadings for paths to all three measures: QL paths independent of EHI; L paths specific to language LI.

Information Criteria:

	df Penalty	Parameters Penalty	Sample-Size Adjusted
AIC:	5562.0751	8922.075	8923.893
BIC:	-489.8496	8976.597	8929.033

3.5 *Language dominance and handedness*
Relevant data on this are now added.
We note also that the lack of association is inconsistent with previous literature which tends to find a significant association between handedness and laterality measures.

3.6 *Numbers with atypical dominance*
CM notes "A further query about validity of the data is raised by the relatively high proportion with atypical lateralisation."

With fTCD, the proportions with atypical lateralisation are entirely dependent on the language measure used, as shown by Woodhead et al (2018). This leads us to conclude that language lateralisation is not a fixed binary property of the brain, but varies in degree according to task demands. Some discussion of this point is now added.

3.7 *Use of the Deppe original method*
Whether means were different was unclear, and a scattergram would have been useful. An advantage is claimed to be that the "bimodality of the laterality index distribution is not seen when the means-based method is used", although that is not self-evidently good when there are strong a priori expectations that laterality indices may well be bimodal.

A scattergram has now been added in the Appendix. We disagree regarding the 'strong a priori expectations': there are some theories that treat handedness as bimodal, but others that do not, and, for language laterality, there is no particular reason to assume bimodality. Data on LIs obtained using the 'peak' method are now added.

CM states "It would be reassuring to know that precisely the same results were obtained when the data were processed with DopOSCCI."
Please see Woodhead et al (2018) where we note: " the R script had been developed in our group to fulfil the need for a reproducible and efficient method for processing large numbers of datasets, without using commercial (Matlab) software that required a licence (see Wilson & Bishop, 2018). As with DopOSCCI the analytic pipeline closely followed procedures developed by Deppe et al. (2004), with one additional option: the possibility of identifying brief periods of signal spiking or dropout and interpolating over these, to avoid rejecting trials. Wilson and Bishop (2018) compared results from DopOSCCI and the R script and found only small differences in the LIs computed by the two methods". Given that we have spent a great deal of time developing scripts that allow us to process the data efficiently in R, and checking that it gives results highly consistent with the prior DopOSCCI approach, we do not think it reasonable to be asked to revert to the prior method. Everyone makes errors of course, and we cannot rule out that there may be a bug somewhere, but we feel it is unreasonable to expect us to keep analysing our data using different methods. We do agree that there is an element of arbitrariness in the data processing pipeline for fTCD – results will vary depending on the sequence of operations, treatment of outliers, the period of interest, and whether peaks or means are used, but in our experience these differences have only minor impact on the final laterality index. We are interested to evaluate other approaches, but the method presented here was judged to be optimal and is the best we can do at present. Our scripts are available, together with the raw data, so others are welcome to try different approaches.

3.8 *Sample*
Overall it seems clear that this population is probably far from representative of the general population.
While this is true, in our prior paper, we showed that there was no difference in cerebral lateralisation for children with and without language problems, and similar results to those obtained by Groen et al (2012) with singleborn children. Prior studies have found no evidence for any genetic link between language disorders and lateralisation (Bishop, 2001;

2005).

Overall
The criticisms offered by the reviewers are generally fair, and we are happy to moderate the way in which the current results are described to clarify the limited conclusions that can be drawn. For the reasons stated at the outset, it is important that null results are published so that future meta-analyses are not biased in favour of positive findings. We hope that there will be more studies on this topic so that in future it might be possible to incorporate these data in a meta-analysis, to give a more precise estimate of heritability of language laterality.

References

Bishop, D. V. M. (1990). How to increase your chances of obtaining a significant association between handedness and disorder. *Journal of Clinical and Experimental Ñeuropsychology, 12*, 786-790.
Bishop, D. V. M. (2001). Individual differences in handedness and specific language impairment: evidence against a genetic link. *Behavior Genetics, 31*, 339-351.
Bishop, D. V. M. (2005). Handedness and specific language impairment: a study of 6-year-old twins. *Developmental Psychobiology, 46*, 362-369.
Bishop, D. V. M. (2020). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research. The 45th Sir Frederic Bartlett Lecture *Quarterly Journal of Experimental Psychology, 73*(1), 1-19. doi:10.1177/1747021819886519
Bryden, M. P. (1975). Speech lateralization in families: A preliminary study using dichotic listening. *Brain and Language, 2*, 201-211.
Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*(3), 891-904. doi:10.1007/s11192-011-0494-7
Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*(1), 1-20.
Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *eLife, 5*, e21451. doi:10.7554/eLife21451.001
Sterling, T. D., Rosenbaum, W. D., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician, 49*(1), 108-112.

Competing Interests: None

Reviewer Report 07 November 2019

https://doi.org/10.21956/wellcomeopenres.16993.r36877

? **Clyde Francks** 🆔

[1] Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Congratulations to the authors for another important study in this field.

I broadly agree with the authors that measures of brain and behavioural asymmetry tend to have low heritabilities. However, it is not clear to me that the heritability of language laterality in this study is negligible or zero, as asserted by the authors. The authors use appropriate caution in many parts of the paper, but other parts seem too strong, such as the current title of the paper, or the abstract, which give the impression that this study found negligible heritability.

The authors report a heritability estimate of zero for their measure of language laterality, based on twin analysis. However, the confidence interval for this estimate is not given. This is not a large study, and the power calculations suggest that the confidence interval must have a substantial range. The authors can only be confident that the heritability falls within the confidence interval, not that it is zero or negligible. How high might it go, for example are the data compatible with 10-20% heritability? This would not be negligible. Related to this, the authors mention that a model with no genetic term gave as good a fit as one including it. Please include the goodness-of-fit statistics.

I would like to make the authors aware of some relevant literature, some of which is from my own group (I may have missed opportunities to make them aware of these studies in recent years). There are two family- and/or twin-based analyses of the heritability of brain structural laterality, that were based on larger sample sizes than those cited by the authors.
Kong X *et al*[ref-1]
Guadalupe *et al*[2]

These studies show significant heritabilites up to 27% for various aspects of brain structural asymmetry, and indicate that gene mapping for these asymmetries may be fruitful.
As regards handedness, there are two recent genome-wide association scan (GWAS) studies of handedness based on the UK biobank data of more than 330,000 subjects.
kovel *et al*[3]
Wiberg *et al*[4]

These studies are over two orders of magnitude larger than the GWAS cited by the authors, and have identified significant genetic associations with left-handedness, offering some glimpses into the biology of the trait. Note that the SNP-based heritability of left-handedness was only around 2% in the UK Biobank data, but still significant due to the large sample size. This shows that a heritability even as low as 2% can be a basis for successful gene mapping, to deliver insights into trait biology, which is a point that the authors could acknowledge to give a balanced impression to the field. If the author's data are compatible with even 5-10% heritability for language laterality (in terms of the confidence interval around their heritability estimate), then future gene mapping is certainly possible, that might help to reveal genetic-developmental mechanisms of laterality formation. My concern is that a title and abstract saying that heritability was negligible does not really capture this possibility helpfully for the field, nor the uncertainty within this study itself.

Methods: I have no experience with data from functional transcranial Doppler sonography. I assume the authors have analyzed this in a correct and state-of-the-art way. However, Somers et al. (2015)[5] found a heritability of 31% for atypical language lateralisation (coded as a binary variable), using functional transcranial Doppler ultrasound in a multi-generational pedigree sample. If the same data processing pipeline from that study would be applied in the current study, might the results no longer be discrepant between the two studies?

In general, measuring asymmetry necessarily involves calculating some kind of difference score, which means that error variance in both left and right measures can affect the asymmetry measure. This may partly explain the low heritabilities of asymmetry indexes, and the authors could acknowledge this. In brain structural analysis (of e.g. grey matter volumes, surface areas, thicknesses), we typically see higher heritabilities for L and R separately than for the asymmetry index (L-R)/(L+R). Might it be informative to calculate heritabilities for L and R, in the context of functional transcranial Doppler sonography? If L and R are not themselves heritable, there may be a problem with the approach.

Abstract: The authors state that 'it is widely assumed that individual differences in language lateralisation have a strong genetic basis'. Are there references to support this? I am not sure that many researchers working on the genetics of laterality have this impression (this one does not).

The Discussion can be revised in parts, to reflect the recent literature indicated above.
The authors wrote that 'Our data are compatible with a more radical model in which language laterality is the consequence of a general population left-sided brain bias for language which does not show any individual variation.' For clarity, it would help to make explicit that genetic variation is meant here. I agree that the author's data are compatible with this model, but the limited sample size means that the data are also compatible with a heritability up to whatever level is encompassed by the confidence interval around the best estimate.

References
1. Kong X, Mathias S, Guadalupe T, ENIGMA Laterality Working Group, et al.: Mapping cortical brain asymmetry in 17,141 healthy individuals worldwide via the ENIGMA Consortium. *Proceedings of the National Academy of Sciences*. 2018; **115** (22): E5154-E5163 Publisher Full Text
2. Guadalupe T, Mathias SR, vanErp TGM, Whelan CD, et al.: Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex.*Brain Imaging Behav*. 2017; **11** (5): 1497-1514 PubMed Abstract | Publisher Full Text
3. de Kovel CGF, Francks C: The molecular genetics of hand preference revisited.*Sci Rep*. 2019; **9** (1): 5986 PubMed Abstract | Publisher Full Text
4. Wiberg A, Ng M, Al Omran Y, Alfaro-Almagro F, et al.: Handedness, language areas and neuropsychiatric diseases: insights from brain imaging and genetics.*Brain*. 2019; **142** (10): 2938-2947 PubMed Abstract | Publisher Full Text
5. Somers J, Wilson LA, Kilday JP, Horvilleur E, et al.: A common polymorphism in the 5' UTR of ERCC5 creates an upstream ORF that confers resistance to platinum-based chemotherapy.*Genes Dev*. 2015; **29** (18): 1891-6 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Human neurogenetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 22 Feb 2020

Dorothy Bishop, University of Oxford, Oxford, UK

Response to reviewers

Our thanks to all three reviewers for their thoughtful comments on this article. There was strong concordance in their critiques, and so we will respond to the general points that were made before going on to cover specific points by individual reviewers.

General comments
This paper, and the criticisms, get to the nub of issues that have become prominent in discussions of the so-called 'reproducibility crisis'. The fundamental question is how do we determine whether a null result, like the one reported here, is a type II error. We agree with the reviewers that it is not enough just to present a null result and conclude that the true effect is zero. A null result could arise because the study is underpowered to detect a true but small effect, or because the methods lack reliability or validity. In the Discussion we considered these possibilities and concluded that we could not draw a strong conclusion about our null finding, though we could specify a likely upper bound for heritability. However, the use of the word 'negligible' in the title attracted criticism from all reviewers, and we accept that this is too value-laden and have deleted it.

Another point is that when evaluating research findings we should never rely on evidence from a single study. Again, we noted in the Discussion how Bishop's earlier findings of

insignificant heritability of handedness had been overturned by subsequent, larger studies, which emphasises the need for caution. The suggestions by reviewers to try other analyses to see if the results look different are consistent with a Bayesian approach that demands especially strong evidence to overturn a strong prior belief.

There is, however, a need for caution here. Prior expectations come from at least two sources. First, there is the issue of needing a plausible mechanism, and we agree that a genetic basis for individual differences in a neurobiological phenotype is plausible. The main source of priors, however, will be previous literature. All three reviewers have cited additional sources for genetic influences on laterality. But there is a question of just how much confidence one should place in prior literature, given that there are many inconsistent findings, plus three systematic biases that distort results: publication bias, p-hacking and citation bias. It is always difficult to discuss these biases, because it looks as if one is singling out other researchers for criticism, and impugning their integrity. Nevertheless, their prevalence is not in doubt (Greenwald, 1975; Sterling, Rosenbaum, & Weinkam, 1995; Fanelli, 2012), and there is circumstantial evidence that they have influenced the field of laterality. Consider, for instance two twin studies, one on handedness by Davis and Annett (1994), and one on structural brain asymmetry by Geschwind et al (2002). Both studies are interpreted by their authors as supporting genetic models of cerebral lateralisation, but neither reported heritability estimates for laterality (or the zygosity-based correlations that underpin these), despite having the data available for doing this. One conclusion is that the heritability estimates were not convincing, and so went unreported.

The role of publication bias in distorting beliefs about solidness of findings was nicely documented in a simulation by Nissen et al (2016). Where this effect is compounded by p-hacking, then we can end up with up with solid beliefs based purely on the fact that we are sampling biased evidence. On top of that we all have a tendency to confirmation bias, which means that even when null findings are published, we tend to disregard them, which further biases the evidence (Bishop, 2020).

The field of laterality research is at particular risk of bias because there are so many different ways of conceptualising the phenotype. This point was made with regard to handedness by Bishop (1990), who noted that if you do not prespecify in advance how you plan to convert a handedness scale into groups, then you raise the chance of finding a 'significant' result to well above 5%. This is equally true for other types of laterality, where there is no agreement about accepted measurement practices. And where laterality measures are part of a larger battery of tests, then it is likely that results on heritability will usually be published only if significant. The reviewers note the need for cautious interpretation of our results, and we agree, but in the absence of pre-registered studies, we also need to adopt a cautious stance to the prior literature, as there is a substantial risk of type I error. Large sample sizes can save us from type II errors but they are no defence against type I errors if p-hacking is possible. Findings that have been replicated using the same methods can be given much more weight than one-off studies.

Interpretation of evidence in this field is further complicated by the fact that there are many different forms of laterality – as well as handedness we have both structural and functional brain laterality. Once we move from handedness, little is known about the reliability of

these different measures of phenotype, but it is clear that they are not interchangeable, and the relationships between them are not clearly understood.

These points are amplified below when dealing with specific points raised by reviewers; the final paragraph of the Discussion has been amended to make it clear that we are not claiming that we have definitely proven a null result, but rather that very low, or even absent heritability of functional language lateralisation should at least be treated as a realistic contender, rather than dismissed as implausible.

Responses to specific comments by reviewers

1. Clyde Francks

1.1. *Request for confidence interval for heritability estimate and goodness of fit statistics*

These have now been provided in Table 2.

1.2. *Additional papers*
Thanks for drawing our attention to these papers that include genetic analysis of structural brain asymmetries. We should not have overlooked these papers which are indeed familiar to the first author, but in mitigation neither paper mention the words genetic or heritable in the title, and only Guadalupe et al mentioned 'heritability' in the keywords, so it is easy to miss these when conducting a systematic search for relevant papers. They are now included in the account of structural asymmetries.

The recent GWA studies of handedness are now mentioned in the Discussion, and we note that while these show very low SNP-based heritability, this does not preclude successful gene mapping – though the sample sizes required make it unlikely this will be feasible using measures of language function based on fMRI or fTCD. Unfortunately, UK Biobank did not include language function activation methods in fMRI.

1.3 *Alternative pipeline for Doppler method*
CF notes: "*Somers et al. (2015)[5] found a heritability of 31% for atypical language lateralisation (coded as a binary variable), using functional transcranial Doppler ultrasound in a multi-generational pedigree sample. If the same data processing pipeline from that study would be applied in the current study, might the results no longer be discrepant between the two studies?*" Somers used the traditional peak-based approach to analysing the fTCD data. Although for some analyses they used a continuous laterality index, heritability was reported only for a binary category of typical vs atypical. Correspondence with Dr Somers confirmed that this method was not suited to continuous data.

Results from analysis of peaks is now added to Table 2 for completeness, but please note that one reason for abandoning the peak approach is that it forces the data into a bimodal distribution. We moved away from this approach when we found that there were some individuals with L-R difference waves that hovered around zero: depending on whether the peak difference was greater for L or R, the peak would be taken at that point, and for those with very slight differences between the sides, this could seem fairly arbitrary. See

Woodhead, Rutherford and Bishop, 2018, for discussion of this point.

1.4 *Heritability of L and R*
CF notes: "In general, measuring asymmetry necessarily involves calculating some kind of difference score, which means that error variance in both left and right measures can affect the asymmetry measure. This may partly explain the low heritabilities of asymmetry indexes, and the authors could acknowledge this. In brain structural analysis (of e.g. grey matter volumes, surface areas, thicknesses), we typically see higher heritabilities for L and R separately than for the asymmetry index (L-R)/(L+R). Might it be informative to calculate heritabilities for L and R, in the context of functional transcranial Doppler sonography? If L and R are not themselves heritable, there may be a problem with the approach."

This is an interesting suggestion, which we have adopted. The correlation between the L and R flow measures *within* individuals is very high (close to .9). As now described in the paper, the raw L and R mean flow measures showed significant twin-twin correlations, but there was no effect of zygosity, suggesting that the similarity between twins was partly due to shared environment, rather than genetic influences. This is an intriguing result, but it is hard to interpret. This is not an age effect: age was not correlated with the flow measures, and residualising scores on age and sex did not make any difference. There have in the past been environmental explanations proposed for handedness, including shared in utero environment, which we have now alluded to. CF implies that a lack of heritability of the individual L and R flow measures may indicate a problem with the approach, but another possibility is that this trait is simply not under genetic control. We are not aware of any previous literature on heritability of cerebral blood flow.

1.5. *Support for assumption of strong heritability*
Abstract: The authors state that 'it is widely assumed that individual differences in language lateralisation have a strong genetic basis'. Are there references to support this? I am not sure that many researchers working on the genetics of laterality have this impression (this one does not).

This comment has been removed, as it is based solely on informal impressions – mainly surprised reactions when these results are discussed.

1.6 *Rewording*
The Discussion can be revised in parts, to reflect the recent literature indicated above.
The authors wrote that 'Our data are compatible with a more radical model in which language laterality is the consequence of a general population left-sided brain bias for language which does not show any individual variation.' For clarity, it would help to make explicit that genetic variation is meant here. I agree that the author's data are compatible with this model, but the limited sample size means that the data are also compatible with a heritability up to whatever level is encompassed by the confidence interval around the best estimate.

We have reworded as requested

2. **Guy Vingerhoets**

2.1 *Methods clarification*

The only detail that puzzles me is the period of interest of the fTCD procedure. In one paragraph a start and stop cue indicating a 10s period is mentioned during which the child describes the 12s cartoon. In another paragraph the period of interest is defined as 4 to 14s after the cue to speak, but this method was abandoned for the 'whole period of interest'. Would that be the 10s period between both cues then?

This has now been written more clearly to explain how the peak method works, and how it differs from the mean method.

2.2 *Validity of fTCD laterality*
fTCD- derived LI's offer only crude estimations of asymmetry as they will pick up velocity changes due to co-activation of many other mental components that take place in the MCA territory and that may not be related to language. The child has just seen a 12s cartoon and now must describe it. Attention, memory, visual imagery, movement(?), receptive and productive language areas all come into play and most of these functions will be associated with activation of the lateral cortex. Besides not being very region-specific, fTCD does not allow for the use of control-tasks that can correct for task-unspecific activation.

This is a point that is often raised by those who work with fMRI, and the Oxford group has given it much consideration – for instance, using tasks based on Mazoyer et al's sentence generation and control tasks, Woodhead et al (2018) computed laterality indices based on the difference in lateralised activation for sentence vs list generation. We showed this made little overall difference, because list generation was not lateralised. The central point is that while, of course, there are numerous nonlinguistic factors involved in performing the animation description task, any nonlateralised activity is subtracted out by our analytic procedure, when we take the difference score. One can see in the waveforms associated with different tasks periods when blood flow increases and decreases – e.g. the blood flow increases just before starting to talk, then falls away. But so long as these are symmetrical effects they do not affect the laterality index, which is based on the difference waveform.

2.3 *Relative size of effect for fTCD and fMRI*
We are currently doing some direct comparisons of laterality indices from fMRI and fTCD with adults, and are finding good levels of agreement of laterality indices for equivalent tasks, but we would caution against attempting direct comparisons of magnitude of effects, because of the very different ways in which brain activation is measured. With fMRI a general linear model is used to measure how the brain activation relates to experimental design variables, and then t-statistics of the voxels within the left and right hemisphere ROIs are computed and thresholded, and either the count (extent) or the sum (magnitude) of supra-threshold t-values is calculated in each hemisphere. With fTCD, the blood flow signal is epoched, normalised and baseline corrected, so that both left and right sensors have an average signal of zero in a resting time period at the start of each trial. The intensities of the left and right signals are then directly compared within a period of interest when the participant is performing the task.

3. Chris McManus

3.1 *Plausibility of laterality based on prior literature*

CM "The use of "assumed" suggests a lack of substance to any genetic basis for lateralisation, but that is surely not the case."

All reviewers challenged this statement, so it has been removed. We would stress that we are not arguing that all laterality is non-heritable, and we agree that as regards handedness, there is enough evidence to be confident in heritability of around .25, and growing evidence of genetic influences on structural brain asymmetry in some regions (as now reviewed more fully).
However, this is not the same as functional language laterality, where there is a paucity of evidence. As far as we are aware, the sum total of prior evidence is contained in 3 studies:
a) Ocklenburg, Bryden, and Somers. CM mentions the Ocklenburg study that we cited. This found zero heritability for the standard dichotic listening task, but heritability of .28 to .36 for a task condition with directed attention. The authors of that study concluded that their results: "implicate a major contribution of non-genetic influences to individual language lateralization."
b) Bryden (now mentioned in our revision), used two measures in both parents and two siblings from 49 families. Although he found one statistically significant association between mother and child, he drew attention to inconsistent findings – not only were the correlations between siblings negative, but one of the highest correlations was between mother and father. As he wryly noted, "This correlation would suggest non-random mating for laterality, a characteristic that one would hardly expect to be of significance in selecting one's spouse. " (p. 206). He concluded: " ..the present study has failed to find any particularly compelling evidence for a genetic basis for speech lateralization. While the problems associated with the use of an indirect measure of only moderate reliability may have doomed this study from the start, it does suggest that one should at least consider seriously the hypothesis that speech lateralization is primarily determined by environmental factors" (p.209). The split half reliability of the measures was .61 and .66.
c) The strongest evidence for heritability of language laterality on a functional brain measure is from the Somers et al study mentioned above using fTCD with a multi-generational pedigree sample from an isolated community. The heritability of atypical language lateralisation (coded as a binary variable) was 0.31. This sample was not at all representative of the general population: it was deliberately selected to have relatively low genetic heterogeneity and to include families with several left-handed members. This might affect generalisability of findings, but the most serious limitation of the sample was that the selection method was biased toward phenotypic similarity of those who were in the sample; i.e. insofar as handedness is related to language laterality, then selecting only families with at least two left-handers per generation could artificially inflate within-family similarity for laterality. On the other hand, a potential advantage of the study was the use of a pedigree-based method of analysis, which gives higher power than a method reliant just on twin pairs.

Taken together, the evidence is far from compelling, with moderate reliabilities, modest samples sizes, and heritability estimates compatible both with zero and with modest values such as .3 or so.

3.2 Other studies
CM notes "Badzakowa-Trakjov et al (2010; not referenced in the current paper)[6] included data on

34 MZ pairs and 11 DZ pairs (with data being available from the authors); they did not however calculate heritabilities."

This was not cited precisely because it is hard to derive any information about heritability from the study, because the sample was not only small, but also was selected to be biased to include discordant pairs (half the pairs studied had discordant handedness). Following the prompt from CM, we downloaded the data (available from PLOS One). The correlation between 34 MZ twin pairs on Word Generation was -.11 and for 11 DZ pairs it was .32. This is hardly encouraging for a genetic theory of language lateralisation, but, for the reasons noted above, it would be rash to draw much of a conclusion from this.

The Human Connectome Project with its 132 MZ and 101 DZ pairs, for which data are available for download, has information on handedness and probably has measures related to language dominance.

As CF points out above, there are some analyses of structural asymmetries. As far as we know there are no data on functional asymmetry

Biobank: again, there are analyses of handedness, but, although there are MRI data on a subset of individuals, the functional MRI in Biobank did not include a language task. We believe there are moves afoot to derive a laterality measure from resting state fMRI, but it is unclear how this would relate to laterality on something like a word or sentence generation task.

3.3 *Need for confidence intervals*
Thanks for the computations of confidence intervals around heritability estimates – these were helpful and in line with our bootstrapped computations, which are now given.

3.4 *Multivariate models*
The second author in fact had suggested we include these, but the first author thought this would be overkill, given that, as the reviewer points out, it would involve testing whether zero heritability in one trait is shared with low heritability in another. These results are given here for completeness, but not incorporated in the main paper. Path estimates are shown for the AE model, as all C estimates were zero.

Unsquared standardized estimates of paths from multivariate model.
Asterisks denote paths that are statistically significant at .05 level.

	Term	Measure	Estimate.EQL	SE.EQL	Estimate.QL	SE.QL	Estimate.L	SE.L
1	A	EHI	0.458*	0.098				
2	A	QHP	0.243	0.124	0.343*	0.093		
3	A	LI	0.123	0.113	-0.152	0.123	0.000	0.213
4	E	EHI	0.889*	0.051				
5	E	QHP	0.401*	0.068	0.814*	0.043		
6	E	LI	-0.134	0.066	0.039	0.064	0.971	0.039

N.B. EQL loadings for paths to all three measures: QL paths independent of EHI; L paths specific to language LI.

Information Criteria:
```
     | df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
AIC:   5562.0751         8922.075              8923.893
BIC:   -489.8496         8976.597              8929.033
```

3.5 *Language dominance and handedness*
Relevant data on this are now added.
We note also that the lack of association is inconsistent with previous literature which tends to find a significant association between handedness and laterality measures.

3.6 *Numbers with atypical dominance*
CM notes "A further query about validity of the data is raised by the relatively high proportion with atypical lateralisation."

With fTCD, the proportions with atypical lateralisation are entirely dependent on the language measure used, as shown by Woodhead et al (2018). This leads us to conclude that language lateralisation is not a fixed binary property of the brain, but varies in degree according to task demands. Some discussion of this point is now added.

3.7 *Use of the Deppe original method*
Whether means were different was unclear, and a scattergram would have been useful. An advantage is claimed to be that the "bimodality of the laterality index distribution is not seen when the means-based method is used", although that is not self-evidently good when there are strong a priori expectations that laterality indices may well be bimodal.

A scattergram has now been added in the Appendix. We disagree regarding the 'strong a priori expectations': there are some theories that treat handedness as bimodal, but others that do not, and, for language laterality, there is no particular reason to assume bimodality. Data on LIs obtained using the 'peak' method are now added.

CM states "It would be reassuring to know that precisely the same results were obtained when the data were processed with DopOSCCI."
Please see Woodhead et al (2018) where we note: " the R script had been developed in our group to fulfil the need for a reproducible and efficient method for processing large numbers of datasets, without using commercial (Matlab) software that required a licence (see Wilson & Bishop, 2018). As with DopOSCCI the analytic pipeline closely followed procedures developed by Deppe et al. (2004), with one additional option: the possibility of identifying brief periods of signal spiking or dropout and interpolating over these, to avoid rejecting trials. Wilson and Bishop (2018) compared results from DopOSCCI and the R script and found only small differences in the LIs computed by the two methods". Given that we have spent a great deal of time developing scripts that allow us to process the data efficiently in R, and checking that it gives results highly consistent with the prior DopOSCCI approach, we do not think it reasonable to be asked to revert to the prior method. Everyone makes errors of course, and we cannot rule out that there may be a bug somewhere, but we feel it is unreasonable to expect us to keep analysing our data using different methods. We do agree that there is an element of arbitrariness in the data

processing pipeline for fTCD – results will vary depending on the sequence of operations, treatment of outliers, the period of interest, and whether peaks or means are used, but in our experience these differences have only minor impact on the final laterality index. We are interested to evaluate other approaches, but the method presented here was judged to be optimal and is the best we can do at present. Our scripts are available, together with the raw data, so others are welcome to try different approaches.

3.8 *Sample*
Overall it seems clear that this population is probably far from representative of the general population.
While this is true, in our prior paper, we showed that there was no difference in cerebral lateralisation for children with and without language problems, and similar results to those obtained by Groen et al (2012) with singleborn children. Prior studies have found no evidence for any genetic link between language disorders and lateralisation (Bishop, 2001; 2005).

Overall
The criticisms offered by the reviewers are generally fair, and we are happy to moderate the way in which the current results are described to clarify the limited conclusions that can be drawn. For the reasons stated at the outset, it is important that null results are published so that future meta-analyses are not biased in favour of positive findings. We hope that there will be more studies on this topic so that in future it might be possible to incorporate these data in a meta-analysis, to give a more precise estimate of heritability of language laterality.

References

Bishop, D. V. M. (1990). How to increase your chances of obtaining a significant association between handedness and disorder. *Journal of Clinical and Experimental Ñeuropsychology, 12*, 786-790.
Bishop, D. V. M. (2001). Individual differences in handedness and specific language impairment: evidence against a genetic link. *Behavior Genetics, 31*, 339-351.
Bishop, D. V. M. (2005). Handedness and specific language impairment: a study of 6-year-old twins. *Developmental Psychobiology, 46*, 362-369.
Bishop, D. V. M. (2020). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research. The 45th Sir Frederic Bartlett Lecture *Quarterly Journal of Experimental Psychology, 73*(1), 1-19. doi:10.1177/1747021819886519
Bryden, M. P. (1975). Speech lateralization in families: A preliminary study using dichotic listening. *Brain and Language, 2*, 201-211.
Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*(3), 891-904. doi:10.1007/s11192-011-0494-7
Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*(1), 1-20.
Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *eLife, 5*, e21451. doi:10.7554/eLife21451.001
Sterling, T. D., Rosenbaum, W. D., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician, 49*(1), 108-112.

Competing Interests: None