Negligible heritability of language laterality assessed by functional transcranial Doppler ultrasound: a twin study [version 1; peer review: awaiting peer review]

Dorothy V.M. Bishop 1, Timothy C. Bates 2

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

Abstract
Background: It is widely assumed that individual differences in language lateralisation have a strong genetic basis, yet prior studies show low heritability (around 0.25) for the related trait of handedness, and two twin studies of structural brain asymmetry obtained similarly low estimates. This report describes heritability estimates from a twin study of language laterality and handedness phenotypes.

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 were consistent with prior research: 0.23 and 0.22 respectively for the EHI and QHP. For the language laterality index, however, the twin-cotwin correlations were very close to zero for both MZ and DZ twins, and the heritability estimate was zero.

Conclusions: A single study showing negligible heritability for language laterality 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 be lower. We cannot rule out the possibility that true heritability of differences in language lateralization is non-zero, but results indicate that the heritability of this trait is low at best. Stochastic variation in neurodevelopment appears to play a major role in determining cerebral lateralisation.

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 95% of right-handers have left-hemisphere language, this is true for around 67% of left-handers (see Bishop, 1990a, for review).

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 zygozity. 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 zygozity. 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 $0.5 \cdot a^2 + c^2$, so we can estimate heritability ($h^2$) 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 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 concordance 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.

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. 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. 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 oversampling of families with several left-handed members, but 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 the effects of 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 has 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). The (numerous) inherited causal mutations would be expected to differ from family to family. However, all would be shared between MZ twins and their cotwins, and in only 50% of cases for DZ twins. 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.

Zygosity determination
(The following paragraph is copied from Newbury et al., 2018. Origene 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 from 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.

Table 1. N twin pairs by zygosity and sex.

<table>
<thead>
<tr>
<th>Twin type</th>
<th>All</th>
<th>With fTCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ female</td>
<td>49</td>
<td>30</td>
</tr>
<tr>
<td>MZ male</td>
<td>47</td>
<td>35</td>
</tr>
<tr>
<td>DZ female</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>DZ male</td>
<td>29</td>
<td>21</td>
</tr>
<tr>
<td>DZ male/female</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>194</td>
<td>141</td>
</tr>
</tbody>
</table>

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

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 interpolate 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.

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. 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.

In the current dataset, the correlation between this mean-based LI and the traditional peak-based LI is very strong: Pearson correlation = 0.907, DF = 306, 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.

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.

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.

### 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 (Rijstjik & 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 (a²), shared environmental (c²) 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 (version 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). 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 0.2 or more, using the power.ACE.test function in umx. This showed that for the larger sample of twins for whom handedness data were available, there is 85% power to detect heritability of 0.2 in an AE model with alpha of 0.05, and for the smaller sample with language laterality data, power of 73%. In this smaller subset with language laterality data, 80% power is obtained for heritability of 0.215, and over 95% power for heritability of 0.33. Thus, even the low heritability handedness estimates reviewed in the Introduction should be detectable with this sample. For language laterality, there is around a 3 in 4 chance of detecting a true but small effect of around 0.2, and strong power to detect the higher heritability reported in the DTI study by Jahanshad et al. (2010).

Results

Figure 1 shows the distributions of scores obtained on the two handedness measures and the language laterality index from fTCD, and Figure 2 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 2 shows that the correlations between two members of a twin pair are low for all measures.

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 0.19 for the Edinburgh Handedness Inventory and 0.17 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.

Because of concerns that the non-normality of the handedness data could distort heritability estimates, analyses of the EHI
The data from handedness measures are jittered; Spearman correlation coefficients are shown.

and QHP were repeated using the umx package (Bates, 2018) to run an ordinal version of ACE analysis. This gave slightly higher 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.

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 3, 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, showed twintwin correlations close to zero for both MZ and DZ twins, and appeared therefore to be determined entirely by chance.

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 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 2, 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

---

**Table 2. Heritability estimates from AE model for different measures.** MZ = monozygotic; DZ = dizygotic.

<table>
<thead>
<tr>
<th>Measure</th>
<th>N MZ</th>
<th>N DZ</th>
<th>$a^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edinburgh Handedness Inventory</td>
<td>96</td>
<td>98</td>
<td>0.190</td>
</tr>
<tr>
<td>Quantification of Hand Preference</td>
<td>96</td>
<td>98</td>
<td>0.170</td>
</tr>
<tr>
<td>Language laterality index</td>
<td>65</td>
<td>76</td>
<td>0.000</td>
</tr>
</tbody>
</table>

---

**Figure 2. Scatterplots by zygosity for handedness and language laterality indices.** The data from handedness measures are jittered; Spearman correlation coefficients are shown.
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 (Carrión-Castillo et al., 2019).

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 also been remarkably difficult to find any genetic variants consistently linked with variations in handedness. It subsequently became clear, however, that there is a genetic effect, but it is small and only clearly detectable in large samples. 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), 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). 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, 1990b).

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. 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. Further research will be needed to determine whether heritability of language lateralisation is low but real, or whether this is the best candidate to date of 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/CPKH (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)

---

**Table 3. Extreme probands, t-tests comparing MZ/DZ co-twins. MZ = monozygotic; DZ = dizygotic; QHP = Quantification of Hand Preference, LI = laterality index.**

<table>
<thead>
<tr>
<th></th>
<th>Edinburgh Handedness</th>
<th>QHP</th>
<th>Language LI</th>
</tr>
</thead>
<tbody>
<tr>
<td>N MZ</td>
<td>21</td>
<td>40</td>
<td>22</td>
</tr>
<tr>
<td>N DZ</td>
<td>40</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>MZ proband mean (SD)</td>
<td>1.43 (1.66)</td>
<td>3.61 (3.86)</td>
<td>-1.06 (2.89)</td>
</tr>
<tr>
<td>DZ proband mean (SD)</td>
<td>1.15 (1.42)</td>
<td>3.89 (4.22)</td>
<td>-1.24 (3.29)</td>
</tr>
<tr>
<td>MZ co-twin mean (SD)</td>
<td>6.86 (3.48)</td>
<td>13.46 (7.14)</td>
<td>1.34 (2.49)</td>
</tr>
<tr>
<td>DZ co-twin mean (SD)</td>
<td>6.58 (3.63)</td>
<td>14.22 (7.19)</td>
<td>2.14 (2.57)</td>
</tr>
<tr>
<td>t</td>
<td>0.30</td>
<td>-0.52</td>
<td>-1.19</td>
</tr>
<tr>
<td>df</td>
<td>42.3</td>
<td>84.4</td>
<td>44.9</td>
</tr>
<tr>
<td>p</td>
<td>0.768</td>
<td>0.606</td>
<td>0.239</td>
</tr>
</tbody>
</table>
• Individual trial LIs.csv (CSV file containing individual language laterality file data)
• Doppler_raw.zip (Zip file containing individual raw Doppler readings)

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

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

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

Acknowledgements
Our thanks to the families who took part in the study, and school staff who helped with assessment arrangements. We also thank research assistants Eleanor Payne, Nicola Gratton, Georgina Holt, Annie Brookman, Elaine Gray, Louise Atkins, Holly Thornton and Sarah Morris for help with data collection and data entry. We are grateful to Paul A. Thompson for help developing data preprocessing scripts.

References

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.


