



# Assessing the reliability of an online behavioural laterality battery: A pre-registered study

Adam J. Parker, Zoe V. J. Woodhead, Paul A. Thompson & Dorothy V. M. Bishop

To cite this article: Adam J. Parker, Zoe V. J. Woodhead, Paul A. Thompson & Dorothy V. M. Bishop (2021) Assessing the reliability of an online behavioural laterality battery: A pre-registered study, *Laterality*, 26:4, 359-397, DOI: [10.1080/1357650X.2020.1859526](https://doi.org/10.1080/1357650X.2020.1859526)

To link to this article: <https://doi.org/10.1080/1357650X.2020.1859526>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 15 Dec 2020.



[Submit your article to this journal](#)



Article views: 1467



[View related articles](#)



[View Crossmark data](#)







Citing articles: 3 [View citing articles](#)

ORIGINAL ARTICLE



## Assessing the reliability of an online behavioural laterality battery: A pre-registered study

Adam J. Parker , Zoe V. J. Woodhead , Paul A. Thompson  and Dorothy V. M. Bishop 

Department of Experimental Psychology, University of Oxford, Oxford, UK



### ABSTRACT


Studies of cerebral lateralization often involve participants completing a series of perceptual tasks under laboratory conditions. This has constrained the number of participants recruited in such studies. Online testing can allow for much larger sample sizes but limits the amount of experimental control that is feasible. Here we considered whether online testing could give valid and reliable results on four tasks: a rhyme decision visual half-field task, a dichotic listening task, a chimeric faces task, and a finger tapping task. We recruited 392 participants, oversampling left-handers, who completed the battery twice. Three of the tasks showed evidence of both validity and reliability, insofar as they showed hemispheric advantages in the expected direction and test-retest reliability of at least  $r = .75$ . The reliability of the rhyme decision task was less satisfactory ( $r = .62$ ). We also confirmed a prediction that extreme left-handers were more likely to depart from typical lateralization. Lateralization across the two language tasks (dichotic listening and rhyme judgement) was weakly correlated, but unrelated to lateralization on the chimeric faces task. We conclude that three of the tasks, dichotic listening, chimeric faces and finger tapping, show considerable promise for online evaluation of cerebral lateralization.

**ARTICLE HISTORY** Received 14 August 2020; Accepted 1 December 2020

**KEYWORDS** Cerebral lateralization; hemispheric asymmetry; language; online behaviour research methods; reliability

Hemispheric specialization, or lateralization, is a ubiquitous characteristic of the human brain, yet many questions remain unanswered as to its causes and consequences. Much of the ambiguity in the published literature may arise from unreliable measures of lateralization and underpowered studies.

**CONTACT** Adam J. Parker  [adam.parker@psy.ox.ac.uk](mailto:adam.parker@psy.ox.ac.uk)  Department of Experimental Psychology, University of Oxford, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

 Supplemental data for this article can be accessed <https://doi.org/10.1080/1357650X.2020.1859526>

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Without well-powered experiments that use reliable tasks, it becomes difficult to address many research questions regarding the relationship between laterality indices and their stability over time. Concerning the relationship between indices, unless a task is reliable, the absence of a correlation cannot be interpreted as evidence for independence of function (Parsons, Kruijt, & Fox, 2019). Reliability can either be assessed within a test session (e.g., split-half reliability, from correlations between odd and even items) or from repeated assessments on different occasions (test-retest reliability). The latter is generally preferred as it includes all test items and will incorporate fluctuations in performance due to changes in state or test conditions from one occasion to another. Assessment of test-retest reliability, however, requires large samples of people who undergo repeated testing, and this is seldom feasible for in-person testing. Online testing allows for the recruitment of much larger samples but raises questions about test validity when opportunities for experimental control are reduced. Here we introduce an online behavioural laterality battery for which we established the test-retest reliability using a sample of 392 participants. The behavioural tasks included were: a novel rhyme decision task, a dichotic listening task (Hugdahl et al., 2009), a chimeric faces task (Burt & Perrett, 1997), and a novel finger tapping task. The validity of the tasks was assessed in terms of whether lateralization was observed in the predicted direction, based on prior literature with in-person testing on analogous tasks. We also examined two pre-registered questions: (1) do extreme left-handers show a higher rate of atypical laterality than right-handers on each behavioural laterality task? (2) do the laterality indices from different tasks correlate?

### ***Behavioural laterality tasks and their reliability***

Behavioural laterality tasks have long been used to establish hemispheric dominance for certain cognitive functions. Results from visual half-field and dichotic listening paradigms support the view that the left hemisphere plays the dominant role in language processing (Hugdahl, 2000; Ocklenburg & Güntürkün, 2018). By contrast, visual half-field paradigms indicate that the right hemisphere is involved in the processing of faces, global feature processing, and visuospatial attention (Brederoo, Nieuwenstein, Cornelissen, & Lorist, 2019). Below, we review how each paradigm has been implemented to study language lateralization and the detection of emotion in faces, and what is known about reliability and validity.

### ***Lateralized visual word processing***

Lateralized visual word processing has been studied using the visual half-field paradigm. In this paradigm, visual stimuli are briefly presented to the left or right of a central fixation stimulus. The visual pathways from the left and right

sides of space to the visual cortex are crossed for non-foveal viewing, and so the right visual-field (RVF) projects directly to the left hemisphere and the left visual-field (LVF) to the right hemisphere. The corpus callosum allows for transfer between hemispheres, but the rationale of the method is that there will be a processing advantage for language stimuli that are directly projected to the language-dominant hemisphere. This method has been used since the 1950s (Mishkin & Forgays, 1952). Bourne (2006) reviewed the literature, which consistently shows superior processing for letter strings presented in the RVF, relative to the LVF, in most people.

Despite wide reporting of a RVF advantage at the population level, the validity of visual half-field language tasks has been questioned. Krach, Chen, and Hartje (2006) compared the laterality indices (LIs) produced during a visual half-field experiment with those produced using functional Transcranial Doppler Sonography (fTCD). The visual half-field task consisted of lexical decisions made towards abstract nouns presented in the LVF and RVF. During the experiment, the blood flow to the left and right hemispheres of 58 participants was measured. In addition, blood flow was measured during a word generation task—the gold standard task for assessing hemispheric dominance for language. Despite showing the expected RVF advantage on the visual half-field task, the behavioural LIs produced on the visual half-field paradigm did not correlate with the fTCD LIs produced during word generation ( $r = .18$ ) or the visual half-field task itself ( $r = -.004$ ). The lack of validity here likely reflects the poor reliability of the visual field task as split-half reliability was very low (response time split-half:  $r = .14$ ; accuracy split-half:  $r = .20$ ). This illustrates the difficulty of interpreting associations between laterality indices when the reliability of measurement is so low.

In recent years it has become clear that the reliability and validity of visual half-field tasks to identify the language-dominant hemisphere is influenced by several methodological aspects. For example, Hunter and Brysbaert (2008) noted stronger and more stable visual field differences when stimuli were presented bilaterally relative to those where only one stimulus was presented either in the LVF or in the RVF (Boles, 1987, 1990, 1994; see also Iacononi & Zaidel, 1996). Hunter and Brysbaert argued that this competition between stimuli during bilateral presentation enhances the advantage to the stimulus presented in the language dominant hemisphere. Reliability and validity also depend on there being a sufficient number of trials and brief stimulus presentation to prevent eye movements. When these methodological aspects are addressed, the validity of visual half-field tasks seems promising. For example, Hunter and Brysbaert (2008) used two visual half-field tasks, word naming (192 trials) and picture naming (160 trials) in which stimuli were bilaterally presented and masked at offset. The two tasks correlated strongly ( $r = .80$ , 95%  $CI = .598-.956$ ) in a sample of 26 left-handers. Encouragingly, the LIs from these visual half-field tasks also

correlated positively with laterality from functional Magnetic Resonance Imaging (fMRI) on a word generation task in 10 left-handers. However, the small sample size meant that the confidence intervals around these estimates were very wide: word naming,  $r = .63$  (95%  $CI = .001-.902$ ), picture naming,  $r = .77$  (95%  $CI = .272-.943$ ). More recently, however, Van der Haegen, Cai, Seurinck, and Brysbaert (2011) reported similar findings in a sample of 50 participants. Again, participants completed behavioural word and picture naming visual half-field tasks and then a word generation task under fMRI. It is reassuring that the LIs for the behavioural measures were correlated with the fMRI LI: word naming,  $r = .64$ , picture naming,  $r = .65$ . Van der Haegen et al. also report a set of analyses involving 250 participants who completed the behavioural visual half-field tasks while their eye movements were monitored. Two sets of LIs were calculated; one that controlled for fixation error and one that did not. Importantly, the LIs which did and did not control for fixation error were highly correlated with word naming,  $r = .94$ , and picture naming,  $r = .98$ , indicating that the inclusion of eye movement control adds little additional value when a visual half-field task employs mechanisms to encourage central fixation and bilateral presentation.

A promising line of work has emerged which further highlights the potential for well-designed visual half-field tasks to assess language lateralization. First introduced by Willemin et al. (2016), the translingual lexical decision task involves participants viewing letter strings either side of fixation and judging whether the left, right, or neither of the letter strings is a valid word. With 100 participants completing 256 trials, Willemin et al. reported a RVF advantage at the population level. Since then this effect has been replicated in a sample of 496 participants covering six international languages, with evidence of a RVF for each language studied (Hausmann et al., 2019). Not only does this task show validity, it seems reliable given Brederoo et al.'s (2020) observation that the task yields good split-half reliability when tested with 122 participants:  $r = 0.86$  for error rates and  $r = 0.80$  for response times.

### *Lateralized auditory verbal processing*

Simultaneous presentation of different auditory stimuli to left and right ears for immediate recall was a method originally used in studies of attention, but it became evident that there was usually better recall of verbal stimuli presented to the right ear (Kimura, 1967). Both ears project to both hemispheres, but the weaker ipsilateral pathways appear to be suppressed when the two sides compete, and so the right ear, with a contralateral projection to the left hemisphere, will have an advantage in tasks that involve language processing.

Over the years, various dichotic listening paradigms have been used (Bryden, 1988; Westerhausen, 2019). As with the visual half-field paradigm,

the test-retest reliability of dichotic listening paradigms has been far from optimal (Kelley & Littenberg, 2019; Voyer, 1998). However, recent data from Westerhausen and Samuelsen (2020) indicated that when designed optimally, the dichotic listening paradigm can be reliable. First, based on a review of the published literature, Westerhausen and Samuelsen outlined eight design features that should be used in an optimal dichotic listening paradigm and explained their justification for each. Their recommendations were: (1) stop consonant–vowel (CV) syllables as stimulus materials; (2) CV stimulus pairs from the same voicing category; (3) alternating trials of voiced and unvoiced stimulus pairs; (4) free-recall instruction; (5) one stimulus pair per trial; (6) a paradigm length of 120 dichotic trials; (7) all stimulus pairs presented at an equal frequency to the two sides; and (8) inclusion of binaural (homonymic) trials. Putting these criteria into practice, Westerhausen and Samuelsen administered a modified dichotic listening paradigm using both a verbal and manual response format to 50 right-handers. Interestingly, the full paradigm yielded test-retest reliability of  $r = .93$  for verbal responses and  $r = .91$  for manual responses. Thus, the dichotic listening task provides a promising paradigm for those wanting to investigate language lateralization without reliance on visual stimuli.

To date, little work has been conducted online with visual stimuli. By contrast, the dichotic listening paradigm has been implemented online with considerable success. Bless et al. (2013) developed a mobile device version of the consonant-vowel dichotic listening paradigm, with 36 simultaneous sound pairs. Under laboratory conditions, with a sample of 76 (33 Norwegian, 43 Australian) participants, they reported test-retest reliability of  $r = 0.78$  when using the mobile app. Furthermore, both subgroups of participants showed clear right ear advantages at the population level. In a subsequent large-scale study, Bless et al. (2015) administered the same application to a sample of over 4,000 participants with more than 60 different language backgrounds. Again Bless et al. (2015) reported a clear right ear advantage at the population-level. They then used the large dataset to examine cultural/linguistic effects on the degree of lateralization; reporting that the degree of lateralization appears to be modulated by linguistic background. Together these studies demonstrate the potential for online laterality research to generate large amounts of data while maintaining a tolerable degree of reliability.

### *Lateralized face processing*

In studies of face processing, it is possible to use a modification of the visual half-field paradigm in which single stimuli are viewed, composed of vertically split chimeric faces, where each half displays a different emotion or identity. Our focus here is on emotion detection in chimeric faces, where bias is usually found towards reporting the emotion shown in the left hemiface.

The reliability of the chimeric faces tasks seems promising. Early work by Levy, Heller, Banich, and Burton (1983) using two vertically presented chimeric faces, found split-half reliability of  $r = .93$ . The validity of the chimeric faces task has been demonstrated using behavioural and fMRI data but in a task of recognizing facial identity rather than emotion. In Yovel, Tambini, and Brandman's (2008) study, participants had to recognize centrally presented chimeric faces, which showed different identities in the right and left visual fields. It is encouraging to note that the behavioural measure significantly correlated with the fMRI laterality index (LI) calculated from activation in the fusiform face area ( $r = .49$ ). However, the sample size used here was small: with 17 participants, the 95% confidence interval around the observed correlation of .49 ranged from .012 to .786.

### ***Behavioural laterality tasks and (a)typical lateralization***

The studies reviewed so far have indicated consistent population biases in lateralization on behavioural tasks, but also there are also substantial individual differences. One factor that has been associated with the degree of lateralization is handedness (for a comprehensive review, see Carey & Johnstone, 2014). Neuroimaging data indicate that for language tasks these differences in lateralization primarily reflect the contribution of a subset of strong left-handers who are right hemisphere dominant for language, rather than indicating that this relationship operates on a continuum (Mazoyer et al., 2014). This association can be used to validate behavioural measures of lateralization: we would expect to see a departure from typical lateralization in strong left-handers.

With regards to visual half-field tasks for studying language lateralization, there is evidence that the visual field advantage varies with handedness (Elias, Bulman-Fleming, & McManus, 1999; Weekes, Capetillo-Cunliffe, Rayman, Iacoboni, & Zaidel, 1999). However, a relatively well-powered study involving 100 participants, conducted by Willemin et al. (2016), reported no hint of any effect of handedness, when coded as left and right, or when divided into left ( $N = 19$ ), mixed ( $N = 24$ ) and right-handed ( $N = 57$ ) groups, on the visual field advantage during a lexical decision task. This result illustrates the recurring challenge of interpreting null results in this field: do they reflect a true lack of relationship, inadequate power to detect small effects, or specific methodological factors? Brederoo et al. (2020) reported good split-half reliability for this task, suggesting the lack of handedness effect is unlikely to be due solely to measurement error. We may note that Willemin et al. asked participants to respond via manual response, whereas other studies have involved an active speech production component; manual responding could interact with handedness to influence the laterality index on this task.

On dichotic listening tasks, the clear right ear (i.e., left hemisphere) advantage that is noted at the population level (see Hugdahl et al., 2009; Westergaard, 2019, for a review) is reduced in left-handers (Bethmann, Tempelmann, De Bleser, Scheich, & Brechmann, 2007; Bryden, 1970, 1988; Hugdahl, 2003). Recently, Karlsson, Johnstone, and Carey (2019) employed a consonant-vowel version of the dichotic listening task where participants reported which one of six syllables that they heard most clearly. In study 1, 87% of right-handers and 78% of non-right-handers showed a right ear advantage. For study 2, 85% of right-handers and 79% of non-right-handers showed a right ear advantage.

For chimeric faces tasks, which typically show a left visual field advantage, there again appears to be a greater departure from the typical pattern of laterality for left-handers (e.g., Gilbert & Bakan, 1973; Levy et al., 1983; Roszkowski & Snelbecker, 1982). This was also studied by Karlsson et al. (2019). When viewing a single face with a different emotion shown in either hemiface, 75% of right-handers and 60% of non-right-handers showed a leftwards bias. For the second chimeric faces task that involved two chimeric faces arranged vertically, a similar pattern emerged: 73% of right-handers and 57% of non-right-handers showed a leftwards bias.

From the above-reviewed studies, it appears that handedness is related to laterality across a range of tasks (for further discussion see Carey & Johnstone, 2014). This relationship points towards a general laterality factor which influences lateralization across tasks and behaviours, as predicted by several theoretical accounts (Behrmann & Plaut, 2015; Hellige, 1993; Kosslyn, 1987). For example, the causal complementarity principle proposes that the lateralization of processes involving the same hemisphere will be lateralized to the homologue in the opposite hemisphere. With reference to language processing and face emotionality processing, it would be predicted that language processing would be left lateralized and face processing would be right lateralized, and an increase of one co-occurs with an increase of another. Indeed, several studies have reported such patterns. In support, Brederoo et al. (2020) and Dundas, Plaut, and Behrmann (2015) have reported significant correlations between LIs for word and face processing. In a study of left-handers using fMRI with five tasks, Gerrits, Verhelst, and Vingerhoets (2020) found that the just under half the participants showed either standard laterality (left-lateralised for language and praxis, right-lateralised for face memory, line-bisection and vocal emotion) or a complete reversal of this pattern, providing clear evidence against the idea that these functions are lateralized independently. Nevertheless, the data indicated that this association is far from perfect (Badzakova-Trajkov, Häberling, Roberts, & Corballis, 2010) and some studies have found no correlation between laterality indices (Bryden, Hécaen, & DeAgostini, 1983; Rosch, Bishop, & Badcock, 2012; Van der Haegen & Brysbaert, 2018; Whitehouse & Bishop, 2009). Our ultimate goal



is to unpick whether dissociations between laterality indices are evidence of multiple independent lateral biases across tasks, or whether there is a more prosaic explanation in terms of poor reliability or validity of tasks.

To date, the focus has been in testing whether lateralization shows complementarity across tasks that are usually mediated by the left and right hemispheres, especially whether there is a relationship between language vs visuospatial functioning. However, emerging lines of evidence have suggested that even within domains there may be dissociated patterns of lateralization across similar tasks. Woodhead, Bradshaw, Wilson, Thompson, and Bishop (2019) reported data for 37 participants engaged in six language tasks under transcranial Doppler sonography (fTCD) and examined the factorial structure of lateralization across tasks using structural equation modelling. They pitted two models against each other. The first assumed that language laterality is a unidimensional trait and a single factor was used to model covariance. The second assumed that two laterality factors underlie the pattern of covariance. The results supported a two-factor model at the population level. However, further examination of the data revealed that this was driven by seven left-handers. Furthermore, Woodhead, Thompson, Karlsson, and Bishop (2020) examined this factor structure in an extended sample of 74 participants (43 right-handers and 31 left-handers). The data this time indicated that a two-factor model was preferred for both right- and left-handers, with the preference for a two-factor model being stronger for left-handers. Together these data suggest that laterality indices within the domain of language may be dissociated.

### *The current study*

The principal goal of the current study was to evaluate the reliability and validity of an online behavioural laterality battery. Specifically, we examined whether it is possible to observe the usual laterality biases at the population level in tests of language and emotion processing and whether it is possible to achieve test-retest reliability of  $r = .65$  or above using online behaviour research methods. This is an arbitrary cutoff, selected to represent a lower bound of reliability for a measure to be potentially useful for analysing individual differences. In addition, for tasks where satisfactory reliability was achieved, we considered how many trials were needed to achieve this. The tasks included in the battery were a novel rhyme decision visual half-field task, a consonant-vowel dichotic listening task, a chimeric faces visual half-field task, and a novel finger tapping task. These behavioural tasks were completed twice on two different occasions.

When selecting specific tasks in our online battery we were strongly guided by pragmatic considerations, i.e., how feasible would it be to administer the task remotely via an online platform. We aimed to include at least

two language measures, one tapping into receptive language and one expressive, to test the notion that language laterality can dissociate across different tasks (cf., Woodhead et al., 2019, 2020). We also aimed to include a task that involved face processing so that we could test for complementarity of lateralization across tasks that usually lateralize to left and right hemispheres. Finally, a behavioural measure of motor laterality was included, to potentially provide an online analogue of relative hand skill similar to that used by Annett (1970).

Ultimately we aimed to develop tasks that would meet three criteria. First, the task should show good test-retest reliability. Second, it would have to show validity as demonstrated by showing significant lateralization at the population level. Third, it should be sensitive to the subtle differences in laterality that have been reported between right- and non-right-handers.

For receptive language, we selected the dichotic listening paradigm, which has already been shown to yield reliable results when used online via a mobile app (Bless et al., 2013, 2015). Furthermore, Karlsson et al. (2019) have shown that this paradigm is sensitive to the subtle differences in language asymmetry between right and non-right handers, making it an obvious choice for inclusion.

For expressive language we aimed to use a visual half-field task. While the task reported by Brysbaert and colleagues looked promising, it involved overt verbal responses which would be difficult to accommodate in an online format. Thus we decided to develop a new rhyme judgement task that was designed to meet the limitations of online testing, and which would require covert naming of stimuli presented in the two half-fields (see below for details).

Given that numerous studies looking at the relationship between lateralization on language and non-language tasks have focused on the use of face stimuli (e.g., Brederoo et al., 2020), we chose to include a face task to conduct similar comparisons. Regarding the chimeric faces task, Levy et al. (1983) reported good split-half reliability while Karlsson et al. had reported that, like the dichotic listening task, the chimeric faces task is sensitive to the differences in perceptual asymmetry in right- and non-right-handers.

Our final online measure was a finger-tapping test used as an online alternative to the well-known Annett (1970) pegboard task. There is suggestive evidence that a continuous measure of relative skill of the two hands may be more effective than more conventional hand preference measures for capturing individual differences in motor laterality that relate to language lateralization (Annett, 2002; Flowers & Hudson, 2013). Performance measures such as this can provide a continuous measure of subtle variation in relative hand skill, and also correlate well with measures of hand preference (e.g., Steenhuis & Bryden, 1999). The pegboard task is one of the most widely used performance tasks (see Hodgson & Hudson, 2018, for discussion), but the requirement

for apparatus makes it unfeasible as an online test. We therefore took the opportunity to explore the reliability and validity of a finger-tapping task that required the participant to repeatedly tap a short sequence on the computer keyboard with the index finger of the preferred or non-preferred hand, which captures one aspect of peg-moving, namely precise aiming of sequential movements.

Our first set of predictions (hypothesis 1) concerned the validity of the tasks as indicators of cerebral lateralization. Based on previous work, we predicted that at the population level: (1) the laterality indices for rhyme decision would show a RVF advantage (indicating left hemisphere dominance for language processing); (2) the laterality indices for the dichotic listening task would show a right ear advantage (indicating left hemisphere dominance for language processing); and (3) the laterality indices for chimeric faces would show a LVF advantage (indicating right hemisphere dominance for face or emotion processing). Predictions regarding the finger tapping task were less clear; our goal was to evaluate a test of relative hand skill that might provide an online surrogate for the kind of peg-moving task used by Annett (1970). Validity of the task would be demonstrated if a right-hand advantage was found in right-handers, and a left-hand advantage in left-handers. (Note that direct comparison with performance on a pegboard was not feasible as the study was conducted during a pandemic when in-person testing could not be done).

Our second prediction (hypothesis 2) related to handedness and atypical lateralization. We predicted that strong left-handers (i.e., those with laterality quotients of  $-90$  or less on the Edinburgh Handedness Inventory; Oldfield, 1971) would be more likely to show atypical laterality (i.e., laterality in the opposite direction to our population-level predictions) than right-handers. This prediction followed from observations by Mazoyer et al. (2014, p. 1) who studied a large group of left- and right-handers using fMRI to measure language laterality. They concluded,

concordance of hemispheric dominance for hand and for language occurs barely above the chance level, except in a group of rare individuals (less than 1% in the general population) who exhibit strong right hemisphere dominance for both language and their preferred hand.

Our final prediction (hypothesis 3) related to the inter-relationships between laterality indices. With online testing, there is the opportunity to test inter-relationships between laterality indices with an adequately powered sample using measures of known reliability. Prior studies have been inconclusive because it has been difficult to gather large amounts of data to establish test-retest reliability. For instance, Van der Haegen and Brysbaert (2018), found that language laterality did not correlate with face laterality in visual half-field tasks. However, their study was restricted to left-handers, who

may be atypical in this regard, and they noted that lack of association was hard to interpret because it may be due to unreliability of the individual measures. In the pre-registration for our current study, we predicted that for right-handers, a two-factor model where the two language tasks are correlated with each other but not with the chimeric faces (model 2F) would be the best fit to the data, compared to either a single factor model in which laterality indices are correlated between the language and chimeric face tasks (model 1F), or a base model where each task is independently lateralized (model 3F). Given our previous findings using fTCD to study dissociations in language laterality (Woodhead et al., 2019, 2020), we further predicted that model fit may differ for right- and left-handers, with more dissociation between tasks in the latter group. While we report this analysis in the Supplementary Materials, we report an analysis of the whole sample in the main body as we found that the results from subgroup analyses were very unstable with the small sample of left-handers who were eligible for analysis, and could favour different models when quite minor changes were made to the sample.

## Method

This project was pre-registered on the Open Science Framework (OSF) before the start of data collection. The registration form along with all task materials, analysis scripts and anonymised data can be found on the OSF at <https://osf.io/qsxcv/>. Working versions of the dichotic listening, chimeric faces, and finger tapping tasks are available on Gorilla Open Materials: <https://gorilla.sc/openmaterials/104636>.

## Participants

As our ultimate goal is to use these measures in individual differences research, we aimed to optimize the task to achieve observed test-retest reliability of at least  $r = .65$ . Our initial power analysis showed that to achieve sufficient precision around the estimate of reliability, we would require 300 participants. Out of the 300 participants, we originally specified that 80 would be left-handers. This would give us power to apply our planned model-fitting analysis to left-handers as a separate group. In practice, our final sample exceeded 300, because we discarded participants from individual tasks for failing to meet prespecified criteria (see below).

In total, 441 native English speakers were initially recruited through Prolific Academic ([www.prolific.co](http://www.prolific.co)) with 411 participants completing both sessions. Prolific is a platform for recruiting and screening paid research participants according to experimenter-specified criteria. Participants were 18- to 50-years-old, had normal or corrected-to-normal vision and reported no

history of neurological disease (e.g., head injury, brain tumour, stroke) or significant hearing loss or impairment.

This study was approved by the University of Oxford's Medical Sciences Inter-Divisional Research Ethics Committee (reference number: R66638/RE002) and conducted following the principles of the Declaration of Helsinki. Participants gave informed consent and were paid £15 for their participation.

### ***Online behavioural laterality battery***

The battery included four measures of background information and four experimental assessments of behavioural laterality that were built and administered using the Gorilla Experiment Builder ([www.gorilla.sc](http://www.gorilla.sc)), a cloud-based tool for collecting data in the behavioural sciences (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). Each task is described below.

### ***Basic demographics questionnaire***

The basic demographics questionnaire asked participants to report age, gender, years in education, and whether they were bilingual. We also asked questions about handedness and footedness. For handedness, we asked: which is your dominant hand (i.e., which hand do you prefer to use for tasks such as writing, cutting, and catching a ball?). For footedness, we asked: which foot do you normally use to step up on a ladder/step? For both questions, participants could answer "left", "right", or "no preference / Don't know". No specific predictions were made about footedness, but it was included to provide a comprehensive picture of behavioural laterality.

### ***Tests of ocular dominance***

Adapted versions of the Miles test (Miles, 1929) and the Porta test (Porac & Coren, 1976) were used to determine each participant's eye dominance in central gaze (i.e., when looking straight ahead). Each test classified participants as being either left or right eye dominant. We did not have specific predictions about ocular dominance but gathered this information to further characterize behavioural laterality of the sample. In practice, the two tests agreed perfectly, so we report here just the Miles test.

### ***Edinburgh handedness Inventory***

The Edinburgh Handedness Inventory (EHI; Oldfield, 1971) was administered to quantify handedness on a continuum. For 10 everyday activities, participants indicated their preferred hand use on a 5-step scale (right hand strongly preferred, right hand preferred, no preference, left hand preferred, left hand strongly preferred). The EHI was scored in the standard way, which involves combining information about direction and exclusiveness of hand preference. For simple contrast between left- and right-handers, we

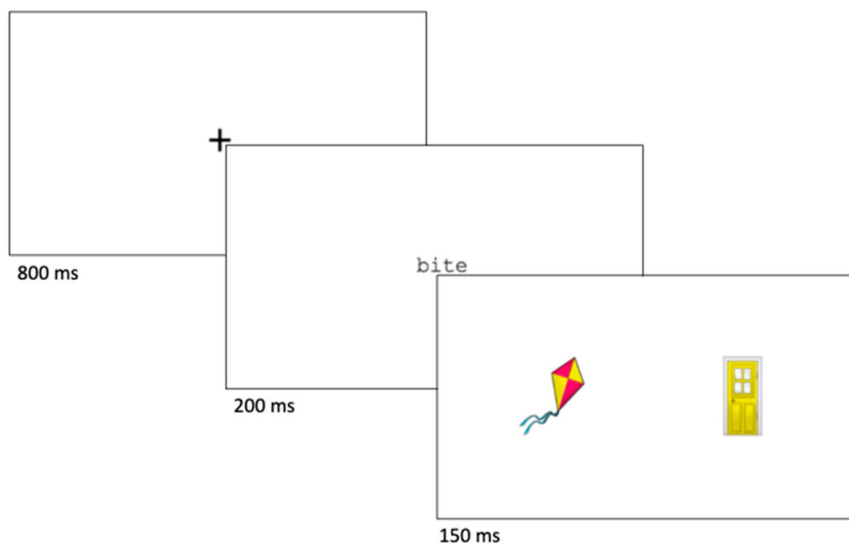
treated indices greater than 0 as right-handed, and indices less than 0 as left-handed. For analysis of extreme left-handers, we focussed on left-handers with an index of  $-90$  or less, which corresponds to near-exclusive use of the left hand.

### *LexTALE*

The Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012) was used to screen for normal-range language ability. Participants saw 60 letter strings and had to indicate which words they knew. Of the 60 strings, 40 were actual English words and 20 were non-words. LexTALE scores were corrected for the unequal proportion of words and non-words by averaging the percentages corrected for these two item types (e.g.,  $((\text{number of words correct}/40 \times 100) + (\text{number of nonwords correct}/20 \times 100))/2$ ). Accordingly, the LexTALE scores ranged from 0 to 100.

### *Rhyme decision task*

The rhyme decision task was developed for this project as a novel visual half-field technique designed to engage lateralized language processes. On each trial, participants were asked to decide whether a foveally presented letter string rhymed with the name of an image shown in either the LVF or RVF. With online presentation, a particular concern is a potential for eye movements away from fixation. We relied on the fact that participants had to focus initially on the central written word to do the task, and thus the task encouraged fixation to a central location at the start of the trial. We subsequently displayed pictures under brief bilateral presentation, which has been shown to give adequate control for eye movements in previous visual half-field experiments with verbal naming (Beaumont, 1982). Studies that directly monitored eye movements in this setting reported failures of fixation on only 0.5% of trials (Geffen, Bradshaw, & Nettleton, 1972; but see also Bourne, 2006). Pictures of familiar objects were presented for rhyme matching to the written word, to ensure that the participant would have to covertly name the picture, thereby engaging left hemisphere speech processing. We used covert rather than overt naming (cf. Hunter & Brysbaert, 2008) to make the task suitable for online presentation. This obviated the need for measuring latencies of oral responses, which would be challenging in the online context, both in terms of the need for a high-quality microphone, and possible intrusion of background noises interfering with recording. Furthermore, the evaluation of oral responses would require additional scoring following data collection which would be a taxing procedure for researchers and could introduce subjectivity into the measure. Voice recording would also require additional ethical approvals with regards to the storing and processing of the data.



**Figure 1.** Schematic illustration of a trial on the rhyme decision task. In the example, the word “bite” is presented along with the two images “kite” and “door”. When making the correct judgement, participants would respond with by pressing “S” to indicate that the rhyme was on the left.

**Materials.** Twenty-six monosyllabic words were presented in foveal vision. Each word was 3–4 letters in length ( $mean = 3.8$ ;  $SD = 0.40$ ) and had an average Zipf frequency [ $\log_{10}(\text{frequency per million words}) + 3$ ] of 4.4 ( $SD = 0.72$ ) when using the SUBTLEX-UK database (Van Heuven, Mander, Keuleers, & Brysbaert, 2014). Each foveal word was paired with an image that corresponded to one of 26 rhyming monosyllabic words which had a mean Zipf frequency of 4.5 ( $SD = 0.48$ ). For example, the word “bite” was paired with the image “kite” and the word “more” was paired with the image “door”. Each word-image pair was then paired with another word-image pair for which there was no inter-pair rhyme (i.e., bite-kite was paired with more-door). This resulted in 13 word-image couplets. For each of the foveally presented words in a given word-image couplet, the corresponding images could appear in either the LVF or RVF. Thus, for each of the 13 couplets, four combinations were possible. For the pairing of bite-kite and more-door the following stimuli were created: kite—bite—door, door—bite—kite, door—more—kite, and kite—more—door. In total, there were 52 unique stimuli. On a given trial, each image was displayed at 2.5 degrees eccentricity of the foveal word (see Figure 1).

Word stimuli were presented in black lowercase Tahoma font on a white background. Font was scaled to 250% equating to 28pt font. The image pairs were presented at approximately 2.5 degrees of visual angle either

side of fixation. As screen size was likely to vary between participants, the size of the stimuli and eccentricity of images relative to fixation was maintained using Gorilla's scaling tool.

**Procedure.** The trials were preceded by a familiarization phase in which each of the stimulus pictures was presented together with its written name. After familiarization, 260 stimuli (52 unique stimuli each shown 5 times) were displayed across four blocks (65 items per block). Each of the four blocks also contained five items for which there was no rhyme. For example, the word "moon" was displayed with the images for kite and door: kite—moon—door. These foil items were included in an attempt to prevent participants from preferentially responding to items in either the LVF or RVF. These items were then pseudo-randomised such that identical image pairs could not appear one after the other.

Each trial began with a fixation cross presented centrally for 800 ms. Participants were instructed to keep their eyes on the central fixation point. Then a foveally presented word would appear for 200 ms. The bilaterally presented images were then presented for 150 ms. At stimulus offset, participants indicated whether the name of the image in the LVF or RVF rhymed with the foveally presented word. Participants were instructed to press "S" if the written word rhymed with the image in the LVF, and "K" if it rhymed with the image in the RVF. If participants did not detect a rhyme, they were instructed to press the space bar. The response triggered the next trial; note that the programme waited until a response occurred: trials did not time-out after a predetermined duration. Accuracy and response times (RT) were recorded.

### ***Dichotic listening task***

The consonant-vowel dichotic stimuli were taken from a dichotic listening task that was developed to work as a smartphone app, and shown to give high levels of test-retest reliability (Bless et al., 2013). These stimuli were the same as those used by Karlsson et al. (2019) and were used with the kind permission of Professor Kenneth Hugdahl.

**Materials.** The consonant-vowel stimuli were formed by pairing six stop-consonants (/b/, /d/, /g/, /p/, /t/, /k/) with the vowel /a/. Each of the six consonant-vowel stimuli (/ba/, /da/, /ga/, /pa/, /ta/, /ka/) were paired and played in each sound channel (e.g., /ba/-/da/), resulting in 36 pairings including homonyms.

**Procedure.** As the dichotic listening task demands the use of stereo-headphones, participants were screened at the start of each session. Screening consisted of two tasks. The first involved participants judging which of three pure tones was quietest, and is described by Woods, Siegel, Traer,



and McDermott (2017) in full detail. This sound discrimination task was designed to be easy to achieve when using headphones but difficult over loudspeakers due to phase-cancellation. Participants were removed from the study if they failed to correctly discriminate the quietest sound on at least five out of six trials, indicating that they were unlikely to be using headphones. The second check involved participants hearing a sound presented in only one of the two channels. Participants were removed if they failed to correctly identify which channel the sound was played on five out of six trials, indicating that they were not using stereo-headphones.<sup>1</sup>

Participants completed four blocks of the dichotic listening task. In each block, participants heard each of the 36 stimuli once. Across the four blocks, they responded to 144 stimuli. At the start of each trial, participants saw a fixation cross for 250 ms, after which a stimulus was played. After hearing each stimulus, participants were instructed to report the syllable they heard or if it seemed like they heard two different sounds, the one they heard most clearly by clicking a button corresponding to one of the six syllables. The response triggered the start of the next trial; trials did not time-out after a predetermined duration. Accuracy and response times (RT) were recorded.

### *Chimeric faces task*

The chimeric faces stimuli were kindly provided by Dr Michael Burt, and have been reported elsewhere (Burt & Perrett, 1997; Innes, Burt, Birch, & Hausmann, 2016; Karlsson et al., 2019).

**Materials.** The faces consisted of symmetrical averaged images created from four male and four female faces. Four emotional expressions were used: anger, disgust, happiness, and sadness. The faces were split vertically down the middle and paired so that one emotive hemiface was attached to another and blended at the midline (see Figure 2 for an example). These were paired in all possible combinations creating 16 individual stimuli. Four stimuli showed identical emotions in each hemiface.

**Procedure.** Participants completed four blocks of the chimeric faces task. In each block, participants viewed each stimulus twice. Thus, they viewed 32 stimuli per block and 128 in total. At the start of each trial, participants fixated a cross for 1000 ms. A chimeric face was then displayed for 400 ms.

---

<sup>1</sup>Hugdahl et al. (2009) recommend excluding participants with a general hearing impairment of more than 20 dB for all frequencies tested, and/or an interaural hearing threshold difference of more than 10 dB as this can influence dichotic listening outcomes. It is important to note that this exclusion was not possible in the current study as we did not include a hearing test. With the current sample size, subtle hearing deficits in one ear will be unlikely to undermine the results. However, it will be important for future studies to incorporate hearing accuracy tests in order to exclude participants with these hearing related difficulties.



**Figure 2.** An example stimulus from the chimeric faces task. In the example, “happiness” is shown on the left and “sadness” on the right. A participant who is right lateralized for emotional processing would be more likely to perceive the emotion in the left hemifield and would, therefore, respond “happiness”.

Participants then reported the emotion seen in the face by clicking a button corresponding to one of the four emotions. The response triggered the start of the next trial; trials did not time-out after a predetermined duration. Accuracy and response times (RT) were recorded.

### *Finger tapping task*

The finger tapping task is a novel behavioural task that we designed as a quantitative index of handedness analogous to a peg-moving task. Participants were required to either press the “W”, “R”, “V”, and “X” keys with the forefinger of their left hand or the “T”, “U”, “M”, “B” keys with their right hands in sequence as many times as possible within a 30-second window. Participants completed four blocks of the finger tapping task in total. Each block required participants to complete the task for both their dominant

and non-dominant hand. Each keypress was recorded along with its time-stamp relative to the beginning of the trial.

### ***General procedure***

The current study implemented a test-retest, within-subjects design across two sessions. Each session was three to 10 days apart. During session 1, participants completed the full battery of tasks. Self-report measures were completed first, followed by the behavioural laterality tasks. During session 2, participants completed only the behavioural laterality tasks. Hence, each participant provided data for the rhyme decision, dichotic listening, chimeric faces, and finger tapping tasks twice. The two sessions lasted approximately 45 and 35 min respectively.

When completing the behavioural tasks, participants were instructed to sit approximately 50 cm from the screen. Gorilla's scaling tool, which requires the participant to hold a credit card on the screen while adjusting size to match an image on the screen, was utilized such that stimuli size and eccentricity was maintained across tasks and participants. During each task participants were instructed to respond as fast and accurately as possible. Stimuli from each task were repeated across four blocks. Within each block, participants completed 70 trials of the rhyme detection task (R); 36 trials of the dichotic listening task (D); 32 trials of the chimeric face task (C); and 2 trials of the finger tapping task (F) for both their dominant and non-dominant hand. The order of tasks with each block was determined as follows RDCF, DCFR, CFRD, FRDC, and the order of tasks remained consistent across blocks for each participant. For each task, trial order was randomized, except for the finger tapping task where participants responded with their dominant hand followed by their non-dominant hand.

### ***Data analysis***

Out of 441 participants recruited, 411 (93.2%) completed both sessions. From the 411 participants who completed both sessions, 19 (4.6%) were removed as their lexTALE score was below 80, indicating that their level of vocabulary knowledge was below the cut off for lower advanced proficiency. This left a sample of 392 participants.

### ***Data cleaning***

For each task, we adopted the following data cleaning procedures. First, for all tasks, we removed trials in which response times were shorter than 200 ms. Response times were log-transformed to reduce skew, as specified in our pre-registration. Outlier response times were identified using Hoaglin and Iglewicz's (1987) procedure, whereby outliers are defined as

response times that are 1.65 times the difference between the first and third quartiles ( $1.65 \times (Q3 - Q1)$ ) below or above the first and third quartile values (e.g: lower limit =  $Q1 - 1.65 \times (Q3 - Q1)$ ; upper limit =  $Q3 + 1.65 \times (Q3 - Q1)$ ). This criterion was applied to all responses, before subdivision into left- and right-sided responses. Our differences criterion was set to 1.65 instead of the typical 2.2 because we expected to obtain asymmetric distributions for response time, with most outliers in one tail of the distribution. The Hoaglin and Iglewicz procedure was applied to each participant's data separately. After applying each data trimming procedure, participants left with less than 80% of trials were removed.

At the next stage, we excluded participants who failed to meet the pre-registered criteria for inclusion. Applying each of these procedures meant that the number of participants left in the final analyses could vary by task. For the chimeric faces task, we removed participants if they scored lower than 75% on trials in which an identical emotion was shown in each hemiface, and for the dichotic listening task, we removed those who scored less than 75% correct on trials where the same sound was played to each ear. These data trimming procedures left data for 304 participants on the dichotic listening task and 305 participants on the chimeric faces.

For the rhyme decision task, we removed participants who scored less than 50% accuracy (chance level) in either visual field. Furthermore, we removed participants who scored less than 75% accuracy on the catch trials (trials in which there was no rhyme). These trimming procedures left data for 270 participants on this task. We considered recruiting further participants to achieve our target sample size of 300 participants. This seemed uneconomic as we could already see that with close to 300 cases, the test-retest reliability of the rhyme detection task was unsatisfactory. With 270 participants, the test-retest reliability was  $r = .62$ , which is smaller than the  $r = 0.65$  that we pre-registered as our criterion for a reliable task. We plan to evaluate modified versions of this task in future but decided to stop data collection at this point and so we present here data from those participants with usable data.

For the finger tapping task, we added a data-cleaning step that had not been pre-registered, to remove participants where there were more than 20 keypresses across left and right hands that did not correspond to the intended sequence. This could reflect that the individual had forgotten the target sequence, or was not co-operating with instructions, but scrutiny of cases where this occurred suggested that it usually arose because of technical problems. For instance, some individuals had repeated recording of the same keypress at short intervals, suggestive of a sticky keyboard. The cut-off of 20 was arbitrarily selected and was successful in excluding cases with abnormally low scores in one session.

The final group consisted of 392 participants (230 female) who had complete data on at least one of the behavioural tasks. They had a mean age of 29.5 years ( $SD = 8.11$ ) and included 133 left-handers (75 female). 30 right-handers were bilingual and 13 left-handers were bilingual.

**Methods for calculating laterality indices, reliability, and population-level analysis.** We pre-registered that we would calculate LIs in two ways: (1) traditionally ( $100 \times (\text{Right} - \text{Left}) / (\text{Right} + \text{Left})$ ); (2) a standardized LI that involved computing a z-score that represented, for each individual, a comparison between left- and right-sided responses. In fact, for RT data we specified a *t*-score, but *t* and *z* converge when sample size exceeds 30, and for simplicity, we refer to these as LI<sub>z</sub> scores. For accuracy data, these LI<sub>z</sub> scores are very highly correlated with the traditional LI measures (correlations typically exceed .95) but can be used to provide a common metric for RT and accuracy measures when identifying significantly lateralized individuals. For example, using a conventional 2-sided alpha of .05, we can categorize individuals with an absolute LI<sub>z</sub> greater than 1.96 as significantly lateralized.

For the rhyme task, we pre-registered that we would calculate LIs based on log RT data, as accuracy during piloting was at ceiling. To obtain LI<sub>z</sub> we ran a *t*-test comparing accurate log RTs for left- and right-sided stimuli for each person. In this case, LI<sub>z</sub> will depend on an individual's RT variability as well as the difference in means between the two sides, and so the correlation with the conventional LI, though positive, will be less strong than for LI<sub>z</sub> scores based on accuracy.

For the dichotic listening and chimeric faces, we specified that LI calculations would be based on trials in which participants correctly identified a stimulus. The count of correct responses that corresponded to each side was used to generate an accuracy LI in two ways: (1) in the traditional fashion ( $100 \times (\text{Right} - \text{Left}) / (\text{Right} + \text{Left})$ ); (2) as a z-statistic for each participant, using  $z = (pR - .5) / \sqrt{pR \times pL / n}$ , where *pR* is the proportion of correct responses in the RVF, *pL* is the proportion of L responses in the LVF, and *n* is total LVF and RVF responses. This method can yield extreme z-scores (for instance, if 83/89 responses are to the right, then  $z = 15.92$ ), so scores were censored to be in the range  $-5$  to  $+5$ . Finally, LIs for the finger tapping task were calculated in the same manner, using the number of sequences generated by the left and right hand as opposed to the proportion of correct responses in each visual field.

To address hypothesis 1 and examine the laterality at the population level, we conducted a series of one-sample *t*-tests to compare mean LIs of the whole sample against zero for each task, using session 1 data.

For the reliability assessment, we first checked LIs for normality using a Shapiro-Wilk test. As most LIs were non-normal, we used a series of Spearman's correlations to examine the correlation of coefficients for session 1 versus session 2. We also conducted an analysis to determine the optimal test length. For tasks where the test-retest reliability was above .75, we considered how many trials were needed to achieve satisfactory reliability (i.e.,  $r > .65$ ). This was accomplished by recalculating LIs using 1 to  $n$  trials, where  $n$  is the total number of trials on a task. This was of interest in indicating the most efficient way to administer these tasks in future while retaining good reliability. However, the full data (including all trials) was used for all subsequent analyses.

***Handedness and atypical laterality.*** To address hypothesis 2, that extreme left-handers will be more likely to show atypical laterality than right-handers, we categorized participants as extreme left-handers if they scored less than  $-90$  on the EHI and right-handers if they scored greater than zero on the EHI. Following our pre-registration, atypicality was defined according to whether the participant's LI<sub>z</sub> was significantly lateralized (with  $\alpha = .05$ ) in the opposite direction from the majority. Chi-square tests were then used to examine the association between handedness and atypicality on the rhyme decision, dichotic listening, and chimeric faces tasks. In our pre-registration we had planned to look at cases who were atypical on the language tasks and the chimeric faces. We did not conduct this analysis because few participants showed atypical lateralization across multiple tasks. In our Supplementary Materials we do, however, compare laterality indices between extreme left-handers and right-handers to further examine the link between handedness and atypical lateralization.

***Relationship between language and non-language LIs.*** We examined the relationship between LIs on the rhyme decision, dichotic listening, and chimeric faces tasks by modelling covariances using a simplified version of Structural Equation Modeling (SEM) that used the sum of LIs from the two sessions of each task to represent each of three factors, minimizing the impact of measurement error. In this approach, we constrain particular covariance patterns and report "best" model fit according to AIC weights (Wagenmakers & Farrell, 2004). Akaike weights are transformations of the AIC values, which can be directly interpreted as conditional probabilities for each model, with the best-fitting model being selected. We compared three models: Model 1F, where all LIs are correlated, Model 2F, where the two language tasks are correlated with each other but not with the chimeric faces, and Model 3F, a base model where the LIs are unrelated. This analysis was conducted using data for both left- and right-handers.

**Table 1.** Count of left- and right-handers who report left or right on self-reported footedness, and eyedness (Miles test).

Foot-Eye	Right hander (N)	Left hander (N)	Right hander (%)	Left hander (%)
RR	141	21	54.4	15.8
RL	29	23	11.2	17.3
LR	51	41	19.7	30.8
LL	18	37	6.9	27.8
Other	20	11	7.7	8.3
Total	259	133	100	100

Note: RR: right foot, right eye; RL: right foot, left eye; LR: left foot, right eye; LL: left foot, left eye. "Other" includes those who responded "No preference / Don't know".

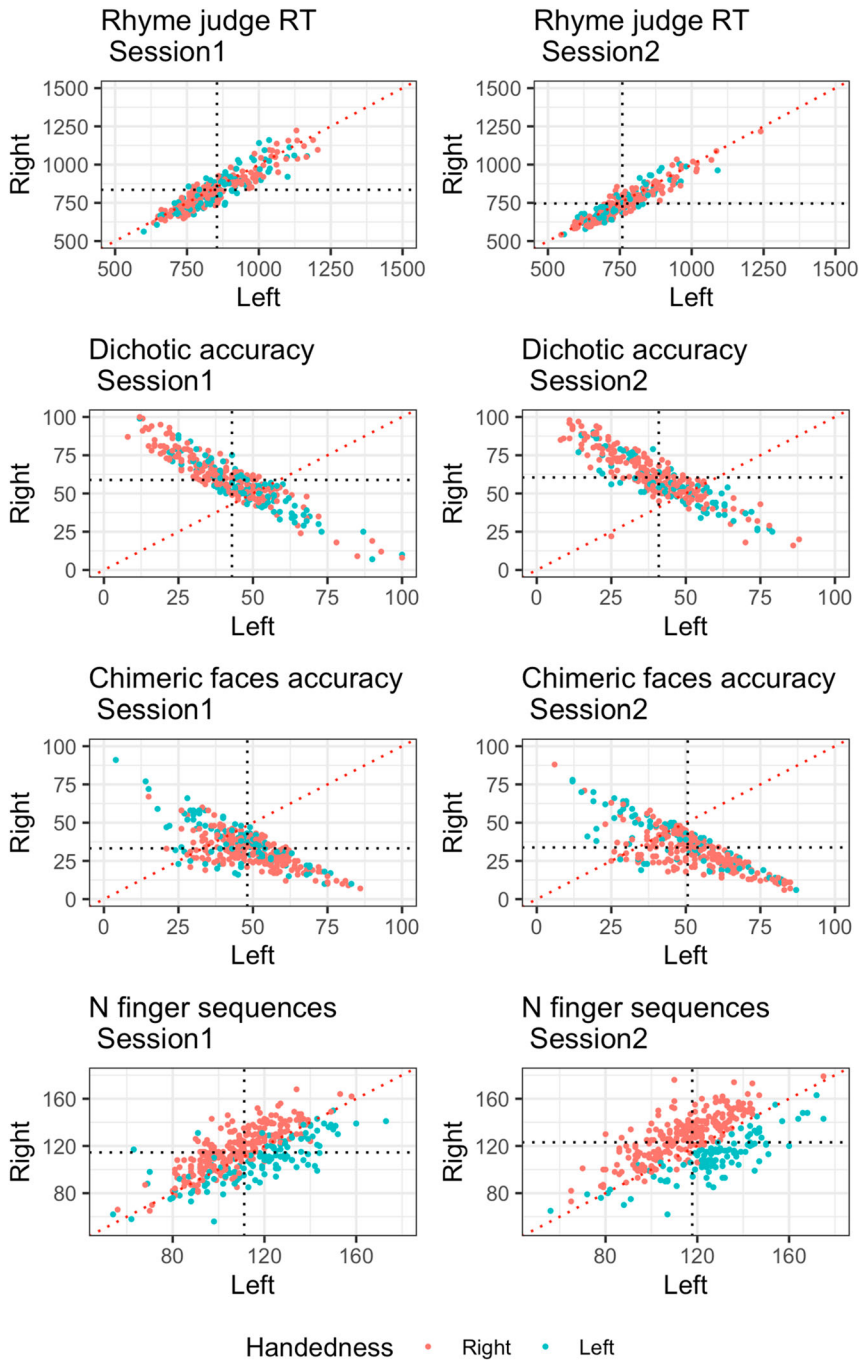
## Results

Descriptive data on foot and eye preference characteristics of the sample after exclusions as noted above are shown in Table 1. As these aspects of laterality are not a focus of this study, we will not elaborate further on these data.

### *The validity of each behavioural task*

Distributions of the raw data are shown in Figure 3. The scatterplots indicate scores on the left and right sides for each task, with self-reported handedness (left vs right) colour-coded. For each task, the mean for each side is shown as vertical and horizontal dotted lines, and equal performance on the two sides corresponds to the sloping red dotted line: thus an impression of whether the task is lateralized can be obtained by considering whether points cluster around the sloping lines symmetrically.

For the rhyme decision task, only correct responses were used for computing mean log RTs (shown transformed back to the original ms scale in the plot). Generally, accuracy on the rhyme decision task was high. For session 1, accuracy was 90.8% ( $SD = 12.5\%$ ) in the LVF and 91.2% ( $SD = 11.1\%$ ) in the RVF. For session 2, accuracy was 91.6% ( $SD = 12.8\%$ ) in the LVF and 91.9% ( $SD = 12.5\%$ ) in the RVF. For dichotic listening and chimeric faces, a correct response can be given from either side; it follows that there will be a negative correlation between the two sides reflecting the fact that the number of selections on the left will be inversely related to the number of correct responses on the right. For these tasks, it is evident by inspection that most responses for dichotic listening cluster above the dotted line, indicating a right ear advantage, whereas most responses on chimeric faces cluster below the dotted line, indicating a left hemi-face advantage. Finally, we show the number of correct sequences with each hand on the finger tapping task. Here, a strong impact of handedness is evident on inspection, with left-handers generally more accurate with the left hand, and right-handers more accurate with the right hand.



**Figure 3.** Left side (L) vs right side (R) scores for sessions 1 and 2 for the behavioural laterality tasks. The diagonal dotted line shows the point of equality: points above this are higher for R than L, and points below are higher for L than R. Mean values for R and L are shown as horizontal and vertical dotted lines. The colour of the points indicates the participants' handedness (pink = right-handed; blue = left-handed).



**Table 2.** Mean LIs and results from one-sample t-tests and Shapiro-Wilks tests for departure from normality for the rhyme decision (RD), dichotic listening (DL), chimeric faces (CF), and finger tapping (FT) tasks at sessions 1 and 2.

Task	N	Mean	SD	One sample <i>v</i>	One sample <i>p</i>	Normality <i>p</i>
RD 1	270	0.18	0.46	25,830.0	<0.001	0.459
RD 2	255	0.14	0.40	22,600.0	<0.001	0.237
DL 1	305	15.96	32.14	36,792.5	<0.001	<0.001
DL 2	287	19.53	32.04	33,568.0	<0.001	<0.001
CF 1	342	-17.97	26.79	8944.5	<0.001	0.001
CF 2	322	-19.08	30.78	8555.5	<0.001	0.001
FT 1	330	1.36	6.90	31,802.5	<0.001	0.015
FT 2	309	2.22	7.96	30,534.5	<0.001	0.001

Note: LIs are scored such that a positive value equates to left lateralization at the brain level.

The distribution of LIs can be used to quantify departures from symmetry. Recall that for all tasks, we computed LIs in a typical way:  $(100 \times (\text{Right} - \text{Left}) / (\text{Right} + \text{Left}))$ , and that because of crossed pathways from hands, ears and eyes to the brain, a positive LI indicates left lateralization and a negative LI indicates right lateralization at the brain level. LIs on the rhyme task were calculated using log RTs for correct responses. LIs for the dichotic listening and chimeric faces were calculated using the count of correct responses for sounds/emotions in each ear/visual. LIs on the finger tapping task were calculated using the number of sequences generated for each hand. [Table 2](#) shows these mean LIs, as well as a one-sample Wilcoxon test comparing the mean LI with zero. Wilcoxon tests were chosen as a series of Shapiro-Wilks tests showed a departure from normality for several tasks. It can be seen that all measures were significantly lateralized in the predicted direction at the population level, with left-hemisphere superiority for rhyme decision, dichotic listening and finger tapping, and right-hemisphere superiority for chimeric faces.

[Table 3](#) shows data from the LI<sub>z</sub> measures, which make it easier to compare different tasks on the same scale. The distributions of LI<sub>z</sub> scores are shown graphically in the supplementary materials. For each task, those with LI<sub>z</sub> scores greater than 1.96 are categorized as left-hemisphere dominant, and those with LI<sub>z</sub> scores less than -1.96 are categorized as right-hemisphere dominant. If there were no population bias, we would expect around 2.5% of individuals to meet each of these criteria. It can be seen that for dichotic listening and chimeric faces, around 47% of individuals have hemispheric dominance in the predicted direction and around 7% have hemispheric dominance in the opposite direction. The proportions showing significant laterality are lower for the rhyme decision task (around 19% left-hemisphere dominant, and 4% right-hemisphere dominant), and for finger tapping, the proportions are only slightly greater than would be expected by chance, although there is a suggestion of increasing laterality on that task across the two sessions. Note, however, that we had oversampled

**Table 3.** Mean LI\_z and percentage of participants who were significantly left lateralized ( $z > 1.96$ ) or right lateralized ( $z < -1.96$ ) on the rhyme decision (RD), dichotic listening (DL), chimeric faces (CF), and finger tapping (FT) tasks at sessions 1 and 2.

Task	N	Mean	SD	L-lateralised %	R-lateralised %
RD 1	270	0.63	1.50	19.3	4.4
RD 2	254	0.62	1.51	16.1	6.7
DL 1	304	1.64	6.12	48.7	9.5
DL 2	287	2.23	6.29	47.0	9.1
CF 1	305	-1.91	3.45	7.9	45.6
CF 2	287	-2.42	4.80	9.4	48.4
FT 1	330	0.21	1.03	3.9	2.4
FT 2	309	0.35	1.24	6.5	3.2

Note: LIs are scored such that a positive value equates to left lateralization at the brain level.

**Table 4.** Test-retest reliability estimates and 95% CIs for the rhyme decision (RD), dichotic listening (DL), chimeric faces (CF), and finger tapping (FT) tasks.

Task	Spearman r	95% CI
RD	0.615	0.52–0.69
DL	0.782	0.71–0.84
CF	0.877	0.84–0.90
FT	0.755	0.68–0.80

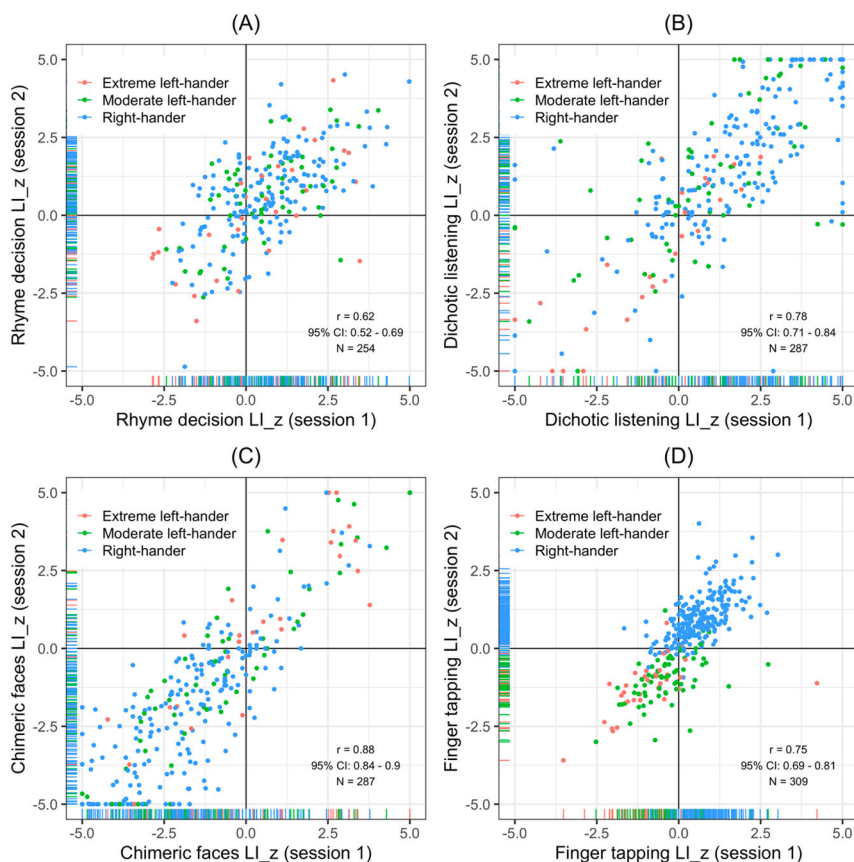
left-handers for this study, and, as is evident from exploratory analyses presented below, finger-tapping was strongly influenced by handedness.

### Test-retest reliability

The scatterplots depicting test-retest reliability are shown in [Figure 4](#), with Spearman test-retest correlations in [Table 4](#). These were based on LI\_z values, but are essentially identical to values from unscaled LI values. Confidence intervals were estimated using the *spearman.ci()* function from the RVAideMemoire package (version 0.9-75; Hervé, 2020) with 10,000 iterations. For all but the rhyme decision task, we obtained good test-retest reliability that exceeded our pre-registered threshold of  $r = .65$ .

### Optimizing the tasks

The dichotic listening and chimeric faces tasks had test-retest reliability above  $r = .75$ , and so we investigated whether the tasks could be made more efficient by reducing the number of trials while retaining good test-retest reliability. Hence, we recalculated LIs for the dichotic listening and chimeric faces tasks using 1 to n trials, where n is the total number of trials on each task. The absolute test-retest correlation coefficients are shown in [Figure 5](#). Peak reliability was reached in around 85 trials with dichotic listening, and a minimum of 60 trials was required to achieve a test-retest correlation of

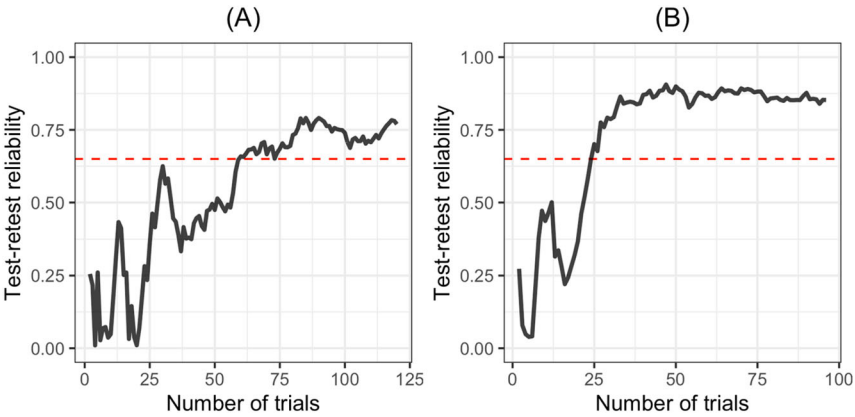


**Figure 4.** Scatter plots for session 1 and session 2 data for LI<sub>z</sub> scores on each laterality task: (A) rhyme decision, (B) dichotic listening, (C) chimeric faces, and (D) Finger tapping. Handedness is colour coded with extreme left-handers identified as those who have a score below  $-90$  on the Edinburgh Handedness Inventory. Note that z-scores are censored at absolute values of 5.

$r = .65$ . For the chimeric faces task, peak reliability was achieved with around 36 trials, and a minimum of 25 trials was necessary to achieve test-retest reliability of  $r = .65$ .

### **Atypical laterality and handedness**

According to Mazoyer et al. (2014), extreme left-handers are more likely than other handedness groups to have reversed laterality compared to the rest of the population. Our pre-registered analysis accordingly focused on comparing extreme left-handers (those with a score of  $-90$  or below on the Edinburgh Handedness Inventory) with right-handers. We defined “atypical laterality” as laterality that is in the opposite direction to the majority of



**Figure 5.** Test-retest correlations for (A) the dichotic listening task and (B) the chimeric faces task when using a variable number of trials to calculate laterality indices. The red dashed line marks the test-retest correlation  $r = .65$ . The correlations coefficients are shown as absolute values.

**Table 5.** Frequencies of laterality typicality for extreme left-handers, moderate left-handers and right-handers across the rhyme decision (RD), dichotic listening (DL), and chimeric faces (CF) tasks.

Task	Sub-group	Left N	Non N	Right N	Left %	Non %	Right %
RD	Extreme left-hander	9	21	6	25.0	58.3	16.7
	Moderate left-hander	14	44	2	23.3	73.3	3.3
	Right-hander	29	141	4	16.7	81.0	2.3
DL	Extreme left-hander	13	21	7	31.7	51.2	17.1
	Moderate left-hander	27	37	11	36.0	49.3	14.7
	Right-hander	118	93	14	52.4	41.3	6.2
CF	Extreme left-hander	10	19	10	25.6	48.7	25.6
	Moderate left-hander	10	33	30	13.7	45.2	41.1
	Right-hander	7	108	116	3.0	46.8	50.2

Note: N is the count of participants. % is the percentage of participants.

participants, with an absolute  $LI_z$  value greater than 1.96. Although the focus of our planned analyses included only right-handers and extreme left-handers, we show data for the remaining left-handers for completeness in Table 5.

The rhyme decision task has relatively few individuals who are significantly lateralized in either direction. We will therefore not consider it further, as our analyses are only justified if the task is an indicator of cerebral lateralization. Both dichotic listening and chimeric faces, on the other hand, have a high proportion of strongly lateralized individuals. The planned analyses, comparing typically lateralized with atypically lateralized individuals, and extreme left vs right-handed individuals, show a significant association between extreme left-handedness and atypical laterality on a chi-square test. For dichotic listening, chi-square = 6.75, d.f. = 1,  $p = 0.019$ ; for chimeric faces,

**Table 6.** Spearman correlations between the LI\_z scores for the different tasks in two Sessions.

	RD 1	RD 2	DL 1	DL 2	CF 1	CF 2
RD 1	1					
RD 2	0.629	1				
DL 1	0.145	0.155	1			
DL 2	0.185	0.200	0.851	1		
CF 1	−0.057	−0.030	−0.039	−0.007	1	
CF 2	−0.059	−0.025	−0.055	−0.017	0.879	1

chi-square = 28.15, d.f. = 1,  $p < .001$ . These data provide statistical support for the prediction from Mazoyer et al. (2014) that extreme left-handers have an increased rate of atypical cerebral lateralization.

The planned analysis looking at cases who were atypical on both dichotic listening and chimeric faces was hampered by the fact that the two tasks were not associated (see below), and consequently, there were only four cases in this category. Nevertheless, it was noteworthy that two of these were extreme left-handers and the other two were less extreme left-handers.

### *Relationship between LIs from different tasks*

Our final question was whether the laterality indices from the language and faces laterality tasks correlated. While we had originally planned to examine the relationship between LI\_z score from different measures separately for left- and right-handers, we report these analyses for a combined group of all participants who did not have missing data (69 left- and 146 right-handers). Our departure from pre-registered analysis followed reviewers concerns about limited variation and restricted ranges of LI\_z scores when looking at intercorrelations within handedness subgroups. Furthermore, examining the relationship between LI\_z scores for the sample as a whole protects against the instability of results from our planned modelling of covariances, which with small sample sizes could favour different models when quite minor changes were made to the sample.

Initial scrutiny of Spearman correlations between LI\_z scores suggests weak associations between LI\_z values for the two language measures (rhyme decision and dichotic listening), but no association between either of these measures and LI\_z on chimeric faces. Note, however, that correlations will be influenced by test reliability; disattenuated correlations (divided by the square root of the product of the reliabilities of the two measures) will be higher (Muchinsky, 1996) (Table 6).

In a more formal test, we compared three models: (2F) where the two language tasks are correlated with each other but not with the chimeric faces (i.e., a language laterality factor); (1F) all LIs are correlated (i.e., a

single laterality factor); and (3F) a base model where the LIs are unrelated (i.e., independent functions). We departed from our pre-registered analysis based on reviewer recommendations. First, we show results for the full sample rather than separately for left- and right-handers. Second, we modified the Akaike weights analysis to remove the effect of measurement error by using factor scores based on the summed scores computed separately for odd and even trials (i.e., split-half) from the two test sessions. The analysis is thus conducted on factor scores with four indicators for each factor.

The Akaike weight for the 1F, 2F, and 3F models were 0.103, 0.610, and 0.287 respectively. Thus the Akaike weights favoured model 2F confirming that there is association for laterality indices for the two language tasks, but not between language laterality and laterality for face processing of emotion.

For completeness, Supplementary Material Table S1 shows the original pre-registered Akaike weights analysis, which did not show a benefit for the two-factor model, but rather gave best fit to a model where there were distinct factors for each laterality test. We regard this model as misleading because it incorporates measurement error, which was particularly large for the rhyme detection task, and so attenuates the relationship between variables.

## ***Exploratory analyses***

### ***Finger tapping and Edinburgh handedness inventory***

To examine the relationship between the finger tapping task and a conventional hand preference battery, we computed Spearman's rank correlation between LI\_z scores produced on the finger tapping task and the Edinburgh Handedness Inventory. For session 1 data, the relationship between the tasks was moderate,  $r = 0.53$ , 95% CI [0.44, 0.61],  $p < 0.001$ . For session 2, the correlation was  $r = 0.63$ , 95% CI [0.55, 0.70],  $p < 0.001$ . This suggests that the influence of hand preference on the tapping task may increase with practice.

### ***Vocabulary knowledge and laterality***

The LexTALE vocabulary measure had been included largely for data exclusion, but we had planned to explore correlations between LexTALE score and the laterality index on the rhyme decision task. However, as the rhyme task was less reliable than other measures, and showed only small asymmetries, we chose to examine the relationship between the LexTALE and the other verbal laterality measure, dichotic listening, instead. For these exploratory analyses we included participants scoring less than 80 on the LexTALE to reduce the effect of a restricted range on the correlational analyses. There was no correlation between the LexTALE score and LIs on the dichotic listening at

session 1,  $r = -0.05$ , 95%  $CI = [-0.16, 0.06]$ ,  $p = 0.681$ , or session 2,  $r = -0.05$ , 95%  $CI = [-0.17, 0.07]$ ,  $p = 0.830$ .

## Discussion

With the advancement of online behaviour research methods, it has become possible to test large samples of participants outside the laboratory. In the current study, we introduced an online behavioural laterality battery that was administered to a relatively large sample of right- and left-handers. In addition to completing self-report measures of handedness, footedness, and sightedness, participants completed a visual half-field rhyme decision task, a dichotic listening task, a chimeric faces task, and a finger tapping task twice. Thus, we were able to examine the test-retest reliability of each task. We also examined the association between handedness and atypical lateralization and examined the intercorrelations between LIs on the dichotic listening, rhyme judgement and chimeric faces tasks.

Each behavioural task yielded a hemispheric advantage in the expected direction, but the size of asymmetries and the test-retest reliability of the different tasks varied. A relatively small proportion of individuals showed significant asymmetry on the rhyme decision task, despite the population as a whole being significantly left lateralized. This pattern of results indicates that while very few participants showed statistically better performance for images in the right visual field (i.e., LI\_z scores above  $|1.96|$ ), many showed more subtle RVF advantages that did not exceed the threshold of statistical significance. When comparing LI\_z scores to zero at the population level there will have been sufficient power to determine whether these small differences were reliably different from zero. This task also yielded the poorest test-retest reliability, just falling short of our target minimum cutoff of  $r = .65$ . Compared to early studies (e.g., Krach et al., 2006) the rhyme decision task used here was more reliable but, compared to other visual half-field tasks, it was far from optimal. Van der Haegen and Brysbaert (2018) have reported that, under lab conditions, visual half-field paradigms can yield test-retest reliabilities of  $r = .77$  when using pictures and  $r = .83$  when using words. Our study was well-powered, so the relatively poor task validity and reliability are likely to be due to methodological factors.

One key difference between van der Haegen and Brysbaert's and our study was that their language-based visual half-field tasks involved overt naming. Speech production is highly lateralized (e.g., Knecht et al., 2000), and requiring speech production should yield a hemispheric advantage. Our task was designed to require covert speech, raising the question of whether an active speech production component might reveal stronger lateralization. However, this explanation seems unlikely, given that recent studies have reported robust right visual-field advantages on a lexical decision task that

does not require active speech production (Brederoo et al., 2020; Hausmann et al., 2019; Willemin et al., 2016).

An alternative explanation for the poor validity and reliability of the rhyme task relates to the control viewing conditions. The visual half-field method depends critically on presenting lateralized visual stimuli that project to one hemisphere. Although we were able to calibrate screen size at the start of the session and to request participants to sit a certain distance from the screen, it is plausible this was not sufficient to achieve adequately lateralized presentation. While we can format images to be consistent across displays, we cannot control participants' behaviour. Outside of the lab, participants can freely move their head which will influence the size of images on the retina, and it is difficult to ensure that images are shown exclusively in parafoveal vision. If a subset of images are not equally displaced in parafoveal vision it could mean that processing is more efficient for the image in its corresponding visual field, reducing the validity of the task and adding considerable noise to response time estimates in each visual field. Without strict monitoring of eye-movements here it is not possible to control for this. In future work, we plan to explore the use of images presented in peripheral vision, as it is unlikely that small head movements would result in the image shifting from peripheral vision. This would increase difficulty in the processing of information in this region, which may also result in stronger visual field advantages.

In contrast to rhyme decision, the other verbal task, dichotic listening, showed excellent validity and reliability. Participants showed a clear right-ear advantage at the population level, and the task exceeded our pre-registered reliability cut-off of  $r = .65$ . Indeed, the test-retest correlation of the dichotic listening task,  $r = .78$ , was identical to that reported by Bless et al. (2013) using the iDichotic application, and numerically higher than that reported for the same dichotic listening task when run under lab conditions (Bless et al., 2013). We further showed (see Figure 5) that there was little decline in reliability with 85 trials, and with as few as 60 trials, the current implementation of the dichotic listening task yielded a test-retest correlation higher than  $r = .65$ , suggesting that those who wish to use this paradigm in the future could use a smaller number of trials. It would be of interest to extend this line of work to use the modified version of dichotic listening described by Westerhausen and Samuelsen (2020), which had even higher reliability. Of course, it is important to screen for adequate use of stereo headphones when devising online versions of the task.

Like the dichotic listening task, the chimeric faces task showed impressive levels of both validity and reliability, this time with a strong right hemisphere bias at the population level. The task has previously been shown to have high split-half reliability (Levy et al., 1983), we show that it can meet the higher



standard of achieving a high test-retest correlation. Our subsequent analysis showed that acceptable levels of reliability can be achieved with this task using very few trials: reliability did not suffer when only 36 trials were used, and test-retest reliability of  $r = .65$  could be found with 25 trials. It seems likely that one reason for the success of the chimeric faces task online is the task's simplistic nature. From the perspective of the participant, it is a straightforward task of reporting the emotion expressed in a single face and does not require any explicit comparison. The finding that good test-retest reliability could be observed with a reduced number of trials will also be welcome for future studies, where there are constraints on testing time.

We also introduced a novel finger tapping task, which we designed to quantify handedness behaviourally. This produced a right-hand advantage at the population level, and test-retest reliability of the LIs was good. The asymmetries were small in absolute terms, but this in part reflected the high proportion of the left-handers in the sample: our exploratory analysis indicated that LIs from this task correlated reasonably well with the Edinburgh Handedness Inventory (see also [Figure 3](#)). Previous studies have reported correlations between hand preference and measures of relative hand skill, but different hand skill measures typically are weakly intercorrelated (Bishop, 1990), and it seems that asymmetries may vary depending on particular aspects of planning and execution of motor coordination that is involved. It would be of interest to compare LIs on this task with those from the pegboard task (e.g., Hodgson, Hirst, & Hudson, 2016), which is widely used in neuropsychological assessment. Future work should also examine how other factors such as the complexity of key combinations and reach across the midline influence the task.

In addition to assessing lateralization of the task using traditional LIs, we also examined the potential to use individual z-scores to quantify lateralization on each task. This method of calculating LIs is perhaps more useful for those interested in looking at atypical lateralization, or for cases where there is a need to compare laterality from different tasks on a common scale. Rather than selecting those who fall either side of an arbitrary cut-off, the LI\_z score allows researchers to assess whether participants are consistently and statistically lateralized in either a typical or atypical manner. The availability of an online measure means that a researcher who is interested in studying the correlates of atypical language lateralization could assess participants with our battery to identify those with LI\_z scores  $> 1.96$  or  $< -1.96$  before inviting them into the lab for more detailed in-person testing. This has the potential to save time and effort by screening participants to identify a sample that is enriched for rare cases of reversed cerebral lateralization.

As well as validating the battery and establishing its test-retest reliability, we addressed two theoretically motivated questions related to handedness and atypicality and intercorrelations between the dichotic listening and chimeric faces task. In our pre-registered analysis of handedness, we categorized participants into right-, left-, and extreme left-handers. The data are consistent in showing that extreme left-handers showed increased atypicality on both the dichotic listening and chimeric faces task. Mazoyer et al. (2014) reported a relationship between the strength of lateralization and hand preference, which was largely driven by the existence of a group with strong left-hand preference and strong right hemisphere dominance for language. They further claimed that this right hemisphere language dominance may be exclusively present in left-handers. Although strong left-handers were over-represented in this category in our data, we did also observe cases of right-handers with right language dominance on the dichotic listening task. This raises the question of whether right language dominance on dichotic listening is equivalent to right language dominance on more direct measures of brain lateralization, such as fMRI.

A common mechanism that co-determines lateralization on a range of skills, extending beyond handedness and language, has been a hotly debated topic in the laterality literature. Several researchers have postulated complementary lateralization for left hemisphere and right hemisphere brain asymmetries, predicting that LIs on verbal and nonverbal tasks should be inversely correlated (Behrmann & Plaut, 2015; Hellige, 1993; Kosslyn, 1987). On this view, a person with atypical right hemisphere language representation should have atypical left hemisphere representation for nonverbal domains such as face processing. However, several studies have reported no such association, or only a weak relationship (Badzakova-Trajkov et al., 2010; Bryden et al., 1983; Rosch et al., 2012; Van der Haegen & Brysbaert, 2018; Whitehouse & Bishop, 2009). Data from the current work supports a view in which laterality on verbal and nonverbal tasks is independent. Not only was our study well powered to examine this relationship, our sample included a large number of left-handers who showed increased variability in their LIs. Specifically, we found no association between the LI<sub>z</sub> indices for the two language tasks and chimeric faces. There was a weak but measurable association between the two language tasks, rhyme decision and dichotic listening. This is of particular interest given the different demands of the two tasks: speech perception using dichotic presentation for the dichotic listening task, vs covert naming of a picture and comparison to the phonology of a written word in the rhyme judgement task. The current result suggests there could be interest in replicating this comparison of the two tasks using a visual-half-field language task that has better reliability and validity.

The lack of association between dichotic listening and chimeric faces is convincing, given that both tasks show strong lateralization at the population level and high test-retest reliability. This is consistent with the view that, at least for some functions, lateralization is determined by independent biases: the left-hemisphere population bias for language is unrelated to the right-hemisphere bias for facial expressions of emotion. Given the range of evidence now available, it may be time to abandon hypotheses that predict that different lateralized skills are either wholly interdependent or wholly independent. Evidence such as that from Gerrits et al. (2020) confirms that lateralization of different functions is associated far more than expected by chance, yet cases of dissociation are not uncommon.

This conclusion is supported by recent work from Brederoo et al. (2020), who reported data for 122 participants. In their study, participants completed a variety of visual half-field tasks covering a range of domains such as word processing, face processing, and spatial frequency processing. At the population level, Brederoo et al. reported lateralization in the expected direction for the majority of tasks. For left-handers and those with right hemisphere dominant speech, there was evidence of a deviation from typical lateralization. Contrary to the current study, Brederoo et al. reported partial evidence to suggest that processes recruiting similar resources will lateralize to homologue areas to promote intra-hemispheric proximity to other cortical areas within their processing hierarchies. In one analysis they found the lateralization of faces and word recognition to be correlated but when they looked at lateralization in those who were right-handed versus those with right dominant language processing there was considerably less evidence for causal complementarity. However, Brederoo et al. reported consistent evidence within the visual domain, where lateralization of low- and high-frequency information processing was associated, but independent of lateralization on other tasks. So it would appear by comparing tasks of a wider variety and considering laterality across domains, we may elucidate the issue of colateralisation and identify how far individual differences in lateralization depend on task demands and shared processing resources. Complementarity may be specific to functions that engage homologous areas in the left and right hemispheres. As such, research in future might explore which functions are associated, and to what extent, rather than making a simple contrast between models that postulate a general answer to explain all of laterality.

To conclude, the present study introduced an online behaviour laterality battery that can be used to examine laterality for speech perception, facial emotion perception, and manual preference with good test-retest reliability. The battery of tasks was administered to a large sample of participants with relative ease and enabled us to examine several questions related to lateralization. We find evidence of increased atypical lateralization in extreme left-handers relative to right-handers. However, we find no support for a

unitary model of cerebral lateralization that postulates simple complementarity between left- and right-sided brain lateralization in individuals.

## Acknowledgements

First, we would like to thank Mike Burt for providing stimuli for the chimeric faces task and Kenneth Hugdahl for providing stimuli for the dichotic listening task. Second, we would like to thank Marco Hirnstein, Marc Brysbaert, and Sanne Brederoo for their helpful comments on the original manuscript of this paper. Third, we would like to thank Rogier Kievit for his advice regarding the modelling of covariances. Portions of this work were presented at the Online Meeting of the Experimental Psychology Society in July 2020.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This study was supported by an Advanced Grant from the European Research Council [grant number 694189].

## Data accessibility

This project was pre-registered on the Open Science Framework (OSF) before the start of data collection. The registration form along with all task materials, analysis scripts and anonymised data can be found on the OSF at <https://osf.io/qsxcv/>. Working versions of the dichotic listening, chimeric faces, and finger tapping tasks are available on Gorilla Open Materials: <https://gorilla.sc/openmaterials/104636>.

## ORCID

Adam J. Parker  <http://orcid.org/0000-0002-1367-2282>

Zoe V. J. Woodhead  <http://orcid.org/0000-0003-0462-6791>

Paul A. Thompson  <http://orcid.org/0000-0001-9940-6913>

Dorothy V. M. Bishop  <http://orcid.org/0000-0002-2448-4033>

## References

- Annett, M. (1970). The growth of manual preference and speed. *British Journal of Psychology*, 61, 545–558. doi:[10.1111/j.2044-8295.1970.tb01274.x](https://doi.org/10.1111/j.2044-8295.1970.tb01274.x)
- Annett, M. (2002). *Handedness and brain asymmetry; The right shift theory*. Hove: Psychology Press.

- Anwyll-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52, 388–407.
- Badzakova-Trajkov, G., Häberling, I. S., Roberts, R. P., & Corballis, M. C. (2010). Cerebral asymmetries: Complementary and independent processes. *PLoS One*, 5, e9682.
- Beaumont, J. G. (1982). Studies with verbal stimuli. In J. G. Beaumont (Ed.), *Divided visual field studies of cerebral organisation* (pp. 58–86). London: Academic Press.
- Behrmann, M., & Plaut, D. C. (2015). A vision of graded hemispheric specialization. *Annals of the New York Academy of Sciences*, 1359, 30–46.
- Bethmann, A., Tempelmann, C., De Bleser, R., Scheich, H., & Brechmann, A. (2007). Determining language laterality by fMRI and dichotic listening. *Brain Research*, 1133, 145–157.
- Bishop, D. (1990). *Handedness and developmental disorder*. Hove: Erlbaum.
- Bless, J., Westerhausen, R., Arciuli, J., Kompus, K., Gudmundsen, M., & Hugdahl, K. (2013). Right on all occasions? On the feasibility of laterality research using a smart-phone dichotic listening application. *Frontiers in Psychology*, 4, 42.
- Bless, J. J., Westerhausen, R., Torkildsen, J. V. K., Gudmundsen, M., Kompus, K., & Hugdahl, K. (2015). Laterality across languages: Results from a global dichotic listening study using a smartphone application. *Laterality: Asymmetries of Body, Brain and Cognition*, 20, 434–452.
- Boles, D. B. (1987). Reaction time asymmetry through bilateral versus unilateral stimulus presentation. *Brain and Cognition*, 6, 321–333.
- Boles, D. B. (1990). What bilateral displays do. *Brain and Cognition*, 12, 205–228.
- Boles, D. B. (1994). An experimental comparison of stimulus type, display type and input variable contributions to visual field asymmetry. *Brain and Cognition*, 24, 184–197.
- Bourne, V. J. (2006). The divided visual field paradigm: Methodological considerations. *Laterality: Asymmetries of Body, Brain and Cognition*, 11, 373–393.
- Brederoo, S. G., Nieuwenstein, M. R., Cornelissen, F. W., & Lorist, M. M. (2019). Reproducibility of visual-field asymmetries: Nine replication studies investigating lateralization of visual information processing. *Cortex*, 111, 100–126.
- Brederoo, S. G., Van der Haegen, L., Brysbaert, M., Nieuwenstein, M. R., Cornelissen, F. W., & Lorist, M. M. (2020). Towards a unified understanding of lateralized vision: A large-scale study investigating principles governing patterns of lateralization using a heterogeneous sample. *Cortex*. doi:10.1016/j.cortex.2020.08.029
- Bryden, M. P. (1970). Laterality effects in dichotic listening: Relations with handedness and reading ability in children. *Neuropsychologia*, 8, 443–450.
- Bryden, M. P. (1988). An overview of the dichotic listening procedure and its relation to cerebral organization. In K. Hugdahl (Ed.), *Handbook of dichotic listening: Theory, methods and research* (pp. 1–43). New York: Wiley.
- Bryden, M. P., Hécaen, H., & DeAgostini, M. (1983). Patterns of cerebral organization. *Brain and Language*, 20, 249–262.
- Burt, D. M., & Perrett, D. I. (1997). Perceptual asymmetries in judgements of facial attractiveness, age, gender, speech and expression. *Neuropsychologia*, 35, 685–693.
- Carey, D. P., & Johnstone, L. T. (2014). Quantifying cerebral asymmetries for language in dextrals and adextrals with random-effects meta analysis. *Frontiers in Psychology*, 5, 1128.
- Dundas, E. M., Plaut, D. C., & Behrmann, M. (2015). Variable left-hemisphere language and orthographic lateralization reduces right-hemisphere face lateralization. *Journal of Cognitive Neuroscience*, 27, 913–925.

- Elias, L. J., Bulman-Fleming, M. B., & McManus, I. C. (1999). Visual temporal asymmetries are related to asymmetries in linguistic perception. *Neuropsychologia*, 37, 1243–1249.
- Flowers, K. A., & Hudson, J. M. (2013). Motor laterality as an indicator of speech laterality. *Neuropsychology*, 27, 256–265.
- Geffen, G., Bradshaw, J. L., & Nettleton, N. C. (1972). Hemispheric asymmetry: Verbal and spatial encoding of visual stimuli. *Journal of Experimental Psychology*, 95, 25–31.
- Gerrits, R., Verhelst, H., & Vingerhoets, G. (2020). Mirrored brain organization: Statistical anomaly or reversal of hemispheric functional segregation bias? *Proceedings of the National Academy of Sciences*, 117, 14057–14065.
- Gilbert, C., & Bakan, P. (1973). Visual asymmetry in perception of faces. *Neuropsychologia*, 11, 355–362.
- Hausmann, M., Brysbaert, M., van der Haegen, L. J., Specht, K., Hirnstein, M., Willemin, J., ... Mohr, C. (2019). Language lateralisation measured across linguistic and national boundaries. *Cortex*, 111, 134–147. doi:10.1016/j.cortex.2018.10.020
- Hellige, J. B. (1993). *Hemispheric asymmetry: What's right and what's left. perspectives in cognitive neuroscience*. Cambridge, MA: Harvard University Press.
- Hervé, M. (2020). *Rvaidemicrosoftware: Testing and plotting procedures for biostatistics* [R package version 0.9-75]. Retrieved from <https://CRAN.R-project.org/package=RVAideMemoire>
- Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 82, 1147–1149.
- Hodgson, J. C., Hirst, R. J., & Hudson, J. M. (2016). Hemispheric speech lateralisation in the developing brain is related to motor praxis ability. *Developmental Cognitive Neuroscience*, 22, 9–17.
- Hodgson, J. C., & Hudson, J. M. (2018). Speech lateralization and motor control. *Progress in Brain Research*, 238, 145–178.
- Hugdahl, K. (2000). Lateralization of cognitive processes in the brain. *Acta Psychologica*, 105, 211–235.
- Hugdahl, K. (2003). Dichotic listening in the study of auditory laterality. In K. Hugdahl, & R. J. Davidson (Eds.), *The asymmetrical brain* (pp. 441–476). Cambridge, MA: MIT Press.
- Hugdahl, K., Westerhausen, R., Alho, K., Medvedev, S., Laine, M., & Hämäläinen, H. (2009). Attention and cognitive control: Unfolding the dichotic listening story. *Scandinavian Journal of Psychology*, 50, 11–22.
- Hunter, Z. R., & Brysbaert, M. (2008). Visual half-field experiments are a good measure of cerebral language dominance if used properly: Evidence from fMRI. *Neuropsychologia*, 46, 316–325.
- Iacoboni, M., & Zaidel, E. (1996). Hemispheric independence in word recognition: Evidence from unilateral and bilateral presentations. *Brain and Language*, 53, 121–140.
- Innes, B. R., Burt, D. M., Birch, Y. K., & Hausmann, M. (2016). A leftward bias however you look at it: Revisiting the emotional chimeric face task as a tool for measuring emotion lateralization. *Laterality: Asymmetries of Body, Brain and Cognition*, 21, 643–661.
- Karlsson, E. M., Johnstone, L. T., & Carey, D. P. (2019). The depth and breadth of multiple perceptual asymmetries in right handers and non-right handers. *Laterality: Asymmetries of Body, Brain and Cognition*, 24, 707–739.
- Kelley, K. S., & Littenberg, B. (2019). Dichotic listening test–retest reliability in children. *Journal of Speech, Language, and Hearing Research*, 62, 169–176.

- Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex*, 3, 163–178.
- Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., ... Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain*, 123, 2512–2518.
- Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, 94, 148–175.
- Krach, S., Chen, L. M., & Hartje, W. (2006). Comparison between visual halffield performance and cerebral blood flow changes as indicators of language dominance. *Laterality: Asymmetries of Body, Brain and Cognition*, 11, 122–140.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44, 325–343.
- Levy, J., Heller, W., Banich, M. T., & Burton, L. A. (1983). Asymmetry of perception in free viewing of chimeric faces. *Brain and Cognition*, 2, 404–419.
- Mazoyer, B., Zago, L., Jobard, G., Crivello, F., Joliot, M., Percey, G., ... Tzourio-Mazoyer, N. (2014). Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. *PLoS one*, 9, e101165.
- Miles, W. R. (1929). Ocular dominance demonstrated by unconscious sighting. *Journal of Experimental Psychology*, 12, 113–126.
- Mishkin, M., & Forgays, D. G. (1952). Word recognition as a function of retinal locus. *Journal of Experimental Psychology*, 43, 43–48.
- Muchinsky, P. (1996). The correction for attenuation. *Educational and Psychological Measurement*, 56(1), 63–75. doi:10.1177/0013164496056001004
- Ocklenburg, S., & Güntürkün, O. (2018). *The lateralized brain: The neuroscience and evolution of hemispheric asymmetries*. London: Academic Press.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Parsons, S., Kruijt, A., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science*. doi:10.1177/2515245919879695
- Porac, C., & Coren, S. (1976). The dominant eye. *Psychological Bulletin*, 83, 880–897.
- Rosch, R. E., Bishop, D. V., & Badcock, N. A. (2012). Lateralised visual attention is unrelated to language lateralisation, and not influenced by task difficulty—a functional transcranial Doppler study. *Neuropsychologia*, 50, 810–815.
- Roszkowski, M. J., & Snelbecker, G. E. (1982). Temporal stability and predictive validity of self-assessed hand preference with first and second graders. *Brain and Cognition*, 1, 405–409.
- Steenhuis, R. E., & Bryden, M. P. (1999). The relation between hand preference and hand performance: What you get depends on what you measure. *Laterality: Asymmetries of Body, Brain and Cognition*, 4, 3–26.
- Van der Haegen, L., & Brysbaert, M. (2018). The relationship between behavioral language laterality, face laterality and language performance in left-handers. *PLoS one*, 13, e0208696.
- Van der Haegen, L., Cai, Q., Seurinck, R., & Brysbaert, M. (2011). Further fMRI validation of the visual half field technique as an indicator of language laterality: A large-group analysis. *Neuropsychologia*, 49, 2879–2888.
- Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176–1190.

- Voyer, D. (1998). On the reliability and validity of noninvasive laterality measures. *Brain & Cognition*, 36, 209–236.
- Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11, 192–196.
- Weekes, N. Y., Capetillo-Cunliffe, L., Rayman, J., Iacoboni, M., & Zaidel, E. (1999). Individual differences in the hemispheric specialization of dual route variables. *Brain and Language*, 67, 110–133.
- Westerhausen, R. (2019). A primer on dichotic listening as a paradigm for the assessment of hemispheric asymmetry. *Laterality: Asymmetries of Body, Brain and Cognition*, 24, 740–771.
- Westerhausen, R., & Samuelsen, F. (2020, January 31). An optimal dichotic-listening paradigm for the assessment of hemispheric dominance for speech processing. doi:10.31234/osf.io/tpjes
- Whitehouse, A. J., & Bishop, D. V. (2009). Hemispheric division of function is the result of independent probabilistic biases. *Neuropsychologia*, 47, 1938–1943.
- Willemijn, J., Hausmann, M., Brysbaert, M., Dael, N., Chmetz, F., Fioravera, A., ... Mohr, C. (2016). Stability of right visual field advantage in an international lateralized lexical decision task irrespective of participants' sex, handedness or bilingualism. *Laterality: Asymmetries of Body, Brain and Cognition*, 21, 502–524.
- Woodhead, Z. V. J., Bradshaw, A. R., Wilson, A. C., Thompson, P. A., & Bishop, D. V. M. (2019). Testing the unitary theory of language lateralization using functional transcranial Doppler sonography in adults. *Royal Society Open Science*, 6, 181801.
- Woodhead, Z. V. J., Thompson, P. A., Karlsson, E., & Bishop, D. V. M. (2020, September 9). Investigating multidimensionality of language lateralisation in left and right handed adults: an update on Woodhead et al. 2019. <https://doi.org/10.31234/osf.io/mrkgf>.
- Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79, 2064–2072.
- Yovel, G., Tambini, A., & Brandman, T. (2008). The asymmetry of the fusiform face area is a stable individual characteristic that underlies the left-visual-field superiority for faces. *Neuropsychologia*, 46, 3061–3068.