Quantitative MRI reveals differences in striatal myelin in children with DLD

Saloni Krishnan[1,2]

Gabriel J. Cler[1,3]

Harriet J. Smith[1,4]

Hanna E. Willis[1,5]

Salomi S. Asaridou[1]

Máiréad P. Healy[1,6]

Daniel Papp[8,9]

Kate E. Watkins[1]

[1] *Wellcome Centre for Integrative Neuroimaging, Dept of Experimental Psychology, University of Oxford, OX2 6GG, Oxford, UK*

[2] *Department of Psychology, Royal Holloway, University of London, Egham Hill, Surrey TW20 0EX*

[3] *Department of Speech and Hearing Sciences, University of Washington, Seattle WA*

[4] *MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, UK*

[5] *Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Headley Way, Headington, Oxford OX3 9DU*

[6] *Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB*

[8] *NeuroPoly Lab, Biomedical Engineering Department, Polytechnique Montreal, 2500 Chemin de Polytechnique, H3T 1J4, Montreal, Canada*

[9] *Wellcome Centre for Integrative Neuroimaging, FMRIB Centre, Nuffield Department of Clinical Neuroscience, University of Oxford, OX3 9DU, Oxford, UK*

Corresponding author:

Dr. Saloni Krishnan

Department of Psychology, Royal Holloway, University of London, Egham Hill, Surrey TW20 0EX

<saloni.krishnan@rhul.ac.uk>

Quantitative MRI reveals differences in striatal myelin in children with DLD

Abstract

Developmental Language Disorder (DLD) is a common neurodevelopmental disorder characterised by receptive or expressive language difficulties or both. While theoretical frameworks and empirical studies support the idea that there may be neural correlates of DLD in frontostriatal loops, findings are inconsistent across studies. Here, we use a novel semiquantitative imaging protocol – multiparameter mapping (MPM) – to investigate microstructural neural differences in children with DLD. The MPM protocol allows us to reproducibly map specific indices of tissue microstructure. In 56 typically developing children and 33 children with DLD, we derived maps of: 1) longitudinal relaxation rate R1 (1/T1); 2) the transverse relaxation rate R2* (1/T2*); and 3) Magnetization Transfer Saturation (MTsat). R1 and MTsat predominantly index myelin, while R2* is sensitive to iron content. Children with DLD showed reductions in MTsat values in the caudate nucleus bilaterally, as well as in the left ventral sensorimotor cortex and Heschl's gyrus. They also had globally lower R1 values. No group differences were noted in R2* maps. Differences in MTsat and R1 were coincident in the caudate nucleus bilaterally. These findings support our hypothesis of corticostriatal abnormalities in DLD and indicate abnormal levels of myelin in the dorsal striatum in children with DLD.

Quantitative MRI reveals differences in subcortical microstructure in children with DLD

Children with Developmental Language Disorder (DLD) struggle to learn their native language for no obvious reason. DLD is an extremely common neurodevelopmental disorder, with recent estimates indicating the prevalence of DLD is 7% (Norbury et al., 2016). DLD has serious economic and social consequences – it is associated with a higher risk for academic underachievement, unemployment, social and behavioural difficulties, and detriment to well-being (Conti-Ramsden et al., 2018). Although we know that DLD does not result from gross neural lesions, we still do not have a clear picture of how brain anatomy differs in children with DLD (Krishnan et al., 2016; Mayes et al., 2015). This is not only practically relevant but would also help us to understand the neural underpinnings of language development. Here, we use a robust new semiquantitative imaging protocol – MPM or Multi-Parameter Mapping (Weiskopf et al., 2013, 2021) – to shed light on microstructural neural differences in a large group of children with DLD.

There is a dearth of literature examining brain structure in children with DLD, which is surprising given the prevalence and impact of DLD. In the available literature, grey matter changes have been noted in the left inferior frontal gyrus and the posterior superior temporal gyrus (Badcock et al., 2012; Gauger et al., 1997; Jancke et al., 2007; Lee et al., 2020; Plante, 1991; Preis et al., 1998). These are core regions for language processing, with language activation in these regions observed across the literature (Price, 2012) and across languages (Ayyash et al., 2021). Differences have also been noted in areas homologous to these language regions, such as the right perisylvian cortex (Girbau-Massana et al., 2014; Jancke et al., 2007; Kurth et al., 2018). However, there is variability in the direction of differences reported in different studies – for instance, both increases and decreases in grey matter have been noted in the left inferior frontal gyrus (Badcock et al., 2012; Gauger et al., 1997; reviewed in Mayes et al., 2015).

In addition to these cortical changes in the language network, we, and others, have hypothesised that the dorsal striatum is important for language learning, and may be abnormal in DLD (Krishnan et al., 2016; Ullman et al., 2020; Ullman & Pierpont, 2005). The dorsal striatum is important for habitual and sequential learning (Graybiel & Grafton, 2015; Yin & Knowlton, 2006), and we hypothesise that it may play an important role in the

acquisition of language because of the complexity of sequencing required for language. In accordance, a series of behavioural studies has suggested that sequential learning in the linguistic and non-linguistic domains is affected in children with DLD (Hsu & Bishop, 2014; Lum et al., 2014; but see West et al., 2018, 2021). In studies that probe complex sequential production in the vocal domain, the dorsal striatum is implicated (Rauschecker et al., 2008; Simmonds et al., 2014; Skipper et al., 2020). The striatum is structurally and functionally connected to regions associated with language production, with the head of the dorsolateral caudate nucleus receiving inputs from inferior frontal cortex, and the putamen receiving inputs from motor, premotor and supplementary motor cortex (Alexander et al., 1986; Jarbo & Verstynen, 2015; Lima et al., 2016). The importance of the dorsal striatum for speech and language was first highlighted by work on the KE family, who have a point mutation in the *FOXP2* gene and a behavioural profile similar to DLD (Watkins, Dronkers, et al., 2002), as well as childhood apraxia of speech. Morphometric studies revealed that affected members of the KE family had reduced grey matter in regions not typically associated with language processing such as the head of the caudate nucleus, areas within the sensorimotor cortex, the posterior inferior temporal cortex and the posterior lobe of the cerebellum, and increased grey matter in the putamen (Argyropoulos et al., 2019; Belton et al., 2003; Watkins, Vargha□Khadem, et al., 2002). These studies indicate that corticostriatal circuits might have a role in co-ordinating and learning the fine auditory-motor sequencing required for language.

Studies of children with DLD analysing standard T1-weighted scans have also indicated reductions in the size of the caudate nucleus (Badcock et al., 2012; Herbert et al., 2003; Jernigan et al., 1991). Other studies do not wholly support the view that the volume of the caudate nuclei is reduced in DLD but do indicate that there is some abnormality. Lee and colleagues (2013) found reduced absolute volumes in the caudate nucleus and thalamus in individuals with DLD. These relationships did not survive correcting for total intracranial volume, which was significantly reduced in their DLD participants. The authors also observed negative relationships between language proficiency and the relative volume of subcortical structures such as the nucleus accumbens, globus pallidus, putamen and hippocampus. Others have suggested that volumetric differences in the caudate nuclei are modulated by age, with only younger children showing differences in volume (Soriano□Mas et al., 2009). Finally, some more recent studies suggest that children with DLD have greater grey matter in the right cerebellum (Pigdon et al., 2019). The interpretation of both cortical and subcortical findings is complicated by the heterogeneity of the DLD populations

sampled, and the small sample sizes investigated. However, another factor that may also contribute to this lack of clarity is the nature of the scans acquired.

Standard structural imaging protocols such as T1-weighted scans reflect a complex mix of tissue properties, or in other words, the contrast between grey and white matter reflects a combination of histological properties such as iron content, myelin, cell density and water. Importantly, these microstructural properties yield regionally-specific contributions to commonly used structural markers such as grey matter volume or cortical thickness, which complicate the interpretation of these markers (Lorio et al., 2014, 2016). As standard T1-weighted imaging protocols are dependent on acquisition parameters that can vary across scanners, they are also often difficult to replicate across studies. More recently, semiquantitative MRI methods have been used to map specific properties of tissue (Weiskopf et al., 2021). Semiquantitative protocols such as MPM can provide specific indices of microstructure, including myelin and macromolecular content of neural tissue, and the resulting maps are highly reproducible across individuals and scanners (Leutritz et al., 2020; Weiskopf et al., 2013). In the MPM quantitative imaging protocol, multiple maps are constructed, which allow us to probe different tissue properties. The generated maps quantify i) the longitudinal relaxation rate R1 [1/T1], ii) the transverse relaxation rate R2* [1/T2*], and iii) Magnetization Transfer Saturation MTsat. The dominant influence on R1 in cortical tissue is myelin (Lutti et al., 2014), although R1 is sensitive to both myelin and iron in subcortical grey matter. R2* is sensitive to iron concentration, especially in ferritin-rich regions such as the striatum (Langkammer et al., 2010). MTsat is sensitive to bound water, and consequently myelin (Schmierer et al., 2004). This quantitative protocol therefore represents an unparalleled means of acquiring time-efficient, multi-modal, whole-brain data with insight into tissue composition. Such semiquantitative maps have been used to delineate heavily myelinated areas such as somatomotor (Carey et al., 2017), visual (Sereno et al., 2013), and auditory cortex (Dick et al., 2012). They have also been used to characterise developmental maturation in adolescence and young adulthood (Carey et al., 2018; NSPN Consortium et al., 2019; Paquola et al., 2019; Whitaker et al., 2016), during aging (Callaghan et al., 2014; Draganski et al., 2011; Steiger et al., 2016), and in pathological populations (Freund et al., 2013; Manara et al., 2019). A growing number of studies use these maps to understand brain-behaviour relationships (Allen et al., 2017; Clark et al., 2020). Most recently, we have used MPMs and found elevated iron levels in the putamen and speech motor network in people who stutter (Cler et al., 2021). However, this novel semiquantitative

protocol has not yet been used to examine microstructure in children with neurodevelopmental disorders. A particular advantage of this protocol is that its sensitivity to cortical myelin can help distinguish two different explanations of developmental change in grey matter. Developmentally, cortical thinning indexed through grey/white matter contrast changes in standard T1-weighted scans could reflect a loss in the number of connections within grey matter, i.e., 'synaptic pruning' (Huttenlocher, 1979), or a gain in the volume of tissue through increased intra-cortical myelination that appears to "whiten" the grey matter on T1-weighted images (Paus, 2005). Recent studies using MPM protocols have shown that cortical thinning during development is associated with increased myelination rather than synaptic pruning (Natu et al., 2019; Whitaker et al., 2016).

In the present study, we used the MPM quantitative imaging protocol to map contrast parameters (R1, R2* and MTsat) in typically developing children and those with DLD. Based on our previous studies (Badcock et al., 2012; Watkins, Vargha□Khadem, et al., 2002), we hypothesised that these indices would reveal that the microstructure of i) the dorsal striatum (the caudate nuclei and the putamen) and ii) the left inferior frontal gyrus was altered in those with DLD.

Results

As part of the Oxford BOLD study, we collected brain imaging data from 162 children between the ages of 10-15 years, as well as performing a detailed characterisation of their language and cognitive skills. All children in the study had a nonverbal IQ > 70. Children were categorised as DLD if they scored 1SD below the mean on two or more language tests (N=57), and typically developing (N=77) if they scored +/- 1 SD of the mean on language tests (see Methods for more detail). Behavioural testing in a further 28 revealed that they did not meet our criteria for DLD but presented with a history of speech and language problems (HSL). After quality control, we retained MPM data from 109 of these children, including 56 typically developing children, 33 children with DLD and 20 children with HSL. The children in the HSL group were excluded from comparisons of TD and DLD, but were included in continuous analyses which allow us to examine language variability. The three groups (TD, HSL, and DLD) did not differ in in terms of age (see Table 1)

Table 1. Descriptive Data for the Typically Developing (TD), Developmental Language Disorder (DLD), and History of Speech and Language Difficulties (HSL) groups. Means are shown below, with standard deviations in parentheses. Language proficiency and memory factor scores are derived from a factor analysis (for a full description, see Krishnan et al., 2021). Nonverbal IQ is a scaled score (M = 10, SD =3) representing an average of performance on block design and matrix reasoning tasks. The last column shows whether there were significant group differences when using t-tests ($p < .05$), no correction for multiple comparisons is applied.

	TD	***DLD***	***HSL***	***Group differences***
Age (years)	12.41 (1.62)	12.48 (1.80)	12.40 (1.67)	None
Gender	28F:28M	11F:22M	3F:17M	N/A
Total Intracranial volume (mm^3)	1329.24 (145.98)	1345.39 (145.43)	1411.51 (158.99)	None
Language Proficiency	0.8 (0.45)	-0.95 (0.55)	-0.06 (0.48)	TD>HSL>DLD
Memory	0.58 (0.80)	-0.74 (0.79)	0.07 (0.77)	TD>HSL>DLD
Nonverbal IQ	12.3 (1.91)	8.65 (2.09)	11.0 (1.83)	TD>HSL>DLD

Whole-brain comparisons of neural microstructure in children with DLD with typically developing (TD) children

We first compared children who met criteria for DLD and TD children. There were no group differences in mean MTsat and R2* values for grey and white matter, or in total intracranial volume. The groups did differ in mean R1 values for grey matter (see Table 2), in that children with DLD had lower R1 than the TD group across all grey matter. We created averages of MTsat, R1, and R2* for each group. Across all three maps, we observed high values in primary motor, visual, and auditory cortex (see Figure 3- figure supplement 1), in line with our expectations. Additionally, we observed a close correspondence between average values of MTsat, R1, and R2* in grey matter and white matter in our two groups and published values (Weiskopf et al., 2013), see Table 2.

Table 2. Group mean and standard deviation of parameter values in grey and white matter in children who were typically developing (TD), had developmental language disorder (DLD), or published mean and standard deviation in adults, mean 24.2 years, S.D. 1.6 years (Weiskopf et al., 2013). Values in bold indicate a global difference between the TD and DLD groups (p<.05), uncorrected for multiple comparisons.

	TD	**DLD**	**Published values**
Grey matter			
MT	0.828 (0.018)	0.821 (0.019)	0.794 (0.014)
R1	***0.615 (0.012)***	***0.608 (0.014)***	0.609 (0.008)
R2*	15.192 (0.437)	15.234 (0.544)	15.2 (0.4)
White matter			
MT	1.72 (0.052)	1.714 (0.053)	1.764 (0.066)
R1	0.973 (0.025)	0.962 (0.025)	1.036 (0.036)
R2*	20.672 (0.659)	20.679 (0.704)	21.0 (0.8)

We then investigated whether there were group differences in neural microstructure by assessing each of the parameter maps using nonparametric permutation methods and established significant clusters using the threshold-free cluster enhancement method, setting a whole-brain corrected threshold of $p<.05$. Children with DLD had lower MTsat values than TD children in the inferior frontal gyrus (pars opercularis), ventral sensorimotor cortex, insular cortex, lateral Heschl's gyrus, planum temporale, and posterior superior temporal sulcus of the left hemisphere, and in portions of lateral and dorsomedial occipital cortex bilaterally. Subcortically, the children with DLD also had reduced MTsat in the dorsal caudate nucleus bilaterally; these differences were seen mainly in the body and were more extensive on the left than the right (Figure 1 and Supplementary file 1a). Consistent with the mean global differences in R1 in children with DLD (see Table 2), examination of the R1 maps revealed widespread reduction over the lateral convexities of the frontal and parietal lobes bilaterally (but slightly more on the right), the medial frontal cortex including SMA and extending to paracentral lobule and left posterior temporal cortex extending from the superior temporal plane to posterior inferior temporal cortex. Subcortically, there were differences in

R1 in the dorsal striatum and thalamus bilaterally, and in anterior portions of the medial temporal lobe (Figure 2). We did not find any significant group differences when examining the R2* maps. There were also no regions where children with DLD showed greater MT, R1 or R2* values relative to TD children.

Given that MTsat and R1 maps are particularly sensitive to myelin, and the R2* maps did not differ between the groups, we then examined if there were regions where we would see convergence of differences across MTsat and R1. A conjunction analysis was performed, where we assessed which voxels showed significant TD > DLD differences in both MTsat and R1 maps. We found conjoint differences in the MT and R1 maps across several brain regions, including the caudate nuclei bilaterally, and in the left ventral sensorimotor cortex, insula, lateral Heschl's gyrus, planum temporale, posterior superior temporal sulcus and middle temporal gyrus (Figure 3 and Supplementary file 1b).

Figure 1. Brain areas showing reduced MTsat in DLD *Coloured maps resampled and overlaid on the fsaverage surface (reconstructed using the HCP workbench) show TD > DLD differences in MTsat values (whole-brain corrected threshold p<.05). Axial slices (coloured maps are overlaid on an average MTsat image from all participants) reveal additional group differences subcortically in the caudate nuclei. The inset shows a boxplot of MTsat values drawn from the cluster in the caudate nuclei by group, with this cluster split into right and left using a hemispheric mask (orange – DLD, green – TD).*

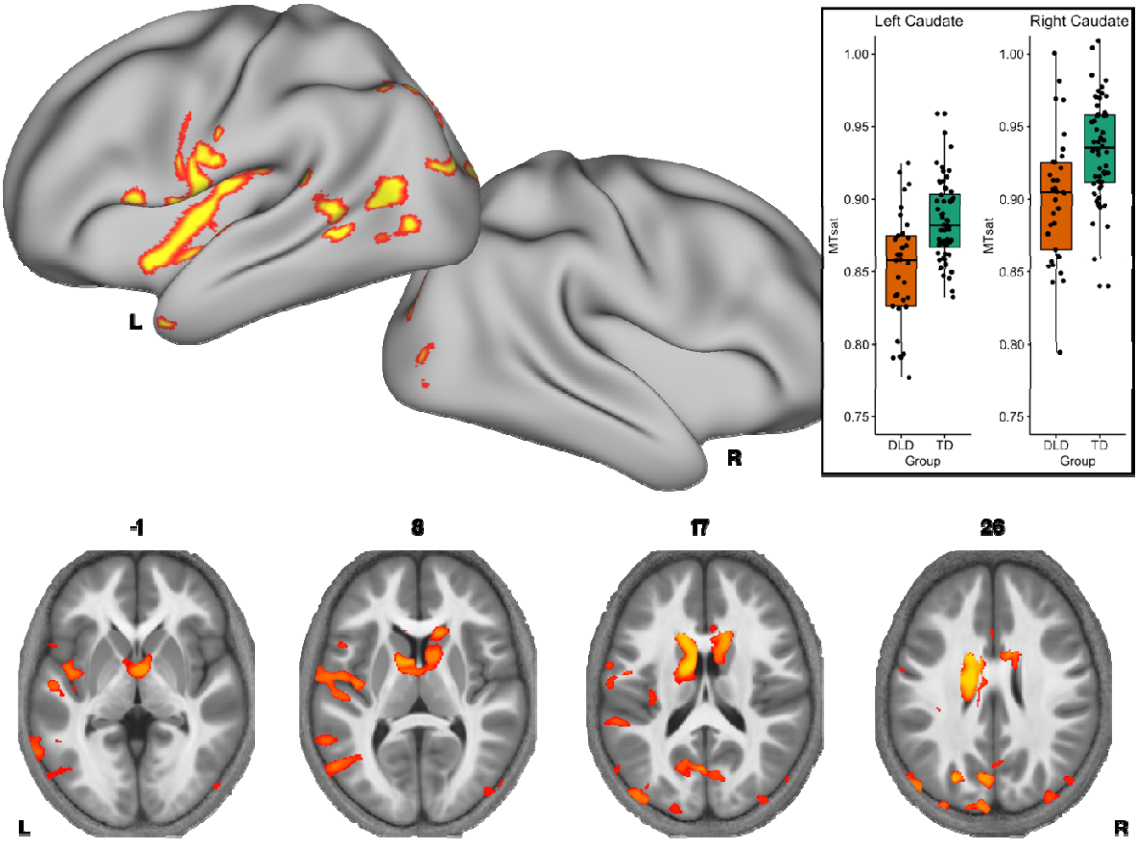


Figure 2. Brain areas showing reduced R1 in DLD *Coloured maps resampled and overlaid on the fsaverage surface (reconstructed using the HCP workbench) show TD > DLD differences in R1 maps (whole brain corrected threshold p<.05). Axial slices (coloured maps are overlaid on an average MTsat image from all participants) reveal additional subcortical group differences in the striatum and thalamus bilaterally.*

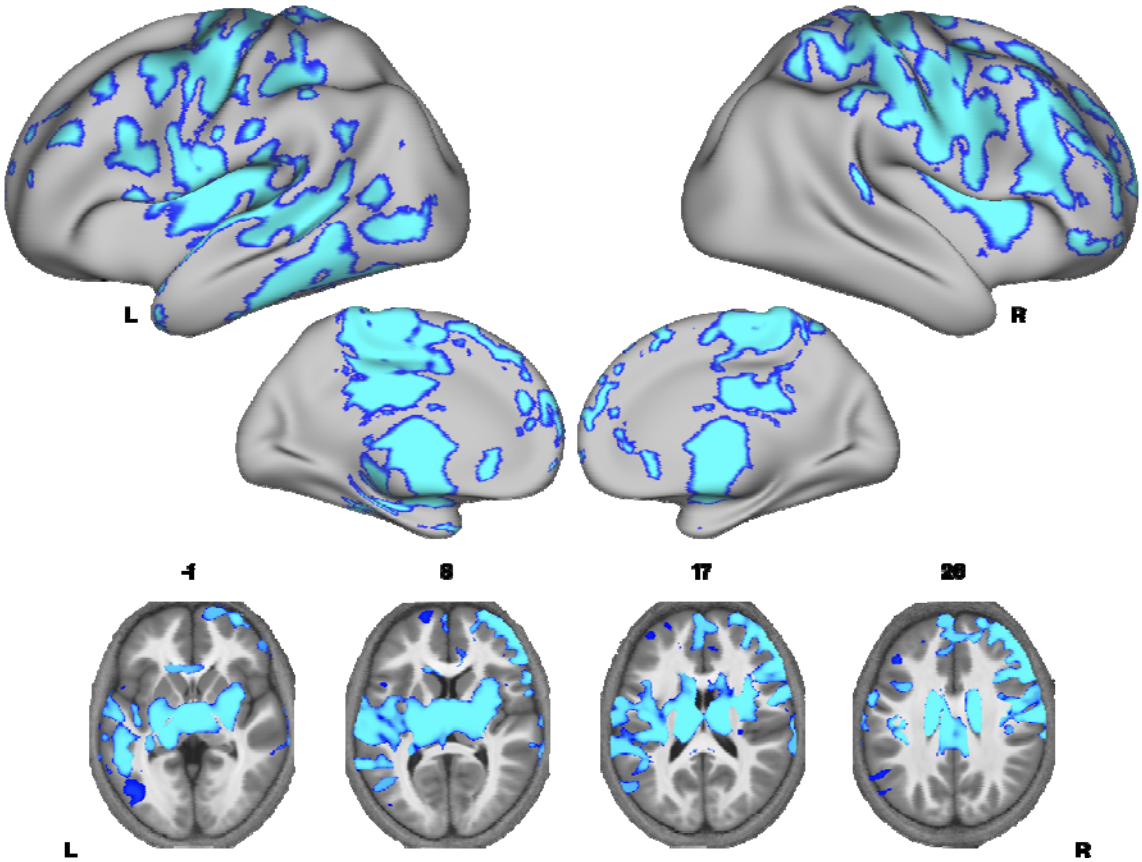
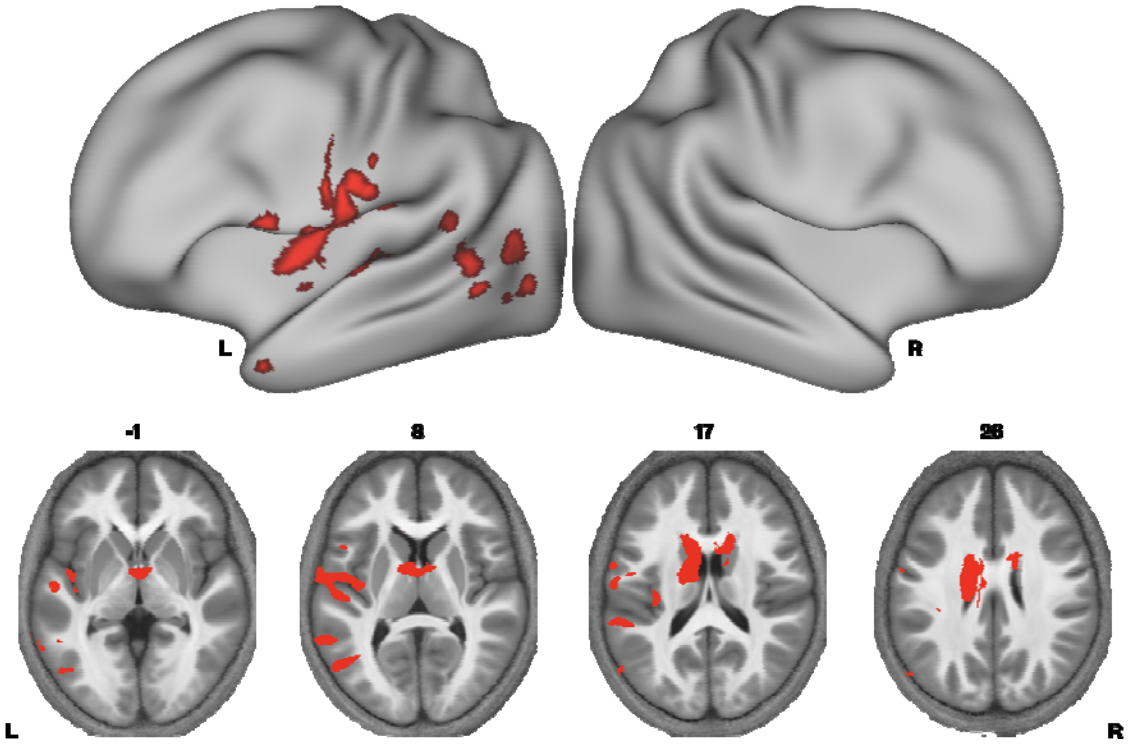


Figure 3. Brain areas showing conjoint differences in MTsat and R1 in DLD *Coloured maps resampled and overlaid on the fsaverage surface (reconstructed using the HCP workbench) show convergence of TD > DLD differences in MTsat and R1 maps thresholded at p<0.05. Axial slices (coloured maps are overlaid on an average MTsat image from all participants) reveal the differences subcortically in the dorsal striatum in particular.*



The CATALISE definition of DLD (Bishop et al., 2016, 2017) rightly removes the criterion requiring a discrepancy between verbal and nonverbal skills, which in practice means broadening the phenotype to include children with low nonverbal IQ. To determine whether our group differences were in fact driven by the inclusion of children with low nonverbal IQs in addition to low language ability, we assessed differences only in children with nonverbal IQ scores > 85. Group differences in MTsat and R1 maps were observed even when removing the five children with DLD who had nonverbal IQs between 70-85 (see Figure 1 – figure supplement 1 & Figure 2 – figure supplement 1), indicating the differences reported here were not driven by the inclusion of children with low nonverbal IQs. Consequently, all following analyses include these children.

Given our interest in the dorsal striatum in relation to DLD (see introduction), we extracted average MTsat values for each participant from the caudate nuclei bilaterally (the region

where we observed TD>DLD differences). Using a hemispheric mask, we separated this cluster into left and right components. We then assessed if group differences in average MTsat values in the left and right caudate nuclei could be accounted for by age or total intracranial volume. Age and total intracranial volume were not significant predictors of MTsat values in the caudate nuclei clusters, see Figure 1 – figure supplement 2.

Whole-brain correlation analysis of neural microstructure with language proficiency

We have previously found continuous measures of language proficiency to be more sensitive to neural differences than diagnostic categories (Krishnan et al., 2020). Using a continuous measure of language ability also gave us the opportunity to include the group of children with a history of speech and language problems who did not meet criteria for DLD on testing (HSL; N=20). We constructed language and memory factor scores from our neuropsychological battery (see Krishnan et al., 2020 for further detail on model construction), using data from the full sample in whom behavioural data were available in the BOLD study. We then assessed if these factors, as well as nonverbal IQ, were predictors of our three parameter values across our whole sample (N=109) using whole-brain analyses. Given the strong correlation between the language and memory factors ($r=0.7$, $p<.001$), we entered these predictors separately into our statistical models. We found that language proficiency was strongly positively correlated with MTsat values focally in the left caudate nucleus (Figure 4). In the R1 maps, poorer language proficiency was once again associated with widespread reduction over the lateral convexities of the frontal and parietal lobes bilaterally (as seen before in the TD > DLD differences, which were slightly more right lateralised), the medial frontal cortex including SMA and extending to paracentral lobule, and left posterior temporal cortex extending from the superior temporal plane to posterior inferior temporal cortex (see Figure 4 – figure supplement 1). Subcortically, language proficiency was positively associated with R1 values in the dorsal striatum and thalamus bilaterally, and in anterior portions of the medial temporal lobe. This pattern of results was similar to those derived in the TD > DLD contrast. Lower R1 values in a more focal but overlapping set of regions (perisylvian cortex including the inferior frontal gyrus, insula, superior temporal gyrus, extending to the anterior temporal pole, bilaterally, but more extensive on the left, and the dorsal striatum bilaterally) were associated with poor memory proficiency, suggesting that these cognitive differences were reflecting the globally lowered values of R1 in children with DLD. There were no significant relationships between R2* values in grey matter and language or memory proficiency.

Associations with language proficiency, memory and IQ in the dorsal striatum

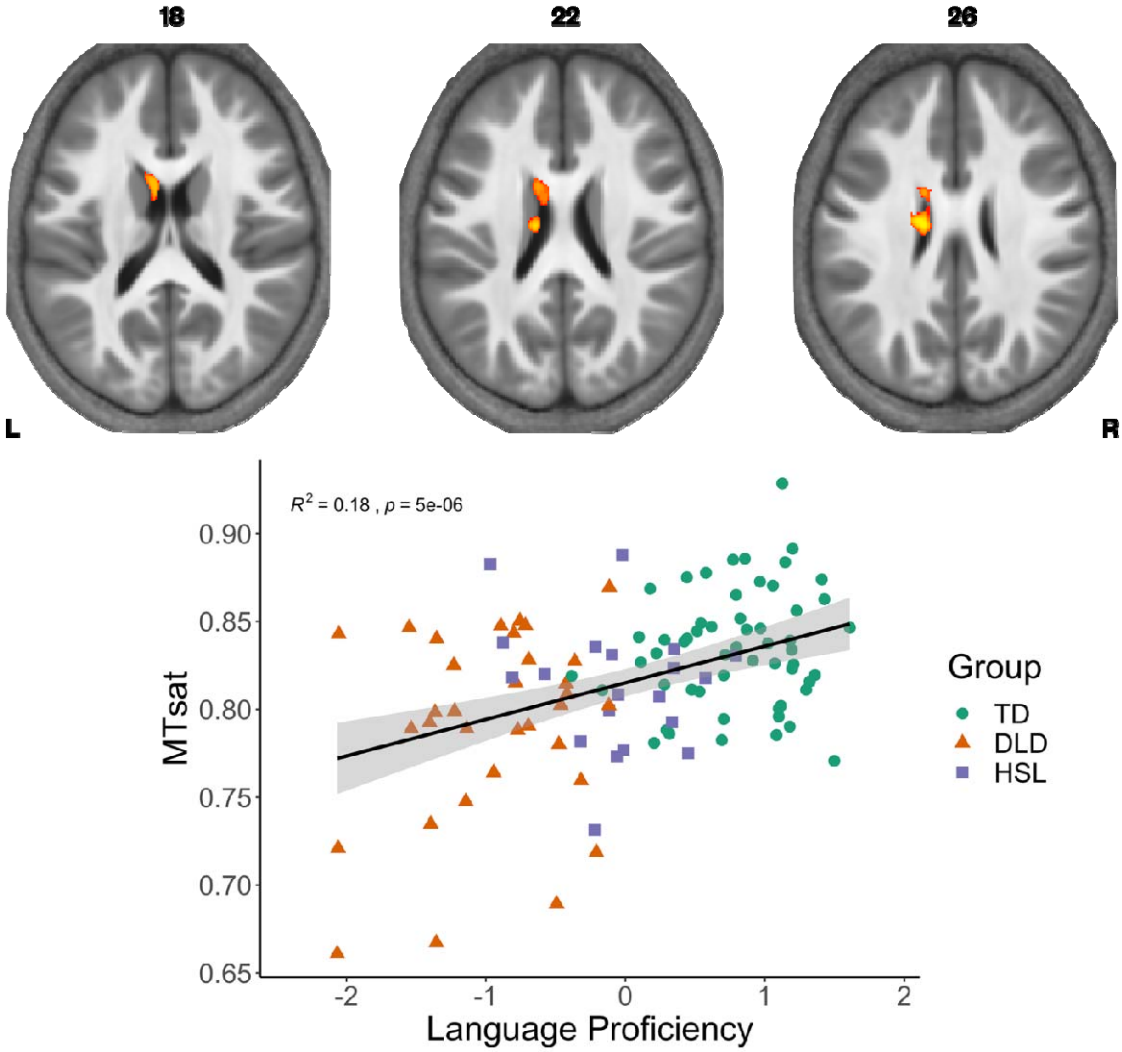
The DLD and TD groups differed on language proficiency, memory, and nonverbal IQ (see Table 1). We consequently assessed whether variation in language proficiency, memory, or nonverbal IQ best explained the variation in R1 and MTsat values in the portions of the caudate nucleus where we observed TD > DLD group differences. We constructed stepwise regressions to evaluate the contribution of each of these factors, including children from the HSL group to maximise power (the pattern of analysis was the same when we limited our analyses to children in the TD or DLD groups).

For MTsat values in the caudate nucleus, we found that a model with language proficiency alone (β=.014, p<.001) was the best fitting model, explaining 14.57% of the variance, $F(1,106)=18.08$, $p<.001$. For R1 values in the caudate nucleus, we again observed that a model with language proficiency alone (β=.015, p<.001) was the best fitting model, explaining 13.19% of the variance, $F(1,106)=16.11$, $p<.001$.

TD vs DLD group differences in voxel-based morphometry

Finally, to assess if the parametric differences reported here were also observed in standard morphometric analysis, we performed a VBM analysis comparing regional amounts of grey matter in TD vs. DLD using T1 scans from the same participants (note these were collected using an MPRAGE T1-weighted sequence and were not the T1-weighted scan from the MPM protocol). We did not observe any group differences that survived our statistical threshold ($p<0.05$), and indeed no group differences were observed at a lower statistical threshold $p<0.2$. Thus, our parametric differences cannot be explained by morphometric differences in the amount of grey matter.

Figure 4. *Correlation between language proficiency and MTsat values in the left caudate nucleus*



Discussion

In this study, we used semiquantitative structural MRI to provide a detailed account of the neural differences in children with developmental language disorder (DLD), taking advantage of the sensitivity of this protocol to understand changes in neural microstructure. We found reduced MTsat and R1 values in the caudate nuclei in children with DLD. We also observed reduced MTsat values in the left inferior frontal gyrus. This offers empirical confirmation of our prediction that cortico-striato-thalamo-cortical loops involved in learning are affected in DLD.

Importantly, these results allow us to understand the cellular mechanisms driving this change. MTsat and R1 are considered in vivo markers of macromolecular content, and within the brain, these measures are particularly sensitive to myelin content in grey matter. Although most of the brain's myelin is found in the white matter where it sheaths the long axons travelling in white matter fibre tracts, it can also be measured in grey matter where it myelinates axons or parts of axons in the cortex and in subcortical structures like the thalamus and basal ganglia (Nieuwenhuys, 2013). Indeed, myelin is a strong contributor to MR signal in these regions (Glasser & Essen, 2011; Sereno et al., 2013; Shafee et al., 2015). Given myelin's role in enabling fast and reliable communication in the brain, reduced myelin content may explain why children with DLD struggle with speech and language processing. Below, we discuss these findings, contextualising why myelin may be altered in these specific regions in DLD.

Interpretation of MPM scans

Myelin is known to increase throughout the brain during adolescence (Paquola et al., 2019; Whitaker et al., 2016), and has been linked to improved behavioural performance (Kwon et al., 2020). Previous morphometric studies have typically drawn inferences about myelin from cortical thinning, but shrinkage of grey matter does not allow us to distinguish if T1 change is due to shorter T1 times associated with reduced synaptic density, or an increase in the proportion of myelinated neurons (Paus, 2005). More recently, MTsat and R1 measures have been used as an in vivo marker for myelin. These have been validated as markers of myelin through post mortem imaging, as well as histological studies of patients with multiple sclerosis (Weiskopf et al., 2021). Importantly, these maps agree with histological maps showing greater myelin in primary motor and sensory cortex (Carey et al., 2017; Dick et al.,

2012; Sereno et al., 2013; Whitaker et al., 2016), as can be seen in our maps as well (Figure 3- figure supplement 1). It is therefore unsurprising to find some convergence over MTsat and R1. R2* is also considered a measure of myelin, but the T2* contrast is particularly sensitive to iron, especially within the basal ganglia. Despite R2* differences being noted in adults with speech problems in the putamen and cortical speech motor network (developmental stuttering, Cler et al., 2021), and observed links between verbal memory and R2* in the ventral striatum (Steiger et al., 2016), we did not see any evidence for R2* differences anywhere in the brain when comparing our TD and DLD groups.

Our findings also strongly suggest that there is an advantage to using multi-parameter mapping to probe microstructure relative to standard T1 measures. While previous studies reported differences using morphometric measures, we did not observe evidence for morphometric differences in our sample. As others have argued, multi-parameter mapping is sensitive to microstructural differences that would not be detected by a standard voxel-based morphometry (VBM) analysis, as VBM is sensitive to differences in regional amounts of grey matter based on T1-weighting rather than quantitative measurements (Lorio et al., 2014, 2016). Using MPMs allows us to more closely interrogate the histological processes involved in neural changes.

Group differences in the striatum

Our findings clearly indicate that there are microstructural abnormalities in the caudate nuclei bilaterally in children with DLD, and more broadly, that reduced myelin in the left caudate nucleus is associated with lower language proficiency. The convergence across MT and R1 differences in the caudate nuclei strongly suggest these are linked to abnormal levels of myelin.

The dorsal striatum has been implicated in learning through practice, particularly habit formation (Skipper et al., 2020; Yin & Knowlton, 2006). A number of groups, including us, have hypothesised that the striatum is a brain region where we might expect to see neural differences in children with DLD (Krishnan et al., 2016; Ullman et al., 2020; Ullman & Pierpont, 2005). Our hypothesis was driven by the idea that children with DLD showed deficits in sequential procedural tasks, which rely on loops through the striatum (Krishnan et al., 2016). Here, using a direct neural measure, we see evidence for structural differences in the striatum, particularly in the caudate nuclei. The functional consequences of these

structural changes might be to make certain aspects of learning, such as the learning of stimulus-response mappings, more challenging. For instance, a reduction in myelin might make sequential learning less efficient. This could explain why children with DLD face difficulties in complex language tasks, such as nonword repetition, where extracting and producing sequential responses are important. However, any speculation about the functional impact of these changes needs careful empirical investigation in future studies.

Our findings are generally consistent with previous smaller-scale work linking volumetric differences in the caudate nucleus to language learning difficulties (Badcock et al., 2012; Watkins, Vargha-Khadem, et al., 2002). They may also offer some resolution to previous work using VBM, where such differences were not observed (Pigdon et al., 2019). Indeed, our own VBM analysis did not reveal any differences in morphometry. However, and perhaps speaking to a slightly different interpretation of our findings, the striatal differences we find are in the body of the caudate nucleus, rather than the head. The head of the caudate nucleus receives input from prefrontal cortex. DTI studies suggest the body of the caudate nucleus gets its input from prefrontal cortex and supplementary motor regions (Lehéricy et al., 2004). Functional connectivity studies indicate that the body of the caudate nucleus also receives projections from temporal association cortex (Choi et al., 2012). Given our findings of differences in the body rather than the head of the caudate nuclei, further studies examining individual differences of microstructure and relevant behavioural tasks (e.g., auditory processing and learning, or processing of rhythm) are warranted.

Although some previous work suggested that striatal differences were normalised by adolescence in children with DLD (Soriano-Mas et al., 2009), our analyses do not show any evidence of group differences being modulated by age (Figure 1 – figure supplement 2). Using the MPM protocols may give us a more sensitive marker of change to differences in DLD. In these analyses, it is perhaps surprising that we did not see age-related change in the striatum within the time frame we sampled. Longitudinal work using MTsat has demonstrated that the striatum continues to mature through adolescence, from 14 to 24 years of age (NSPN Consortium et al., 2019). However, it may be that the time frame we sampled was too short, or that within-subject longitudinal studies, which are more sensitive to changes over time, are needed.

Genetic mechanisms underlying group differences in the striatum

The genetic mechanisms that might drive these striatal changes are currently unclear. DLD is a highly heritable disorder but identification of genetic pathways has been difficult due to the broad phenotype and changes in diagnostic criteria(for a recent review, see Mountford et al., 2022). In terms of monogenic disorders, the best known case is that of the KE family, the affected members of which had a rare point mutation in the gene *FOXP2* causing verbal dyspraxia as well as the kinds of language learning difficulties seen in DLD. The mutation in *FOXP2* was associated with reduced volumes of the caudate nuclei bilaterally in the KE family (Watkins, Vargha Khadem, et al., 2002), but it is unknown whether this was related to a quantitative difference in myelin content as seen here in DLD. *FOXP2* is strongly expressed in the medium-spiny neurons of the striatum in humans and other species, and in many other neural and non-neural tissues(Lai, 2003). *FOXP2* is a transcription factor affecting the function of many downstream target genes, including *CNTNAP2* and *FOXP1,* which in turn have been linked to phenotypic features of DLD (Lozano et al., 2015; Vernes et al., 2008). Aside from identifying rare monogenic causes of DLD, other approaches have yielded common risk variants that explain variances in phenotypic features of DLD, including phonological skills (Newbury et al., 2009). A recent meta-analysis of DLD genome-wide association studies (GWAS) of reading and language traits with an N >30,000 found significant genome-wide association with word reading and distinct genetic variation of word and nonword reading, spelling and phoneme awareness that correlated with cortical surface area of the left superior temporal sulcus (Eising et al., 2021). Interestingly, in our study, the posterior part of the superior temporal sulcus in the left hemisphere showed significantly reduced MTsat and R1 values in children with DLD (see Figures 1, 2 and 3).

Neural differences beyond the striatum

We also observed MTsat changes indicative of myelin-related differences in the posterior part of the inferior frontal gyrus, ventral sensorimotor cortex, insula, planum temporale, lateral Heschl's gyrus, and the superior temporal sulcus, all in the left hemisphere. Regions such as the inferior frontal cortex and posterior superior temporal cortex are considered core parts of the speech and language network (Ayyash et al., 2021; Fedorenko et al., 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009), and ventral sensorimotor cortex and insula are regions that are clearly implicated in speech motor control (Bouchard et al., 2013; Carey et al., 2017; Dronkers, 1996; Wise et al., 1999). Our findings here of abnormal microstructure in these regions are somewhat consistent with those from previous studies in DLD that report structural differences in perisylvian regions (Badcock et al., 2012; Gauger et

al., 1997; Jancke et al., 2007; Plante, 1991; Preis et al., 1998) with two notable differences – one, that these differences did not emerge in a VBM analysis, and two, on average, children with DLD had lower MTsat or R1 values than typically developing children , indicative of slower maturation or abnormal developmental trajectories. Such differences in these regions therefore may be correlates of either auditory or motor inefficiency or both that have been observed in some children with DLD (Halliday et al., 2017; Hill, 2001; McArthur & Bishop, 2004).

While the TD>DLD differences observed in the MTsat appeared left-lateralised and focal, differences in the R1 map were widespread, observed in both hemispheres quite symmetrically. Interpreting the differences in these findings offers a paradox. On the face of it, left-lateralised myelin reduction in brain regions known to contribute to speech and language processing seems very plausible, as this would be entirely consistent with the behavioural profile of DLD. This would fit with a popular theoretical view, i.e., that the left hemisphere is uniquely privileged to support language (Vargha-Khadem et al., 1985). However, this does not fit with the developmental literature on children with early brain lesions. Children with perinatal focal brain lesions, even those encompassing the entire left hemisphere, have fairly good language skills, and they typically perform better than children with DLD on language tasks (Asaridou et al., 2020; Bates & Dick, 2002; Thal et al., 1991). Right hemisphere homologues of language areas are able to support language reorganisation when early damage is sustained (Bates & Dick, 2002; Newport et al., 2017). This has led to theoretical speculation that abnormalities affect both hemispheres in children with DLD preventing this form of brain plasticity. In this vein, the widespread differences seen in the R1 maps, or the bilateral abnormalities observed in the caudate nuclei, might point to why organisation of language is not maximally efficient. Another possibility is that MTsat mainly indexes myelin differences, whereas R1 could be sensitive to other microstructural features such as iron and neuronal fibres (Edwards et al., 2018). This might suggest that the myelin differences observed in R1, that are not observed in MTsat, indicate further sources of neural difference in DLD. Further work is needed to understand what the divergent TD>DLD differences across the R1 and MTsat maps might reflect.

It is also notable that many of the MTsat and R1 differences we observe are in primary motor and sensory areas, or closely adjacent areas. Myelin content is high in primary sensory (due to dense thalamo-cortical projections) and motor areas (due to the large axons of cortico-

spinal projections) and therefore peaks at earlier stages of development (Natu et al., 2019; Paquola et al., 2019; Whitaker et al., 2016). As seen from our average parameter maps (Figure 3 – figure supplement 1), we see the expected strong myelination in these regions, and we may therefore have stronger signal in these regions of the brain to evaluate group differences. In contrast, a relative lack of differences in association areas may be a true finding or may reflect reduced sensitivity to measure this change because there is less myelin content there. For instance, in our maps, we do not see very strong myelination in some areas, such as the inferior frontal gyrus, which might limit the ability to see group differences. In typically developing teenagers, myelogenesis is highest in association areas (Whitaker et al., 2016). Longitudinal studies are therefore necessary to evaluate whether differences in myelin persist in the same areas in children with DLD, i.e., regionally specific changes, or whether differences in myelin would be seen in association areas at later stages of development.

The broad constellation of challenges faced by children with DLD might also explain some of the group differences in MTsat and R1 values. For example, many children with DLD have co-occurring motor challenges (Hill, 2001; Sanjeevan & Mainela-Arnold, 2019), which may be reflected in changes in the motor network. This might also account for differences in regions we did not have clear hypotheses about, for example, in the occipital lobe. Many children in our study faced reading challenges, and left occipital-temporal hypoactivation is associated with dyslexia (Paulesu et al., 2014; Richlan, 2012) and now genetic variation in literacy skills (see above, Eising et al., 2021). One way to disentangle the contributions of these different cognitive traits would be to conduct large-scale studies including those with multiple diagnoses and use a transdiagnostic approach to establish specific relationships between brain and behaviour (Siugzdaite et al., 2020).

Finally, it is important to note that we previously hypothesised that the microstructure of grey and white matter in the medial temporal lobe would be relatively normal in children with DLD (Krishnan et al., 2016). However, we observed differences in R1 values in these regions (see Figure 2, medial surface of the left hemisphere). This fits with the emerging picture that children with DLD can struggle with aspects of learning thought to depend on the medial temporal lobe, e.g. so-called declarative memory tasks such as list learning (Bishop & Hsu, 2015; Earle & Ullman, 2021; Jackson et al., 2020; McGregor et al., 2017).

Implications for theories of DLD

The striatal differences we report are broadly consistent with our views on the dorsal striatum being involved in language learning (Krishnan et al., 2016), as well as with theories such as the procedural deficit hypothesis (Lum et al., 2014; Ullman & Pierpont, 2005), and the procedural deficit circuit hypothesis (Ullman et al., 2020). This may be linked to difficulties extracting and retaining the sequential regularities for language, or challenges in automatising the use of language rules. Our sample with DLD shows differences in linguistic tasks that draw upon these abilities, such as nonword repetition or oromotor sequencing, even though they were not selected on this basis. Orpella and colleagues (2021) recently demonstrated that the dorsal and ventral striatum were crucial for developing and acting upon predictions in a statistical learning task. The use of such sensitive functional tasks may be important to demonstrate the functional consequences of these changes. Understanding such learning mechanisms may also help us understand how to design better intervention for children with DLD, for example, this may be a reason to simplify and repeat rules during learning, providing greater opportunities for practice and habituation. We are now testing the role of repetition in enhancing language comprehension in DLD (Parker et al., 2021).

Our more global differences in myelination would also be consistent with theoretical views postulating that children with DLD show differences in the speed of processing, perhaps due to inefficiencies in information transfer through the brain (Kail, 1994; Miller et al., 2001). However, more recent theories point to more specific lexical processing differences in DLD. Children with DLD have been shown to have greater difficulty with lexical selection and inhibition of competitors, rather than the initial perceptual or phonological processing (Nation, 2014). Computational models suggest this could be due to increased levels of lexical delay (McMurray et al., 2010). Abnormalities in myelination could result in differing levels of inhibition for targets and competitors. Techniques like MEG would be better suited to test this hypothesis.

There has also been significant theoretical debate about whether DLD represents the tail end of a spectrum of language ability or if represents a group or several subgroups that are biologically distinguishable from those with typical language ability (Tomblin, 2011). Broadly speaking, there is now consensus in the field that these children represent the extreme end of a distribution (Bishop et al., 2016; Lancaster & Camarata, 2019; Tomblin & Zhang, 1999). We examined if our data was consistent with this idea by examining the distribution of our data. When plotted, the average values of striatal myelination for

individual participants indicated overlap between the DLD and TD groups (see Figure 1). So not all children with DLD had values outside of the range of values for the TD groups, as might be expected if we were observing a distinct sub-type. However, our continuous analysis does indicate that those who have the lowest values for striatal myelination have the lowest language factor scores (see Figure 4). Our limited exploration of how these measures are distributed both within the DLD group and across the cohort is consistent with the idea that DLD is a spectrum disorder. Even so, further theoretically driven analyses that relate specific behaviour to these quantitative measures of brain microstructure might provide insight into whether there are distinct sub-types within DLD. Future studies should focus on understanding the full spectrum of DLD using large datasets, establishing if there are distinct brain-behaviour relationships within this broad category.

Limitations and future directions

The differences we see here are observed at a group level. In other words, as noted above, lower MT or R1 values in the caudate nuclei are not observed in every child with DLD. In future studies, we are keen to use structural connectivity analyses, as they will allow us to understand how differences across a network of brain areas may make children susceptible to DLD. It is also unclear whether the neural changes we observe are the cause of DLD, or a consequence of having a language disorder. Longitudinal studies where children are followed over time are the best way to shed light on this issue. A pertinent issue when considering longitudinal studies is the amount of data we were able to retain from children we tested (between approximately 65-80% based on whether children were in the TD, HSL or DLD group, see Methods). We were concerned that systematic biases might affect retention, with children with severe language problems being more likely to be excluded. Consequently, in our sample, we assessed if there were any differences between children whose scans were excluded, relative to those we retained for our analysis. We found that children with DLD whose scans were excluded were younger than those who were selected, but they were not more severely affected in terms of their language learning (Supplementary file 1c). However, it is worth noting our analysis does not account for the children who we recruited but were unable to scan. Dropout and data quality are factors to consider if scanning younger children with this protocol. Finally, the relationship between structure and function is complex. We need to understand how these structural differences might affect specific aspects of function. For instance, we did not observe differences in dorsal striatal activity for a simple language task of verb generation in the same group of children (Krishnan et al., 2020). This may signal

the task was not sensitive to differences, or alternately, that we need to tap different aspects of language processing.

Summary and conclusions

Understanding the neural basis of DLD is particularly challenging given the developmental nature of the disorder, as well as the lack of appropriate animal models for understanding language. Novel semiquantitative MPM protocols allow us an unparalleled *in vivo* method to investigate microstructural neural changes in these children. Our findings using this protocol suggest that the caudate nucleus, as well as regions in the wider speech and language network, show alterations in myelin in children with DLD. These findings strongly point to a role for the striatum in the development of DLD. This role is likely to be in the learning of habits and sequences, but future work is necessary to test this hypothesis given the anatomical localisation in our study. Additionally, myelin patterns can change over development, and myelination can be observed after successful training. In next steps, it is important to assess whether these differences in myelin persist over development in DLD, and if they can be targeted through training using behavioural interventions.

Methods

Ethics

This study was approved by the Medical Sciences Interdivisional Research Ethics Committee at the University of Oxford (R55835/RE002). Before enrolling participants in the study, we obtained written informed consent from parents/guardians, and written assent from children.

Participants

As part of the Oxford BOLD study, we recruited and tested 175 children between the ages of 10 and 15 years. All children had to meet certain inclusion/exclusion criteria; specifically, they had to have: (i) normal hearing (defined as passing audiometric screening at 25 dB at 500 Hz, 1000 Hz, and 2000 Hz, in the better ear); (ii) a nonverbal IQ > 70 (assessed using the WISC-IV Matrix Reasoning and Block Design Tests – Wechsler, 2004); and (iii) have grown up in the UK speaking English. Children were excluded if they had another neurodevelopmental disorder such as autism or attention-deficit hyperactivity disorder, or history of neurological disorder. Participants who met inclusionary/exclusionary criteria were categorised as having DLD if they presented with a history of language problems and scored

at least 1 SD below the normative mean on two or more standardised tests of language ability. Children were categorised as HSL if they presented with a history of speech and language problems but did not meet criteria for DLD. Those who were categorised as TD had no history of speech and language problems. If these children scored 1 SD or more below the mean on more than one standardised test score of language ability, they were excluded from the TD group. Of the 175 children we recruited, a total of 162 children completed both behavioural testing and MRI scans and met our inclusionary criteria (77 typically developing children, 57 children with DLD, and 28 who had a history of speech and language difficulties (HSL) but did not meet our criteria for DLD at time of testing), for further details, see (Krishnan et al., 2020). From this sample, we acquired MPM data in 72 TD children, 51 children with DLD, and in 25 children with HSL.

Data acquisition

MR data were collected with a 3T Siemens Prisma scanner with a 32-channel head coil. Participants wore noise-cancelling headphones (Optoacoustics OptoActive II Active Noise Cancelling Headphones). Foam padding was placed around the head for comfort and to restrict movement; the headphones were held in place with inflatable pads.

Whole brain images at an isotropic resolution of 1mm were obtained using an MPM quantitative imaging protocol (Lutti et al., 2014; Weiskopf et al., 2013). This protocol consisted of the acquisition of three multi-echo gradient acquisitions with proton density (PD), T1, or MT weighting. Each acquisition had a TR of 25ms, field of view = 256 x 224 x 176 mm^3, readout bandwidth 488 Hz/pixel, and slab rotation of 30^0. Flip angle for MT and PD weighted acquisitions was 6°, and 21° for T1 weighted acquisitions. MT weighting was achieved by using a Gaussian RF pulse 2kHz off resonance with 4ms duration and a nominal flip angle of 220° prior to excitation. To speed up data acquisition, a GRAPPA acceleration factor of 2 was applied, with 40 references lines in each phase encoding direction. Eight echoes were acquired for the T1 and PD weighted contrasts, and six echoes were acquired for the MT contrast. Each sequence took approximately 5 minutes to acquire. In addition, data to calculate a radiofrequency (RF) transmit field map was acquired at the start of the session, using a 3D echo-planar imaging spin-echo/stimulated echo method (Lutti et al. 2010, 2012; FOV = 256 × 192 × 192 mm^3, matrix = 64 × 64 × 48 mm^3, TE = 39.06, mixing time = 33.8 ms, TR = 500 ms, nominal α varying from 115° to 65° in steps of

5°, acquisition time 4 min 24 s). In total, the MPM protocol took approximately 20 minutes to acquire.

We did collect other imaging data as part of the Oxford BOLD study, including fMRI data (Krishnan et al., 2020). Notably, we also obtained a T1-weighted MPRAGE scan (magnetization prepared low angle spoiled gradient echo, TR 1900ms, TE 3.97ms, flip angle 8°, field of view 208 × 256 × 256 mm) with 1 mm in-plane resolution and 1 mm slice thickness in this set of scans.

Procedure

MPM data were collected at the end of the scanning session. The session also included two task fMRI scans, a resting state scan, and a diffusion weighted scan; these data are not reported here. During the MPM scans, participants were given the option of either closing their eyes or watching an animated film; nearly all participants chose the film.

Participants also completed a comprehensive neuropsychological battery outside of the scanner, focusing on their linguistic and cognitive abilities (see Krishnan et al., 2020 for further details). In brief, language ability was assessed using five tests, assessing aspects of expressive and receptive grammar, narrative, and vocabulary. Specifically, grammatical comprehension was assessed using the Test for Reception of Grammar – 2 or its electronic counterpart (TROG-E, Bishop, 2005). Expressive grammar was evaluated using the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals – 4th Edition (CELF-4; Semel et al., 2004). Children's narrative production and comprehension was assessed using the Expression, Reception and Recall of Narrative Instrument (ERNNI; Bishop, 2004). Receptive and expressive vocabulary were assessed using the Receptive One-Word Picture Vocabulary Test – 4th Edition (ROWPVT-4; Martin and Brownell, 2011b) and Expressive One-Word Picture Vocabulary Test – 4th Edition (EOWPVT-4; Martin and Brownell, 2011a) respectively. In addition to the language measures, children also completed the phonological decoding and sight word reading efficiency subtests of the Test Of Word Reading Efficiency (TOWRE; Torgesen et al., 1999); the block design, matrix reasoning, and coding subtests of the Wechsler Intelligence Scale for Children - 4th Edition (WISC-IV; Wechsler, 2004), the forward and backward digit span subtests, as well as the word lists subtest, from the Children's Memory Scale (CMS; Cohen, 1997), a nonword repetition test (Norbury et al.,

2016), and the oromotor sequences sub-test of the NEuroPSYchology (NEPSY) test battery (Korkman et al., 1998).

Data pre-processing

Data were processed using the hMRI toolbox within SPM12 (Balteau et al., 2018; Tabelow et al., 2019). The default toolbox settings were used. This processing results in the MT saturation, R1 and R2* maps, which index different aspects of tissue microstructure. Briefly, R1 (1/T1) maps were estimated from the PD and T1 weighted images using the process described in Weiskopf et al. (2013), extended by using correction for radio-frequency transmit field inhomogeneities and imperfect spoiling. Regression of the log signal from the signal decay over echoes across all three MPM contrasts was used to calculate a map of R2* (=1/T2*) (Weiskopf et al., 2013). RF transmit field maps were calculated from the 3D EPI acquisition and corrected for off-resonance effects as described in (Lutti et al., 2012). The semi-quantitative MT saturation parameter (MTsat) calculated is relatively robust against differences in relaxation times and RF transmit and receive field inhomogeneities, and small residual higher dependencies are further corrected for within the toolbox.

Using quality assessment metrics obtained from the toolbox, we removed images where the SD R2* (a measure of image degradation) was greater than 3 times the interquartile range (IQR) from the group mean. Data from 9 children with DLD and 2 TD children were excluded on this basis. We also removed scans where interscan movement exceeded 2mm. Data from a further 4 children with DLD, 5 TD children, and 1 child with HSL were excluded on this basis. We then conducted a visual inspection of the R1, MTsat, and R2* maps, and excluded data from a further 5 children with DLD, 4 children with HSL, and 9 TD children where image artifacts were observed. We retained data from 56 TD children, 33 children with DLD, and 20 children with HSL. This equates to data retention of 77.78% in the TD population, 64.71% in the DLD population, and 80% in the HSL population.

Using pipelines implemented in the hMRI toolbox, MTsat maps from each participant were further segmented into grey and white matter probability maps. These grey and white matter maps were used to create a DARTEL template. Each participants' MTsat, R1 and R2* maps were registered to this DARTEL template and were then normalised to a standard MNI template. A tissue weighted smoothing kernel of 6-mm full-width-at-half-maximum was

applied using the voxel-based quantification approach (Draganski et al., 2011)., which aims to preserve quantitative values for interpretation.

Data analyses

We analysed group differences in MTsat, R1 and R2* maps using FSL's *randomise* tool using 5000 permutations. For assessing statistical differences across groups, we employed threshold-free cluster enhancement, setting $p<0.05$ as our threshold. Data from ROIs of interest were extracted using *fslstats*, and further analyses were carried out using R (R Core Team, 2020).

For the behavioural data, we constructed factor scores for use in continuous analyses to minimise the number of comparisons in statistical tests. These factors were based on analysis of the whole cohort of children who contributed behavioural data to our study and therefore includes data from children who did not complete the MPM scans and children in whom we excluded MPM data as describe above. The measures from the language and memory tests described above were entered into a pre-registered factor analysis to identify the best weighted combination of measures to give a language factor score, and a memory factor score. The approach we adopted to factor analysis was E-CFA (Brown, 2006), implemented in *lavaan* (Rosseel, 2012) in the R programming language (R Core Team, 2020). E-CFA is a hybrid exploratory-confirmatory approach to factor analysis where a model is specified with an 'anchor' measure or two anchor measures. As anchor measures, we used the list learning standard score from the CMS for the memory factor, and expressive vocabulary for the language factor. We planned to compare this two-factor model to a single factor model accounting for language proficiency alone. However, our two preregistered models were not a good fit to the data. Consequently, as detailed in Krishnan et al. (2020), we accounted for strong correlations between expressive and receptive vocabulary scores, as well as the two narrative production measures in modified models, and found the modified two-factor model to be a better fit to the data than the modified single-factor model. We consequently derived language and memory proficiency scores using this modified two-factor model.

Data availability

The data that support the findings of this study are openly available on the OSF (https://osf.io/d93gq/?view_only=c48c989c574a49cba6eba1c413f185bb). Statistical maps can also be viewed on Neurovault (https://neurovault.org/collections/DUGBDBPH/)

Acknowledgements

We thank all of our participants and their families, without whom this work would not be possible. We would also like to acknowledge the many individuals and organisations that helped us with recruitment (https://boldstudy.wordpress.com/acknowledgements/). We especially thank Professor Dorothy Bishop for her support, insight, and discussions throughout OxBOLD. We also thank members of the Wellcome Centre for Integrative Neuroimaging, especially the MRI team at the Oxford Centre for Human Brain Activity: Sebastian Rieger, Juliet Semple, Nicky Aikin, Nicola Filippini, Eniko Zsoldos, and Emily Hinson. We are grateful to Professor Fred Dick, as well as members of the Oxford Speech and Brain Group for helpful discussions and support.

Funding

The Oxford Brain Organisation in Language Development or OxBOLD study was funded by the Medical Research Council MR/P024149/1 and supported by the NIHR Oxford Health Biomedical Research Centre. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z).

Competing interests

The authors report no competing interests.

References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Neuroscience*, *9*(1), 357–381. https://doi.org/10.1146/annurev.ne.09.030186.002041

Allen, M., Glen, J. C., Müllensiefen, D., Schwarzkopf, D. S., Fardo, F., Frank, D., Callaghan, M. F., & Rees, G. (2017). Metacognitive ability correlates with hippocampal and prefrontal microstructure. *NeuroImage*, *149*, 415–423. https://doi.org/10.1016/j.neuroimage.2017.02.008

Argyropoulos, G. P. D., Watkins, K. E., Belton-Pagnamenta, E., Liégeois, F., Saleem, K. S., Mishkin, M., & Vargha-Khadem, F. (2019). Neocerebellar Crus I Abnormalities Associated with a Speech and Language Disorder Due to a Mutation in FOXP2. *The Cerebellum*, *18*(3), 309–319. https://doi.org/10.1007/s12311-018-0989-3

Asaridou, S. S., Demir-Lira, Ö. E., Goldin-Meadow, S., Levine, S. C., & Small, S. L. (2020). Language development and brain reorganization in a child born without the left hemisphere. *Cortex*, *127*, 290–312. https://doi.org/10.1016/j.cortex.2020.02.006

Ayyash, D., Malik-Moraleda, S., Gallée, J., Affourtit, J., Hoffman, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2021). *The universal language network: A cross-linguistic investigation spanning 45 languages and 11 language families* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.07.28.454040

Badcock, N. A., Bishop, D. V. M., Hardiman, M. J., Barry, J. G., & Watkins, K. E. (2012). Co-localisation of abnormal brain structure and function in specific language impairment. *Brain and Language*, *120*(3), 310–320. https://doi.org/10.1016/j.bandl.2011.10.006

Balteau, E., Tabelow, K., Ashburner, J., Callaghan, M. F., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Lutti, A., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Weiskopf, N., Ziegler, G., & Mohammadi, S. (2018). *HMRI – A toolbox for using quantitative MRI in neuroscience and clinical research*. 37.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology*, *40*(3), 293–310. https://doi.org/10.1002/dev.10034

Belton, E., Salmond, C. H., Watkins, K. E., Vargha□Khadem, F., & Gadian, D. G. (2003). Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Human Brain Mapping*, *18*(3), 194–200. https://doi.org/10.1002/hbm.10093

Bishop, D. V. M. (2004). *Expression, Reception and Recall of Narrative Instrument: ERRNI.* Harcourt assessment.

Bishop, D. V. M. (2005). *Test for Reception of Grammar, Version 2—Electronic (TROG–E).* (2nd ed.). The Psychological Corporation.

Bishop, D. V. M., & Hsu, H. J. (2015). The declarative system in children with specific language impairment: A comparison of meaningful and meaningless auditory-visual paired associate learning. *BMC Psychology*, *3*(1), 3. https://doi.org/10.1186/s40359-015-0062-7

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & and the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLOS ONE*, *11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327–332. https://doi.org/10.1038/nature11911

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* Guilford Press.

Callaghan, M. F., Freund, P., Draganski, B., Anderson, E., Cappelletti, M., Chowdhury, R., Diedrichsen, J., FitzGerald, T. H. B., Smittenaar, P., Helms, G., Lutti, A., & Weiskopf, N. (2014). Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. *Neurobiology of Aging*, *35*(8), 1862–1872. https://doi.org/10.1016/j.neurobiolaging.2014.02.008

Carey, D., Caprini, F., Allen, M., Lutti, A., Weiskopf, N., Rees, G., Callaghan, M. F., & Dick, F. (2018). Quantitative MRI provides markers of intra-, inter-regional, and age-related differences in young adult cortical microstructure. *NeuroImage*, *182*, 429–440. https://doi.org/10.1016/j.neuroimage.2017.11.066

Carey, D., Krishnan, S., Callaghan, M. F., Sereno, M. I., & Dick, F. (2017). Functional and Quantitative MRI Mapping of Somatomotor Representations of Human Supralaryngeal Vocal Tract. *Cerebral Cortex*, *27*(1), 265–278. https://doi.org/10.1093/cercor/bhw393

Choi, E. Y., Yeo, B. T. T., & Buckner, R. L. (2012). The organization of the human striatum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *108*(8), 2242–2263. https://doi.org/10.1152/jn.00270.2012

Clark, I. A., Callaghan, M. F., Weiskopf, N., & Maguire, E. A. (2020). Hippocampal grey matter tissue microstructure does not explain individual differences in hippocampal-dependent task performance. *BioRxiv*, 2020.08.18.255992. https://doi.org/10.1101/2020.08.18.255992

Cler, G., Krishnan, S., Papp, D., Wiltshire, C., Chesters, J., & Watkins, K. E. (2021). *Elevated iron concentration in putamen and cortical speech motor network in developmental stuttering*. OSF Preprints. https://doi.org/10.31219/osf.io/ypq8s

Cohen, M. (1997). *Children's memory scale.* Psychological Corporation.

Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., & Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language & Communication Disorders*, *53*(2), 237–255. https://doi.org/10.1111/1460-6984.12338

Dick, F., Taylor Tierney, A., Lutti, A., Josephs, O., Sereno, M. I., & Weiskopf, N. (2012). In Vivo Functional and Myeloarchitectonic Mapping of Human Primary Auditory Areas. *Journal of Neuroscience*, *32*(46), 16095–16105. https://doi.org/10.1523/JNEUROSCI.1712-12.2012

Draganski, B., Ashburner, J., Hutton, C., Kherif, F., Frackowiak, R. S. J., Helms, G., & Weiskopf, N. (2011). Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *NeuroImage*, *55*(4), 1423–1434. https://doi.org/10.1016/j.neuroimage.2011.01.052

Dronkers, N. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*, 159–161.

Earle, F. S., & Ullman, M. T. (2021). Deficits of Learning in Procedural Memory and Consolidation in Declarative Memory in Adults With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *64*(2), 531–541. https://doi.org/10.1044/2020_JSLHR-20-00292

Edwards, L. J., Kirilina, E., Mohammadi, S., & Weiskopf, N. (2018). Microstructural imaging of human neocortex in vivo. *NeuroImage*, *182*, 184–206. https://doi.org/10.1016/j.neuroimage.2018.02.055

Eising, E., Mirza-Schreiber, N., de Zeeuw, E. L., Wang, C. A., Truong, D. T., Allegrini, A. G., Shapland, C. Y., Zhu, G., Wigg, K. G., Gerritse, M., Molz, B., Alagöz, G., Gialluisi, A., Abbondanza, F., Rimfeld, K., van Donkelaar, M., Liao, Z., Jansen, P. R., Andlauer, T. F. M., … Fisher, S. E. (2021). *Genome-wide association analyses of*

individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people [Preprint]. Genomics. https://doi.org/10.1101/2021.11.04.466897

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428–16433. https://doi.org/10.1073/pnas.1112937108

Freund, P., Weiskopf, N., Ashburner, J., Wolf, K., Sutter, R., Altmann, D. R., Friston, K., Thompson, A., & Curt, A. (2013). MRI investigation of the sensorimotor cortex and the corticospinal tract after acute spinal cord injury: A prospective longitudinal study. *The Lancet Neurology*, *12*(9), 873–881. https://doi.org/10.1016/S1474-4422(13)70146-7

Gauger, L. M., Lombardino, L. J., & Leonard, C. M. (1997). Brain morphology in children with specific language impairment. *Journal of Speech, Language, and Hearing Research; Rockville*, *40*(6), 1272–1284. https://search.proquest.com/docview/232349104/abstract/BD6AB18BF9924CCBPQ/1

Girbau-Massana, D., Garcia-Marti, G., Marti-Bonmati, L., & Schwartz, R. G. (2014). Gray–white matter and cerebrospinal fluid volume differences in children with Specific Language Impairment and/or Reading Disability. *Neuropsychologia*, *56*, 90–100. https://doi.org/10.1016/j.neuropsychologia.2014.01.004

Glasser, M. F., & Essen, D. C. V. (2011). Mapping Human Cortical Areas In Vivo Based on Myelin Content as Revealed by T1- and T2-Weighted MRI. *Journal of Neuroscience*, *31*(32), 11597–11616. https://doi.org/10.1523/JNEUROSCI.2180-11.2011

Graybiel, A. M., & Grafton, S. T. (2015). The Striatum: Where Skills and Habits Meet. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021691. https://doi.org/10.1101/cshperspect.a021691

Halliday, L. F., Tuomainen, O., & Rosen, S. (2017). Auditory processing deficits are sometimes necessary and sometimes sufficient for language difficulties in children: Evidence from mild to moderate sensorineural hearing loss. *Cognition*, *166*, 139–151. https://doi.org/10.1016/j.cognition.2017.04.014

Herbert, M. R., Ziegler, D. A., Makris, N., Bakardjiev, A., Hodgson, J., Adrien, K. T., Kennedy, D. N., Filipek, P. A., & Caviness, V. S. (2003). Larger brain and white matter volumes in children with developmental language disorder. *Developmental Science*, *6*(4), F11–F22. https://doi.org/10.1111/1467-7687.00291

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hill, E. L. (2001). Non-specific nature of specific language impairment: A review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, *36*(2), 149–171. https://doi.org/10.1080/13682820010019874

Hsu, H. J., & Bishop, D. V. M. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science*, *17*(3), 352–365. https://doi.org/10.1111/desc.12125

Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex—Developmental changes and effects of aging. *Brain Research*, *163*(2), 195–205. https://doi.org/10.1016/0006-8993(79)90349-4

Jackson, E., Leit, ão S., Claessen, M., & Boyes, M. (2020). Working, Declarative, and Procedural Memory in Children With Developmental Language Disorder. *Journal of*

Speech, Language, and Hearing Research, *63*(12), 4162–4178. https://doi.org/10.1044/2020_JSLHR-20-00135

Jancke, L., Siegenthaler, T., Preis, S., & Steinmetz, H. (2007). Decreased white-matter density in a left-sided fronto-temporal network in children with developmental language disorder: Evidence for anatomical anomalies in a motor-language network. *Brain and Language*, *102*(1), 91–98. https://doi.org/10.1016/j.bandl.2006.08.003

Jarbo, K., & Verstynen, T. D. (2015). Converging Structural and Functional Connectivity of Orbitofrontal, Dorsolateral Prefrontal, and Posterior Parietal Cortex in the Human Striatum. *The Journal of Neuroscience*, *35*(9), 3865–3878. https://doi.org/10.1523/JNEUROSCI.2636-14.2015

Jernigan, T. L., Hesselink, J. R., Sowell, E., & Tallal, P. A. (1991). Cerebral structure on magnetic resonance imaging in language- and learning-impaired children. *Archives of Neurology*, *48*(5), 539–545.

Kail, R. (1994). A Method for Studying the Generalized Slowing Hypothesis in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *37*(2), 418–421. https://doi.org/10.1044/jshr.3702.418

Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY: Neuropsychological assessment of Children*. Psychological Corporation.

Krishnan, S., Asaridou, S. S., Cler, G. J., Smith, H. J., Willis, H. E., Healy, M. P., Thompson, P. A., Bishop, D. V. M., & Watkins, K. E. (2020). Functional organisation for verb generation in children with developmental language disorder. *NeuroImage*, 117599. https://doi.org/10.1016/j.neuroimage.2020.117599

Krishnan, S., Watkins, K. E., & Bishop, D. V. M. (2016). Neurobiological Basis of Language Learning Difficulties. *Trends in Cognitive Sciences*, *20*(9), 701–714. https://doi.org/10.1016/j.tics.2016.06.012

Kurth, F., Luders, E., Pigdon, L., Conti-Ramsden, G., Reilly, S., & Morgan, A. T. (2018). Altered gray matter volumes in language-associated regions in children with developmental language disorder and speech sound disorder. *Developmental Psychobiology*, *60*(7), 814–824. https://doi.org/10.1002/dev.21762

Kwon, D., Pfefferbaum, A., Sullivan, E. V., & Pohl, K. M. (2020). Regional growth trajectories of cortical myelination in adolescents and young adults: Longitudinal validation and functional correlates. *Brain Imaging and Behavior*, *14*(1), 242–266. https://doi.org/10.1007/s11682-018-9980-3

Lai, C. S. L. (2003). FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain*, *126*(11), 2455–2462. https://doi.org/10.1093/brain/awg247

Lancaster, H. S., & Camarata, S. (2019). Reconceptualizing developmental language disorder as a spectrum disorder: Issues and evidence. *International Journal of Language & Communication Disorders*, *54*(1), 79–94. https://doi.org/10.1111/1460-6984.12433

Langkammer, C., Krebs, N., Goessler, W., Scheurer, E., Ebner, F., Yen, K., Fazekas, F., & Ropele, S. (2010). Quantitative MR Imaging of Brain Iron: A Postmortem Validation Study. *Radiology*, *257*(2), 455–462. https://doi.org/10.1148/radiol.10100495

Lee, J. C., Dick, A. S., & Tomblin, J. B. (2020). Altered brain structures in the dorsal and ventral language pathways in individuals with and without developmental language disorder (DLD). *Brain Imaging and Behavior*. https://doi.org/10.1007/s11682-019-00209-1

Lee, J. C., Nopoulos, P. C., & Bruce Tomblin, J. (2013). Abnormal subcortical components of the corticostriatal system in young adults with DLI: A combined structural MRI and DTI study. *Neuropsychologia*, *51*(11), 2154–2161. https://doi.org/10.1016/j.neuropsychologia.2013.07.011

Lehéricy, S., Ducros, M., Van De Moortele, P.-F., Francois, C., Thivard, L., Poupon, C., Swindale, N., Ugurbil, K., & Kim, D.-S. (2004). Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans: DTI Corticostriatal Fibers. *Annals of Neurology*, *55*(4), 522–529. https://doi.org/10.1002/ana.20030

Leutritz, T., Seif, M., Helms, G., Samson, R. S., Curt, A., Freund, P., & Weiskopf, N. (2020). Multiparameter mapping of relaxation (R1 , R2 *), proton density and magnetization transfer saturation at 3 T: A multicenter dual□vendor reproducibility and repeatability study. *Human Brain Mapping*, *41*(15), 4232–4247. https://doi.org/10.1002/hbm.25122

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. *Trends in Neurosciences*, *39*(8), 527–542. https://doi.org/10.1016/j.tins.2016.06.003

Lorio, S., Kherif, F., Ruef, A., Melie□Garcia, L., Frackowiak, R., Ashburner, J., Helms, G., Lutti, A., & Draganski, B. (2016). Neurobiological origin of spurious brain morphological changes: A quantitative MRI study. *Human Brain Mapping*, *37*(5), 1801–1815. https://doi.org/10.1002/hbm.23137

Lorio, S., Lutti, A., Kherif, F., Ruef, A., Dukart, J., Chowdhury, R., Frackowiak, R. S., Ashburner, J., Helms, G., Weiskopf, N., & Draganski, B. (2014). Disentangling in vivo the effects of iron content and atrophy on the ageing human brain. *NeuroImage*, *103*, 280–289. https://doi.org/10.1016/j.neuroimage.2014.09.044

Lozano, R., Vino, A., Lozano, C., Fisher, S. E., & Deriziotis, P. (2015). A de novo FOXP1 variant in a patient with autism, intellectual disability and severe speech and language impairment. *European Journal of Human Genetics*, *23*(12), 1702–1707. https://doi.org/10.1038/ejhg.2015.66

Lum, J. A. G., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex*, *51*, 1–10. https://doi.org/10.1016/j.cortex.2013.10.011

Lutti, A., Dick, F., Sereno, M. I., & Weiskopf, N. (2014). Using high-resolution quantitative mapping of R1 as an index of cortical myelination. *NeuroImage*, *93*, 176–188. https://doi.org/10.1016/j.neuroimage.2013.06.005

Manara, R., Ponticorvo, S., Tartaglione, I., Femina, G., Elefante, A., Russo, C., Carafa, P. A., Cirillo, M., Casale, M., Ciancio, A., Di Concilio, R., De Michele, E., Weiskopf, N., Di Salle, F., Perrotta, S., & Esposito, F. (2019). Brain iron content in systemic iron overload: A beta-thalassemia quantitative MRI study. *NeuroImage: Clinical*, *24*, 102058. https://doi.org/10.1016/j.nicl.2019.102058

Martin, N. A., & Brownell, R. (2011a). *Expressive One-Word Picture Vocabulary Test-4 (EOWPVT-4).* (4th ed.). Academic Therapy Publications.

Martin, N. A., & Brownell, R. (2011b). *Receptive One-Word Picture Vocabulary Test-4 (ROWPVT-4).* (4th ed.). Academic Therapy Publications.

Mayes, A. K., Reilly, S., & Morgan, A. T. (2015). Neural correlates of childhood language disorder: A systematic review. *Developmental Medicine & Child Neurology*, *57*(8), 706–717. https://doi.org/10.1111/dmcn.12714

McArthur, G. M., & Bishop, D. V. M. (2004). Which People with Specific Language Impairment have Auditory Processing Deficits? *Cognitive Neuropsychology*, *21*(1), 79–94. https://doi.org/10.1080/02643290342000087

McGregor, K. K., Gordon, K., Eden, N., Arbisi, -Kelm Tim, & Oleson, J. (2017). Encoding Deficits Impede Word Learning and Memory in Adults With Developmental

Language Disorders. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2891–2905. https://doi.org/10.1044/2017_JSLHR-L-17-0031

McMurray, B., Samelson, V. M., Lee, S. H., & Bruce Tomblin, J. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of Processing in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *44*(2), 416–433. https://doi.org/10.1044/1092-4388(2001/034)

Mountford, H. S., Braden, R., Newbury, D. F., & Morgan, A. T. (2022). The Genetic and Molecular Basis of Developmental Language Disorder: A Review. *Children*, *9*(5), 586. https://doi.org/10.3390/children9050586

Nation, K. (2014). Lexical learning and lexical processing in children with developmental language impairments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1634), 20120387. https://doi.org/10.1098/rstb.2012.0387

Natu, V. S., Gomez, J., Barnett, M., Jeska, B., Kirilina, E., Jaeger, C., Zhen, Z., Cox, S., Weiner, K. S., Weiskopf, N., & Grill-Spector, K. (2019). Apparent thinning of human visual cortex during childhood is associated with myelination. *Proceedings of the National Academy of Sciences*, *116*(41), 20750–20759. https://doi.org/10.1073/pnas.1904931116

Newbury, D. F., Winchester, L., Addis, L., Paracchini, S., Buckingham, L.-L., Clark, A., Cohen, W., Cowie, H., Dworzynski, K., Everitt, A., Goodyer, I. M., Hennessy, E., Kindley, A. D., Miller, L. L., Nasir, J., O'Hare, A., Shaw, D., Simkin, Z., Simonoff, E., … Monaco, A. P. (2009). CMIP and ATP2C2 Modulate Phonological Short-Term Memory in Language Impairment. *The American Journal of Human Genetics*, *85*(2), 264–272. https://doi.org/10.1016/j.ajhg.2009.07.004

Newport, E. L., Landau, B., Seydell-Greenwald, A., Turkeltaub, P. E., Chambers, C. E., Dromerick, A. W., Carpenter, J., Berl, M. M., & Gaillard, W. D. (2017). Revisiting Lenneberg's Hypotheses About Early Developmental Plasticity: Language Organization After Left-Hemisphere Perinatal Stroke. *Biolinguistics*, *11*, 407–422. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6291004/

Nieuwenhuys, R. (2013). The myeloarchitectonic studies on the human cerebral cortex of the Vogt–Vogt school, and their significance for the interpretation of functional neuroimaging data. *Brain Structure and Function*, *218*(2), 303–352. https://doi.org/10.1007/s00429-012-0460-z

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

NSPN Consortium, Ziegler, G., Hauser, T. U., Moutoussis, M., Bullmore, E. T., Goodyer, I. M., Fonagy, P., Jones, P. B., Lindenberger, U., & Dolan, R. J. (2019). Compulsivity and impulsivity traits linked to attenuated developmental frontostriatal myelination trajectories. *Nature Neuroscience*, *22*(6), 992–999. https://doi.org/10.1038/s41593-019-0394-3

Orpella, J., Mas-Herrero, E., Ripollés, P., Marco-Pallarés, J., & Diego-Balaguer, R. de. (2021). Language statistical learning responds to reinforcement learning principles rooted in the striatum. *PLOS Biology*, *19*(9), e3001119. https://doi.org/10.1371/journal.pbio.3001119

Paquola, C., Bethlehem, R. A., Seidlitz, J., Wagstyl, K., Romero-Garcia, R., Whitaker, K. J., Vos de Wael, R., Williams, G. B., NSPN Consortium, Vértes, P. E., Margulies, D. S., Bernhardt, B., & Bullmore, E. T. (2019). Shifts in myeloarchitecture characterise

adolescent development of cortical gradients. *ELife*, *8*, e50482. https://doi.org/10.7554/eLife.50482

Parker, A. J., Krishnan, S., Thompson, P. A., Watkins, K. E., & Bishop, D. (2021). Impact of stimulus variability on understanding of reversible sentences in children with Developmental Language Disorder: Learning vs generalisation. Stage 1 Registered Report. *Language Development Research*. https://doi.org/10.17605/OSF.IO/V3EBF

Paulesu, E., Danelli, L., & Berlingeri, M. (2014). Reading the dyslexic brain: Multiple dysfunctional routes revealed by a new meta-analysis of PET and fMRI activation studies. *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/article/10.3389/fnhum.2014.00830

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, *9*(2), 60–68. https://doi.org/10.1016/j.tics.2004.12.008

Pigdon, L., Willmott, C., Reilly, S., Conti-Ramsden, G., Gaser, C., Connelly, A., & Morgan, A. T. (2019). Grey matter volume in developmental speech and language disorder. *Brain Structure & Function*, *224*(9), 3387–3398. https://doi.org/10.1007/s00429-019-01978-7

Plante, E. (1991). MRI findings in boys with specific language impairment. *Brain and Language*, *41*(1), 52–66. https://doi.org/10.1016/0093-934X(91)90110-M

Preis, S., JÄncke, L., Schittler, P., Huang, Y., & Steinmetz, H. (1998). Normal intrasylvian anatomical asymmetry in children with developmental language disorder. *Neuropsychologia*, *36*(9), 849–855. https://doi.org/10.1016/S0028-3932(98)00033-5

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping*, *29*(11), 1231–1242. https://doi.org/10.1002/hbm.20460

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331

Richlan, F. (2012). Developmental dyslexia: Dysfunction of a left hemisphere reading network. *Frontiers in Human Neuroscience*, *6*. https://www.frontiersin.org/article/10.3389/fnhum.2012.00120

Rosseel, Y. (2012). *lavaan: An R package for structural equation modeling and more Version 0.5-12 (BETA)*. 37.

Sanjeevan, T., & Mainela-Arnold, E. (2019). Characterizing the Motor Skills in Children with Specific Language Impairment. *Folia Phoniatrica et Logopaedica*, *71*(1), 42–55. https://doi.org/10.1159/000493262

Schmierer, K., Scaravilli, F., Altmann, D. R., Barker, G. J., & Miller, D. H. (2004). Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Annals of Neurology*, *56*(3), 407–415. https://doi.org/10.1002/ana.20202

Semel, E. M., Wiig, E. H., & Secord, W. (2004). *Clinical Evaluation of Language Fundamentals.* (4th ed.). PsychCorp.

Sereno, M. I., Lutti, A., Weiskopf, N., & Dick, F. (2013). Mapping the Human Cortical Surface by Combining Quantitative T1 with Retinotopy†. *Cerebral Cortex*, *23*(9), 2261–2268. https://doi.org/10.1093/cercor/bhs213

Shafee, R., Buckner, R. L., & Fischl, B. (2015). Gray matter myelination of 1555 human brains using partial volume corrected MRI images. *NeuroImage*, *105*, 473–485. https://doi.org/10.1016/j.neuroimage.2014.10.054

Simmonds, A. J., Leech, R., Iverson, P., & Wise, R. J. S. (2014). The response of the anterior striatum during adult human vocal learning. *Journal of Neurophysiology*, *112*(4), 792–801. https://doi.org/10.1152/jn.00901.2013

Siugzdaite, R., Bathelt, J., Holmes, J., & Astle, D. E. (2020). Transdiagnostic Brain Mapping in Developmental Disorders. *Current Biology*, *30*(7), 1245-1257.e4. https://doi.org/10.1016/j.cub.2020.01.078

Skipper, J. I., Aliko, S., Brown, S., Jo, Y. J., Lo, S., Molimpakis, E., & Lametti, D. R. (2020). *Reorganization of the neurobiology of language after sentence overlearning* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.09.11.293167

Soriano□Mas, C., Pujol, J., Ortiz, H., Deus, J., López□Sala, A., & Sans, A. (2009). Age-related brain structural alterations in children with specific language impairment. *Human Brain Mapping*, *30*(5), 1626–1636. https://doi.org/10.1002/hbm.20620

Steiger, T. K., Weiskopf, N., & Bunzeck, N. (2016). Iron Level and Myelin Content in the Ventral Striatum Predict Memory Performance in the Aging Brain. *Journal of Neuroscience*, *36*(12), 3552–3558. https://doi.org/10.1523/JNEUROSCI.3617-15.2016

Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M. F., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Lutti, A., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Weiskopf, N., Ziegler, G., & Mohammadi, S. (2019). HMRI – A toolbox for quantitative MRI in neuroscience and clinical research. *NeuroImage*, *194*, 191–210. https://doi.org/10.1016/j.neuroimage.2019.01.029

Thal, D. J., Marchman, V., Stiles, J., Aram, D., Trauner, D., Nass, R., & Bates, E. (1991). Early lexical development in children with focal brain injury. *Brain and Language*, *40*(4), 491–527. https://doi.org/10.1016/0093-934X(91)90145-Q

Tomblin, B. (2011). Co-morbidity of autism and SLI: Kinds, kin and complexity. *International Journal of Language & Communication Disorders*, *46*(2), 127–137. https://doi.org/10.1111/j.1460-6984.2011.00017.x

Tomblin, B., & Zhang, X. (1999). Language Patterns and Etiology in Children with Specific Language Impairment. In *Neurodevelopmental disorders* (p. 361). MIT Press.

Torgesen, J. K., Rashotte, C. E., & Wagner, R. K. (1999). *TOWRE: Test of word reading efficiency*. Pro-ed.

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual Review of Psychology*, *71*(1), 389–417. https://doi.org/10.1146/annurev-psych-122216-011555

Ullman, M. T., & Pierpont, E. I. (2005). Specific Language Impairment is not Specific to Language: The Procedural Deficit Hypothesis. *Cortex*, *41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Vargha-Khadem, F., O'Gorman, A. M., & Watters, G. V. (1985). Aphasia and handedness in relation to hemispheric side, age at injury and severity of cerebral lesion during childhood. *Brain*, *108*(3), 677–696. https://doi.org/10.1093/brain/108.3.677

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., Alarcón, M., Oliver, P. L., Davies, K. E., Geschwind, D. H., Monaco, A. P., & Fisher, S. E. (2008). A Functional Genetic Link between Distinct Developmental Language Disorders. *New England Journal of Medicine*, *359*(22), 2337–2345. https://doi.org/10.1056/NEJMoa0802828

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain*, *125*(3), 452–464. https://doi.org/10.1093/brain/awf058

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., Frackowiak, R. S. J., Mishkin, M., & Gadian, D. G. (2002). MRI analysis of an inherited speech and language disorder: Structural brain abnormalities. *Brain*, *125*(3), 465–478. https://doi.org/10.1093/brain/awf057

Wechsler, D. (2004). *WISC-IV: Wechsler Intelligence Scale for Children: Technical and Interpretive Manual.* Psychological Corporation.

Weiskopf, N., Edwards, L. J., Helms, G., Mohammadi, S., & Kirilina, E. (2021). Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. *Nature Reviews Physics*, 1–19. https://doi.org/10.1038/s42254-021-00326-1

Weiskopf, N., Suckling, J., Williams, G., Correia, M. M., Inkster, B., Tait, R., Ooi, C., Bullmore, E. T., & Lutti, A. (2013). Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: A multi-center validation. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00095

West, G., Melby-Lervåg, M., & Hulme, C. (2021). Is a procedural learning deficit a causal risk factor for developmental language disorder or dyslexia? A meta-analytic review. *Developmental Psychology*, *57*(5), 749–770. https://doi.org/10.1037/dev0001172

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: We see some problems. *Developmental Science*, *21*(2), e12552. https://doi.org/10.1111/desc.12552

Whitaker, K. J., Vértes, P. E., Romero-Garcia, R., Váša, F., Moutoussis, M., Prabhu, G., Weiskopf, N., Callaghan, M. F., Wagstyl, K., Rittman, T., Tait, R., Ooi, C., Suckling, J., Inkster, B., Fonagy, P., Dolan, R. J., Jones, P. B., Goodyer, I. M., the NSPN

Consortium, & Bullmore, E. T. (2016). Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proceedings of the National Academy of Sciences*, *113*(32), 9105–9110. https://doi.org/10.1073/pnas.1601745113

Wise, R., Greene, J., Büchel, C., & Scott, S. (1999). Brain regions involved in articulation. *The Lancet*, *353*(9158), 1057–1061. https://doi.org/10.1016/S0140-6736(98)07491-1

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476. https://doi.org/10.1038/nrn1919

Figure legends

Figure 1.
Brain areas showing reduced MTsat in DLD *Coloured maps resampled and overlaid on the fsaverage surface (reconstructed using the HCP workbench) show TD > DLD differences in MTsat values (whole-brain corrected threshold p<.05). Axial slices (coloured maps are overlaid on an average MTsat image from all participants) reveal additional group differences subcortically in the caudate nuclei. The inset shows a boxplot of MTsat values drawn from the cluster in the caudate nuclei by group, with this cluster split into right and left using a hemispheric mask (orange – DLD, green – TD).*

Figure 2.
Brain areas showing reduced R1 in DLD *Coloured maps resampled and overlaid on the fsaverage surface (reconstructed using the HCP workbench) show TD > DLD differences in R1 maps (whole brain corrected threshold p<.05). Axial slices (coloured maps are overlaid on an average MTsat image from all participants) reveal additional subcortical group differences in the striatum and thalamus bilaterally.*

Figure 3.
Brain areas showing conjoint differences in MTsat and R1 in DLD *Coloured maps resampled and overlaid on the fsaverage surface (reconstructed using the HCP workbench) show convergence of TD > DLD differences in MTsat and R1 maps thresholded at p<0.05. Axial slices (coloured maps are overlaid on an average MTsat image from all participants) reveal the differences subcortically in the dorsal striatum in particular.*

Figure 4.
Correlation between language proficiency and MTsat values in the left caudate nucleus

Figure supplement legends

Figure 1 – figure supplement 1
Relationship between MTsat values in the caudate nucleus by age and total intracranial volume (TIV) in TD children and those with DLD

Figure 1 – figure supplement 2
Whole brain TD > DLD differences (thresholded at p<0.05) when excluding children with nonverbal IQs < 85 in MTsat values

Figure 2 – figure supplement 1
Whole brain TD > DLD differences (thresholded at p<0.05) when excluding children with nonverbal IQs < 85 in R1 values

Figure 3 – figure supplement 1.
Average MTsat, R1, and R2* values in TD and DLD groups *are shown using a coloured heat map (hotter colours show higher values). Maps are overlaid on the fsaverage surface reconstructed using the HCP workbench.*

Figure 4 – figure supplement 1
Whole brain correlation of MPM values with Language Proficiency (thresholded at p<0.05)

Supplemental file legends

Supplementary file 1a. TD > DLD differences in MTsat maps. Nonparametric randomisation analysis with threshold-free cluster enhancement was used to compare groups. A whole brain corrected threshold of $p<.05$ was used.

Supplementary file 1b. TD > DLD conjoint differences in R1 and MTsat

Supplementary file 1c. Differences in age and language scores between the selected and excluded children who were typically-developing (TD) or had developmental language disorder (DLD)