Heterogeneous relationships between white matter and behaviour

Alberto Lazari [*1], Piergiorgio Salvan[1], Michiel Cottaar[1], Daniel Papp[1], Olof Jens van der Werf[3,4], Ainslie Johnstone[1, 5], Zeena-Britt Sanders[1], Cassandra Sampaio-Baptista[1], Nicole Eichert[1], Kentaro Miyamoto[6], Anderson Winkler[7], Martina F. Callaghan[8], Thomas E. Nichols[9], Charlotte J Stagg[1,2,10], Matthew Rushworth[6], Lennart Verhagen[6,11], and Heidi Johansen-Berg [†1]

[1] *Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford*
[2] *Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford*
[3] *Section Brain Stimulation and Cognition, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands*
[4] *Maastricht Brain Imaging Centre (MBIC), Maastricht University, Maastricht, The Netherlands*
[5] *Department of Clinical and Movement Neuroscience, Institute of Neurology, University College London, WC1N 3BG*
[6] *Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford*
[7] *National Institute of Mental Health, National of Health, Bethesda, MD, USA*
[8] *Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, UCL, London, UK*
[9] *Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford*
[10] *MRC Brain Network Dynamics Unit, University of Oxford, Oxford, OX1 3TH, UK*
[11] *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands*

*alberto.lazari@univ.ox.ac.uk
†heidi.johansen-berg@ndcn.ox.ac.uk

1

Abstract

Several studies have established specific relationships between White Matter (WM) and behaviour. However, these studies have typically focussed on fractional anisotropy (FA), a neuroimaging metric that is sensitive to multiple tissue properties, making it difficult to identify what biological aspects of WM may drive such relationships. Here, we carry out a pre-registered assessment of WM-behaviour relationships across multiple behavioural, anatomical and biological domains.

Surprisingly, we find support for predicted relationships between FA and behaviour only in one of three pre-registered tests.

We also find no evidence for consistent multimodal signatures across neuroimaging markers with different biological sensitivity, which suggests there is no common biological substrate for WM-behaviour relationships.

These results demonstrate that FA-behaviour relationships from the literature may not be easily generalisable across domains. They also highlight a broad heterogeneity in WM's relationship with behaviour, indicating that variable biological effects may be shaping their interaction.

Highlights

- Systematic, pre-registered testing of microstructural imaging across modalities to test WM-behaviour relationships.

- Partial support for FA-behaviour relationships hypothesised based on previous literature.

- WM-behaviour relationships are heterogeneous: no common microstructural signature in WM correlates with behaviour.

- Variable biological effects may be shaping WM-behaviour relationships.

Introduction

Once thought to be a static, uninformative relay of information between cortical areas of grey matter, the past decade has shown that WM, and in particular the myelinated structures that dominate it, have more varied functions than previously thought, from trophic support of axons (Fünfschilling et al., 2012; Nave, 2010) to active regulation of physiological and behavioural processes (Kaller et al., 2017; Lazari et al., 2018; Steadman et al., 2019). These basic biology findings have opened the door to the possibility of studying WM as a way to better understand brain physiology and behaviour, and also the chance of targeting WM for therapeutic gain in brain-related disorders.

In humans, much evidence on the role of WM has come from a large body of studies linking behaviour to fractional anisotropy (FA), a metric derived from diffusion weighted imaging that is sensitive to features of WM microstructure (Boekel et al., 2015; Johansen-Berg, 2010; Lazari and Lipp, 2020; Roberts et al., 2013). While these studies have provided seminal evidence for a link between WM and human behaviour, two key issues persist regarding WM-behaviour relationships, and thus limit our knowledge on the role of human WM.

First, many previous studies, especially early studies using FA, have focussed on exploratory analyses of relatively small datasets (Boekel et al., 2015). This means issues like publication bias (Turner et al., 2008) and underpowered analyses (Button et al., 2013) in the published literature may have inadvertently inflated the effect size of FA-behaviour relationships. There-

fore, confirmatory, well-powered evidence on FA-behaviour relationships is needed (Munafò et al., 2017; Nosek et al., 2018).

Second, FA-behaviour relationships are difficult to interpret on a biological level. Diffusion signals are sensitive to a broad range of tissue properties, including myelination levels, fiber orientation, axon diameter, astrocyte and vascular morphology (Farquharson et al., 2013; Sampaio-Baptista and Johansen-Berg, 2017; Stolp et al., 2018). Therefore, a given FA-behaviour correlation could arise from a diversity of microstructural patterns (Zatorre et al., 2012).

In recent years, an increasing number of techniques (Figure 1) have been successfully applied to the study of WM (Heath et al., 2018). As WM is dominated by myelinating oligodendrocytes, many of these techniques have focused on detecting direct signals from myelin or from iron, which is enriched in the cell body of oligodendrocytes. Magnetisation Transfer-based techniques, for example, quantify the fraction of macromolecule-bound water protons, and have been shown to relate strongly to myelination in a number of validation studies (Deloire-Grassin et al., 2000; Dousset et al., 1992, 1995). R2* mapping, on the other hand, quantifies local field distortion caused by iron, and has been confirmed as an iron marker by several validation studies (Langkammer et al., 2010; Sun et al., 2015). R1 has gained attention recently as a quantitative metric for myelination, and although its effectiveness as a WM myelin marker has not been directly tested, it has been shown to detect spatial distributions of myelin in grey matter (Lutti et al., 2014; Stüber et al., 2014). In addition to the development of new MR techniques, new statisti-

5

cal tools, such as joint inference permutation testing (Winkler et al., 2014, 2016), facilitate the integration of MR techniques to clarify the biological interpretation of MRI-measured effects in white matter.

Applying these approaches to studying WM microstructural techniques could be helpful for clarifying the mechanisms behind WM-behaviour relationships. It could allow disentangling whether myelination, oligodendrocytes, or fiber orientation, or a combination of them, are key in driving reported FA-behaviour correlations. In turn, if all WM-behaviour relationships are driven by a common biological mechanism, then establishing recurrent multimodal patterns that correlate with behaviour could uncover it, with powerful implications for future studies looking at WM-behaviour relationships and biomarker development.

To tackle these open questions regarding WM-behaviour relationships, we set out to:

1) Perform systematic, pre-registered testing of FA-behaviour relationships.

2) Perform systematic, pre-registered testing of relationships between behaviour and microstructural imaging across neuroimaging modalities.

3) Identify multimodal microstructural signatures which may provide insights into the underlying biology of WM-behaviour relationships.
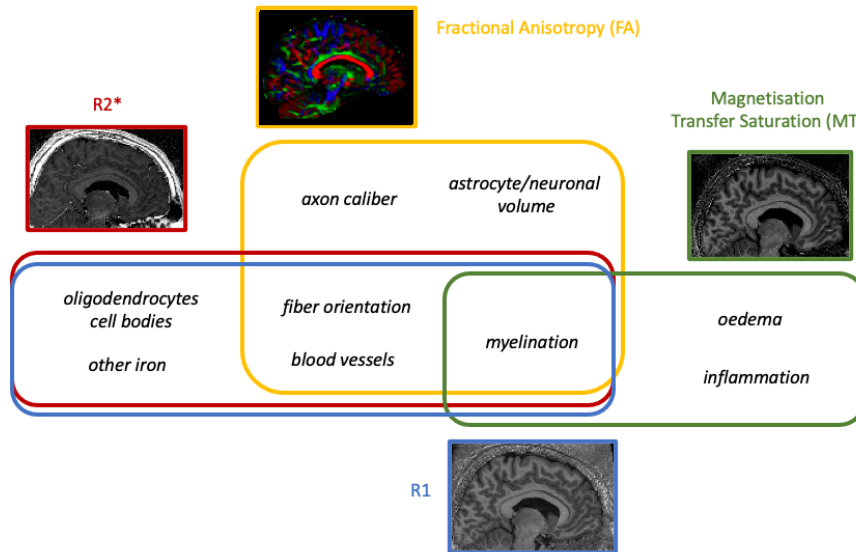
Figure 1: Each neuroimaging modality is sensitive, but not specific, to different features of the biological tissue.
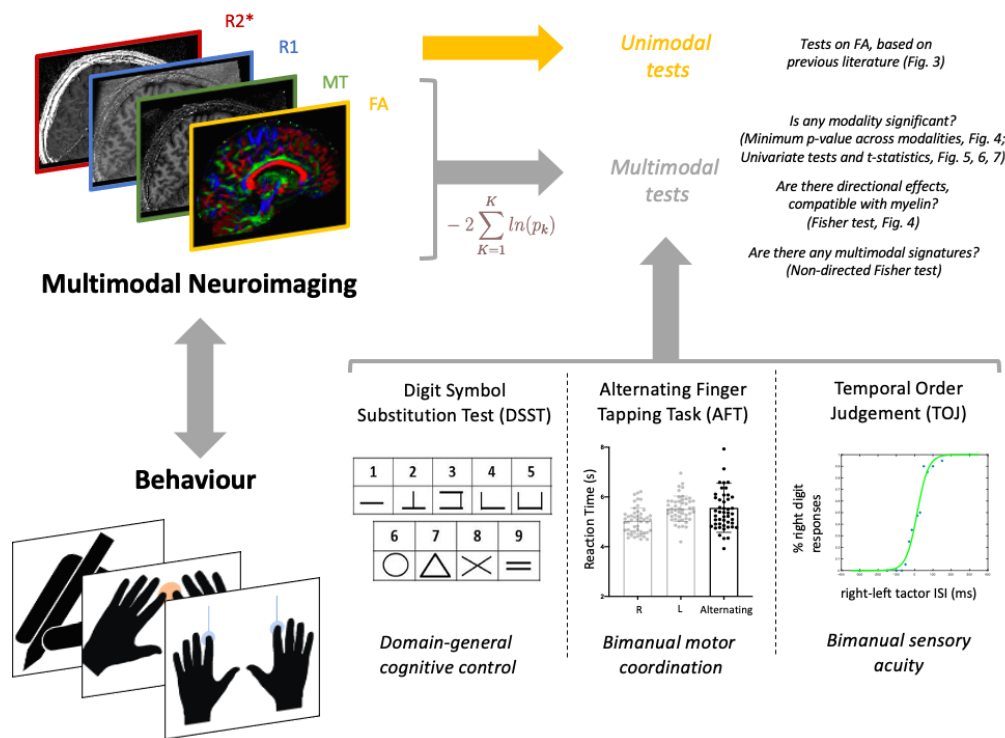
Methods

Figure 2: Summary of MRI and behavioural data acquisitions.

Participants Figure 2 summarises the study design. 50 healthy participants (25 female; aged 18-38 years, mean 26.2 years, median 26 years) underwent a single session of behavioural testing and MRI on the same day. All participants were self-assessed right-handed and their handedness was further assessed through the Edinburgh Handedness Inventory (Oldfield, 1971) (score range 60-100, mean 87.2, median 90). All participants were screened for MRI safety, received monetary compensation for their participation, and gave their informed consent to participate in this study. All study procedures

followed the Declaration of Helsinki, and were reviewed and approved by the local ethics committee at the University of Oxford.

Preregistration Details of the task data collection and analysis plans were pre-registered on the Open Science Framework website (full pre-registration available here: https://osf.io/ar7zs/).

We report here relevant text from the pre-registration: "Overall aim: testing whether previously reported correlations between behavioural measures and fractional anisotropy (FA) measures in long-range projections obtained using diffusion-weighted magnetic resonance imaging (dw-MRI) are related to indices of myelin content obtained using novel quantitative magnetic resonance imaging (qMRI) protocols. To this end, we aim to replicate a sample of previous studies, and extend these FA/behaviour analyses to myelin qMRI/behaviour analyses".

Specific brain/behaviour predictions were made for each task, listed in the analysis section below.

Behavioural tasks A set of behavioural tasks was selected to build on prior studies reporting relationships between behaviour and WM microstructure.

The presence of FA-behaviour relationships has been particularly clear for the corpus callosum and for the cingulum. The cingulum has been often implicated in cognitive control (Bathelt et al., 2019), and cingulum FA has been found to strongly correlate with performance on neuropsychologi-

9

cal tasks (Metzler-Baddeley et al., 2012). The corpus callosum, on the other hand, allows the nodes of the motor network in each hemisphere to communicate with one another, and both positive and negative relationships have been widely reported between callosal FA and various types of bimanual performance ((Johansen-Berg et al., 2007; Muetzel et al., 2008; Sullivan et al., 2001) and (Gooijers and Swinnen, 2014) for a comprehensive review of callosal-bimanual behaviour relationships).

FA-behaviour relationships have also been thoroughly explored in behavioural paradigms beyond the motor system. As mentioned above, bimanual motor performance has been the subject of much literature, and so has bilateral sensory processing. In the visual domain, topographic organisation and visuospatial capacity have both been shown to relate to callosal microstructure (Saenz and Fine, 2010; Todorow et al., 2014). In the auditory domain, relationships have been established between perceptual acuity and WM microstructure, although mostly in pathology (Husain et al., 2011; Lin et al., 2008; Wang et al., 2019). While there have been no previous studies on WM relationships with somatosensory acuity, it would be logical to expect a similar relationship between somatosensory perceptual acuity and microstructure of WM in relevant tracts.

Specifically, we assessed three task domains: (1) testing for a relationship between callosal FA and bimanual motor performance using the Alternating Finger Tapping task aimed to directly replicate a series of previous studies (reviewed by (Gooijers and Swinnen, 2014)); (2) testing for a relationship between cingulum FA and performance using the Digit Symbol Substitution

10

Test aimed to extend a previous finding in older adults (Metzler-Baddeley et al., 2012) to a younger population; (3) testing for a relationship between FA in somatosensory tracts and somatosensory perceptual acuity using the Temporal Order Judgement Task aimed to extend previous findings in the visual and auditory domain, to the sensory system. These three tasks are described in detail below.

Digit Symbol Substitution Test (DSST) A paper-based Digit Symbol Substitution Test (DSST) was conducted as per `https://healthabc.nia. nih.gov/sites/default/files/dsst_0.pdf`. After training on substituting 10 digits for symbols, participants were asked to sequentially fill in the remaining 90 symbol-digit boxes in 90 seconds.

Analysis of the DSST The score was calculated as the total number of symbols filled in correctly by the end of the task. Two participants were identified as outliers (>3 SD away from the mean) and thus excluded from further analyses.

Alternating Finger Tapping (AFT) task The finger tapping task aimed to test the participants' bimanual coordination. The task was based on (Muetzel et al., 2008) and (Pelletier et al., 1993) and ran as follows: three blocks were repeated four times (the first one for training purposes): during the first block, participants were asked to tap their right index finger on a buttonbox (Current Designs, Inc., Philadelphia, PA) 30 times, as fast as they could (right monomanual condition); during the second block, participants

11

were asked to tap their left index finger (left monomanual condition); during the third block, participants were asked to alternate between right and left index finger button presses (bimanual condition). For each block, after the 30 button presses were finished, the total elapsed time was fed back on the computer screen. The experimenter inspected the participant movement by eye to ensure they were correctly switching between fingers and that they were moving the finger rather than the hand. Participant posture and hand position was carefully kept constant throughout all blocks. One participant did not carry out the AFT due to a hardware problem.

Analysis of the AFT task Alternating Finger Condition (AFC) duration was extracted, i.e. average total time needed for 30 taps on the alternating finger condition (Muetzel et al., 2008). Two participants were identified as outliers (>3 SD away from the mean) and thus excluded from further analyses. Total time needed for 30 taps on the monomanual conditions was used as a covariate in group-level analyses, (Pelletier et al., 1993), together with age and gender.

Temporal Order Judgement (TOJ) task The Temporal Order Judgement (TOJ) task aimed to test participants' capacity to discriminate between two closely timed tactile stimuli delivered to the fingertips. The task was based on a previous investigation of the functional activity associated with such behaviour (Kolasinski et al., 2016) and ran as follows. A PC running a PsychoPy script delivered, via a USB 6501 card (National Instruments)

and an amplifier (Tactamp, Dancer Design), two asynchronous pulses to two vibrotactile stimulators (also known as tactors, Dancer Design) positioned within holes in a foam pad. The participant was asked to keep their hands relaxed on the foam pad, with their index fingers gently lying on the tactors. A piece of cardboard was used to block visual input from the tactors; similarly, headphones playing low levels of pink noise were used to block the auditory input from the tactors. Participants performed a two alternative forced choice (2AFC) task and were asked to press on one of two foot pedals, depending on the side of the pulse that they thought had come first. Participants were asked to respond within 2 seconds. If they did not respond within this time then no response was recorded and a new trial was started. They were also instructed that if it was hard to judge which pulse came first, they should just make their best guess. Intervals between pulses ranged from 0 to 300 ms. The task featured a practice session with 10 trials and a full session with 280 trials, for a total duration of roughly 12 minutes.

Analaysis of the TOJ task After trials with no response were discarded, the number of correct pedal responses were plotted as a function of inter-stimulation interval and a logistic regression was fitted to the data. At this stage, six participants were excluded as the logistic regression failed to fit the data correctly. The slope of the curve and the Just Noticeable Difference (JND) were used as key metrics of performance on the task (Kolasinski et al., 2016; Shore et al., 2005).

MRI data collection   Magnetic Resonance Imaging (MRI) data were collected with a 3.0-T Prisma Magnetom Siemens scanner, software version VE11C (Siemens Medical Systems, Erlangen, Germany). Participants were asked to keep their head still and to wear earplugs during scanning in order to reduce the impact of MRI-related noise. The sequences were collected as follows: T1-weighted structural imaging (T1w), resting-state fMRI (rs-fMRI), Multi-Parameter Mapping (MPM) and Diffusion-Weighted Imaging (DWI). MRI scan pre-processing, analysis and statistical comparisons were performed using FMRIB Software Library (FSL, v6.0), except for the MPM quantitative map estimation step which was carried out using the hMRI toolbox implemented in Matlab-based SPM, as described in (Tabelow et al., 2019).

The T1w sequence had a TR of 1900 ms, TE of 3.96 ms, a 1mm isotropic resolution and a large Field of View (FOV, 256 mm^3) to allow for the nose to be included in the image and thus facilitate neuronavigation later on in the paradigm. The sequence used GRAPPA with an acceleration factor of 2.

The diffusion-weighted Echo-planar imaging (EPI)sequence had TR=3070 ms, TE=85 ms, FOV=204mm^3, voxel size=1.5mm isotropic, multiband factor of 4. Diffusion scans were collected for two b-values (500 and 2000 s/mm^2), over 281 directions. An additional 23 volumes were acquired at b=0, 15 in AP phase-encoding direction and 8 in the PA phase-encoding direction.

The MPM protocol (as per (Weiskopf et al., 2013)) included three multi-echo 3D FLASH (fast low-angle shot) scans with varying acquisition param-

14

eters, one RF transmit field map (B1+map) and one static magnetic (B0) field map scan, for a total acquisition time of roughly 22 minutes. To correct for inter-scan motion, position-specific receive coil sensitivity field maps, matched in FOV to the MPM scans, were calculated and corrected for (Papp et al., 2016). The three types of FLASH scans were designed to be predominantly T1-, PD-, or MT-weighted by changing the flip angle and the presence of a pre-pulse: 8 echoes were predominantly Proton Density-weighted (TR = 25ms; flip angle = 6 degrees; TE = 2.3-18.4ms), 8 echoes were predominantly T1-weighted (TR = 25ms; flip angle = 21 degrees; TE = 2.3-18.4ms) and 6 echoes were predominantly Magnetisation Transfer-weighted (MTw, TR = 25ms; flip angle = 6 degrees; TE = 2.3-13.8ms). For MTw scans, excitation was preceded by off-resonance Gaussian MT pulse of 4 ms duration, flip angle of 220 degrees, 2 kHz frequency offset from water resonance. All FLASH scans had 1 mm isotropic resolution and field of view (FOV) of 256x224x176 mm. The B1 map was acquired through an EPI-based sequence featuring spin and stimulated echoes (SE and STE) with 11 nominal flip angles, FOV of 192x192x256 mm and TR of 500 ms. The TE was 37.06 ms, and the mixing time was 33.8 ms. The B0 map was acquired to correct the B1+ map for distortions due to off-resonance effects. The B0 map sequence had a TR of 1020.0 ms, first TE of 10 ms, second TE of 12.46 ms, field of view (FOV) of 192x192x256 mm and read-out bandwidth of 260 Hz/pixel.

MRI preprocessing An automated custom pipeline based on existing FSL tools (Smith et al., 2004) was developed for our diffusion sequence.

15

The topup tool was run on average images of AP b0 volumes and PA b0 volumes. The resulting susceptibility-induced off-resonance field was used as an input for the eddy tool (Andersson and Sotiropoulos, 2016), which was run with options optimised for multiband diffusion data to correct for eddy currents and subject movement. To generate Fractional Anisotropy (FA) maps, a diffusion tensor model was fit to each voxel through DTIFIT.

Magnetisation Transfer saturation (MT), R1 and R2* quantitative maps were estimated through the hMRI toolbox (Tabelow et al., 2019), with default settings including ESTATICS modelling (Weiskopf et al., 2014). In order to register MPM volumes to FA volumes, we used the following steps. Boundary-Based Registration was used to calculate a DWI-to-T1w registration using preprocessed b0 images (with high tissue boundary contrast). A customised pipeline was used to apply the fslreorient2std tool to the MPM maps and register them to T1w space. At this stage, 1 participant was excluded as the MPM-derived maps were heavily corrupted due to movement artefacts; 1 participant was excluded due to lower quality signal in the MPM scan, which resulted in poor registration with other modalities. Once registration matrices for MPM-T1w and DWI-T1w were calculated, they were inverted, concatenated and applied as needed to bring MPM volumes into DWI space with minimal interpolation. Registrations were assessed manually and one participant was excluded due to poor registration across all analyses.

Multimodal voxelwise analysis To bring all volumes into a common space, native FA volumes were skeletonised with Tract-Based Spatial Statistics (TBSS (Smith et al., 2006)), and the skeletonisation transforms were subsequently applied to MPM-to-DWI registered volumes. Group-level analyses were then conducted in skeleton space for all data.

All behavioural performance measures were normalised (through z-scoring, or rank-based inverse-normal transformation if not normally distributed) and correlations between MRI metrics and behaviour were assessed for each behavioural measure separately.

Relevant text from the preregistered analysis plan is as follows:

Cingulum and DSST: "We aim to replicate a reported relationship between [...] number of substituted digits in the Digit Substitution test and cingulum FA (Metzler-Baddeley et al., 2012) in a sample of participants of on average a younger age range than previously reported, and to extend the protocol to investigate qMRI myelin/behaviour relationships."

Callosum and AFT: "We aim to replicate a reported relationship between callosal FA and AFC duration in the finger tapping task (Sullivan et al. 2001; Muetzel et al., 2008). We further aim to test for a relationship between myelin metrics in the corpus callosum and AFC duration"

Sensorimotor tracts and TOJ: "Performance on the temporal order judgement task is not associated with integrity of a single specific white matter tract, but rather with a set of tracts involving multiple sensorimotor areas. Accordingly, we plan to run exploratory analyses across the whole brain, testing for associations between JND/slope values and FA/qMRI."

17

Covariates of age, sex, and performance on control tasks (unimanual finger tapping speed for the AFT, and visuomotor speed for DSST) were included. For each behavioural assay, voxelwise analyses were restricted to voxels within a predefined anatomical mask chosen from standard atlases included in FSL and based on the a priori hypotheses: a cingulum mask for DSST, a callosal mask for AFT and a mask of cortico-cortical and ascending sensorimotor tracts for TOJ. The masks were derived from the JHU ICBM-DTI-81 Atlas, the JHU White-Matter Tractography Atlas and the Human Sensorimotor Tracts Atlas, respectively. Within these masks, multimodal analyses were conducted with voxelwise maps of FA, MT, R1 and R2*. Voxelwise inference across these MRI modalities, testing for correlations between each MRI modality and behavioural measures, was performed using the tool Permutation Analysis of Linear Models (PALM) (Winkler et al., 2014). Joint voxelwise inference using Non-Parametric Combination (NPC), as implemented in PALM (Winkler et al. 2016), was used to combine evidence over modalities. At each voxel, combined evidence of effects over K modalities was assessed with Fisher's p-value combining method:

$$-2\sum_{K=1}^{K} ln(p_k) \tag{1}$$

With this approach, evidence can be assessed for either directional or non-directional effects: combining one-sided p-values (based on prior expected directions of effects) will test for directional effects; combining two-sided p-values will provide sensitivity to non-directional effects (i.e., combination of

18

either direction) as well. We used the distribution of the maximum test statistic for correcting for multiple testing across space and modalities, thus controlling the familywise error rate, and also across space for NPC. We used a cluster-forming threshold of t>1.7 (equivalent to p<0.05, based on the degrees of freedom) and a familywise error significance threshold of 5%.

Results

We first used unimodal analyses to test for correlations between DWI-derived FA and behaviour, based on previously reported literature (Figure 3). No relationships were found between behaviour and FA within tracts of interest for either TOJ or DSST (TOJ: peak p_{corr}=0.08; DSST: peak p_{corr}=0.49). For AFT, a significant correlation was found between callosal FA and AFT performance (peak p_{corr}=0.016).
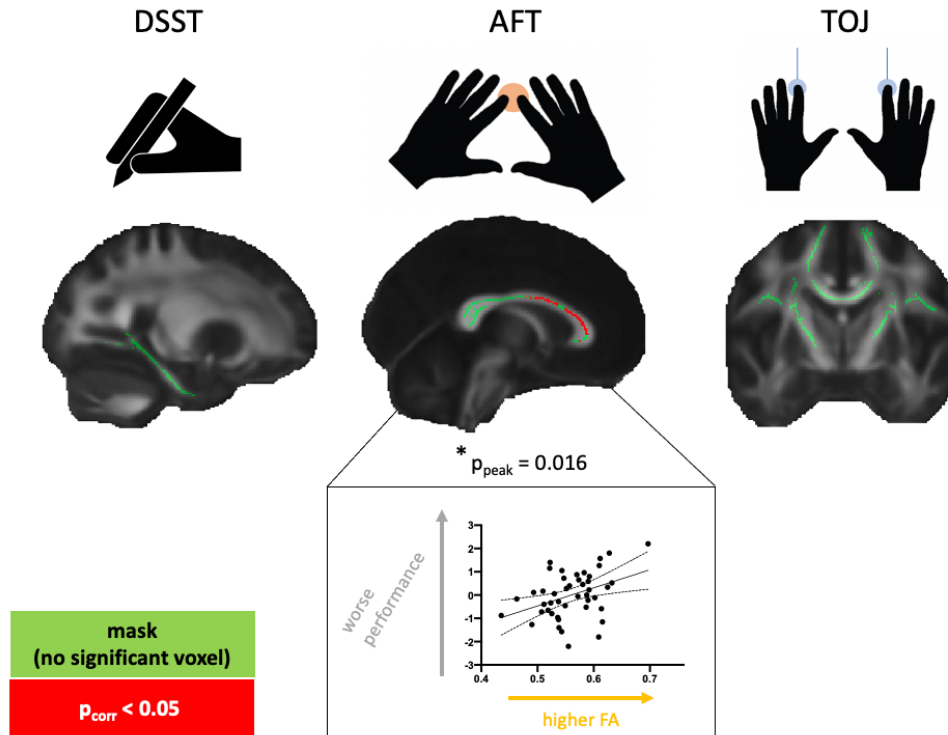
Figure 3: **FA and behaviour.** Unimodal relationships between FA and behaviour were tested across anatomical masks (shown in green) that were selected for each task. Results highlight that the Alternating Finger Tapping task (AFT), but not Temporal Order Judgement task (TOJ) and Digit Symbol Substitution Test (DSST) has a significant relationship with FA (red cluster shows voxels with corrected p-values below 0.05). Within that cluster, mean FA is extracted for each subject and plotted against performance in the scatterplot (with line of best fit and 95% confidence bands), that is for visual assessment of the correlation, rather than for statistical inference.

We then tested whether *any* modality (FA, MT, R1 or R2*) strongly correlated with behaviour (Figure 4), by considering p-value across both voxels and modalities for each WM-behaviour relationship. No relationships were found between behaviour and multimodal MRI metrics within tracts of interest for either TOJ or AFT (TOJ: peak p_{corr}=0.339; or callosum:

21

peak p_{corr}=0.09). For DSST, a significant correlation was found between parahippocampal cingulum and DSST (peak p_{corr}=0.038), driven entirely by R1 (only modality with any voxel of p_{corr} <0.05, Figure 5).

While single-modality tests allow to identify the strongest correlations across modalities, they cannot identify multimodal trends, which can be particularly informative of the underlying biology. For instance, a positive trend across all modalities (which are known to positively correlate with myelin content of the tissue) would indicate that tissue myelination may be related to behavioural performance. Likewise, trends in discordant directions could also be informative, as they could unveil multimodal signatures related to other biological tissue properties such as vasculature and fiber orientation.
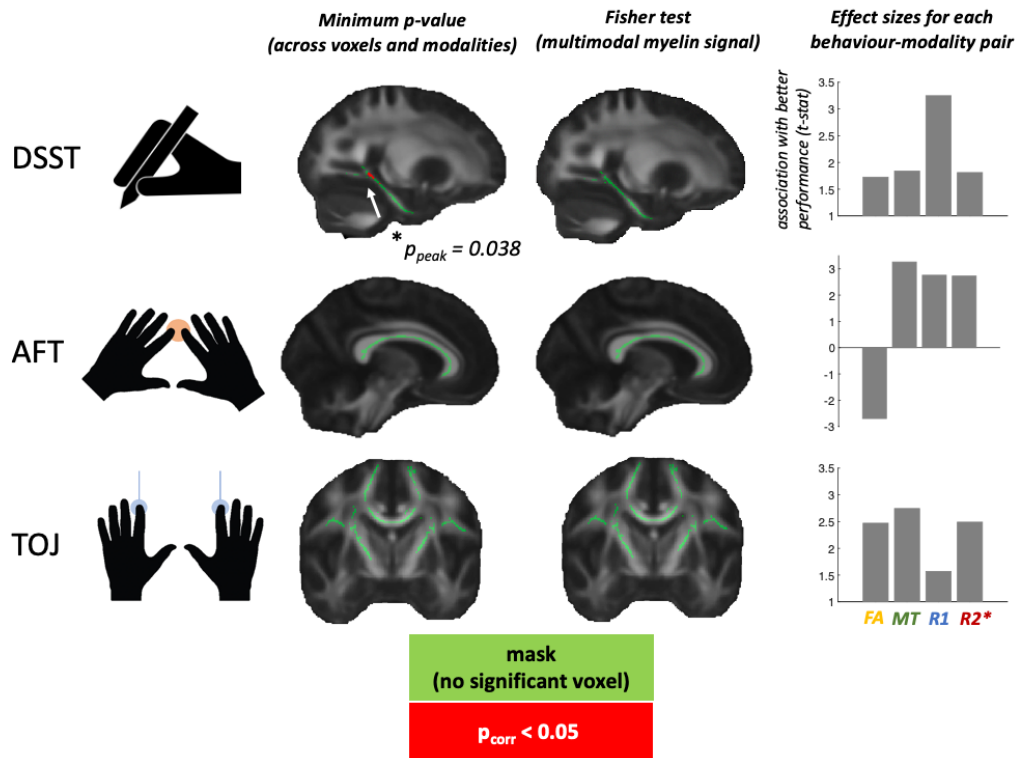
Figure 4: **Multimodal imaging and behaviour.** Multimodal relationships between WM and behaviour across MRI metrics (FA, MT, R1 and R2*) and across Digit Symbol Substitution Test (DSST), Alternating Finger Tapping task (AFT) and Temporal Order Judgement task (TOJ). Only the DSST has a significant relationship with cingulum WM when considering FWER-corrected p-values (2nd column, red cluster shows voxels with corrected p-values below 0.05). A Fisher test was used to test for multimodal microstructural signatures relating WM to behavior but no significant effects were found (3rd column). On the right column, effect sizes are reported for each modality-behaviour correlation, as measured by the top 5% t-statistic within peak Fisher clusters. This analysis was carried out to provide a clear visualisation of peak effect size for each pair of MR modality and behaviour, rather than for statistical inference.

Fisher tests were used to detect multimodal trends between behavioural measures and MRI metrics (FA, MT, R1 and R2*). With the usual (directed, positive) Fisher test (Figure 4), no relationships were found between behaviour and multimodal MRI metrics within tracts of interest (TOJ: peak

23

p_{corr}=0.532; AFT: peak p_{corr}=0.184; DSST: peak p_{corr}=0.2). With a non-directed Fisher test (results not shown), once again no relationships were found between behaviour and multimodal MRI metrics within tracts of interest. (TOJ: peak p_{corr}=0.82; AFT: peak p_{corr}=0.11; DSST: peak p_{corr}=0.29) Taken together, these two tests argue against the presence of multimodal microstructural signatures related to myelination or to other biological tissue properties.

For completeness, we also report analyses of this dataset using conventional univariate approaches, considering each modality separately (Figures 5, 6 and 7) and not correcting across modalities. We find that if each modality-behaviour correlation was run as a separate analysis, each behaviour would show a correlation with at least one modality. Strikingly, different behaviours correlate most strongly with different modalities (DSST with R1 (Figure 5); AFT with FA and MT (Figure 6); TOJ with R2* (Figure 7)) thus there is no evidence for a common microstructural signature across behaviours.
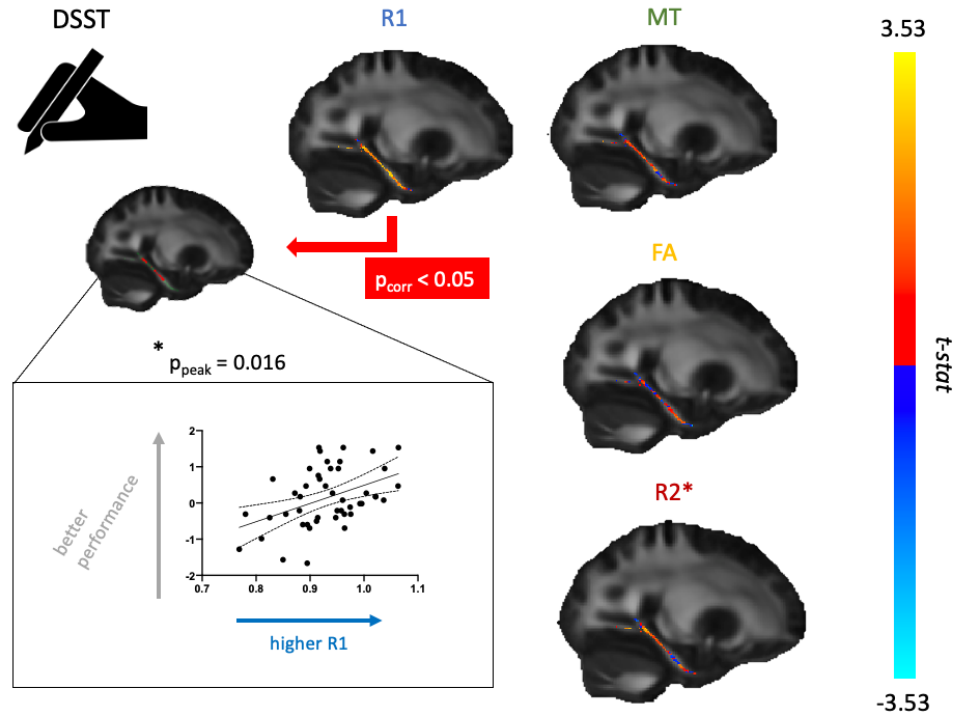
Figure 5: **Correlation between DSST performance and cingulum microstructure, reported as univariate results.** For each modality, unthresholded t-statistics are visualized according the colour bar (right). For R1 only, a cluster of voxels survived the threshold of p<0.05. Average R1 values within that cluster are shown against performance score in the scatterplot (with line of best fit and 95% confidence bands), which is presented for visualisation and is not used for statistical inference.

The lack of a common microstructural signature is also apparent when considering the top 5th percentile t-statistics (Figure 4) and the t-statistics maps for each task (Figures 5, 6 and 7), where peaks are not consistent across modalities. This further confirms the negative Fisher tests, as there is no common trend across modalities within each group of WM-behaviour tests.
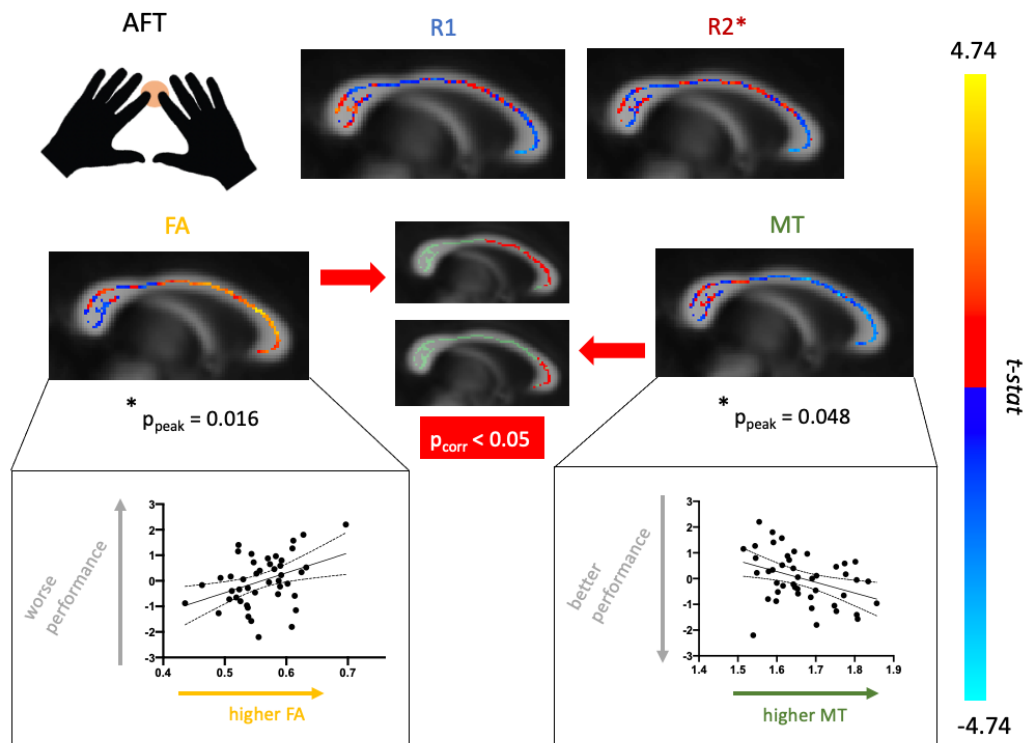
Figure 6: **Correlation between AFT and callosal microstructure, reported as univariate results.** For each modality, unthresholded t-statistics are visualized according the colour bar (right). For FA and MT, clusters of voxels survived the threshold of p<0.05. Average FA/MT values within that cluster are shown against performance score in the scatterplots (with line of best fit and 95% confidence bands), which are presented for visualisation and are not used for statistical inference.
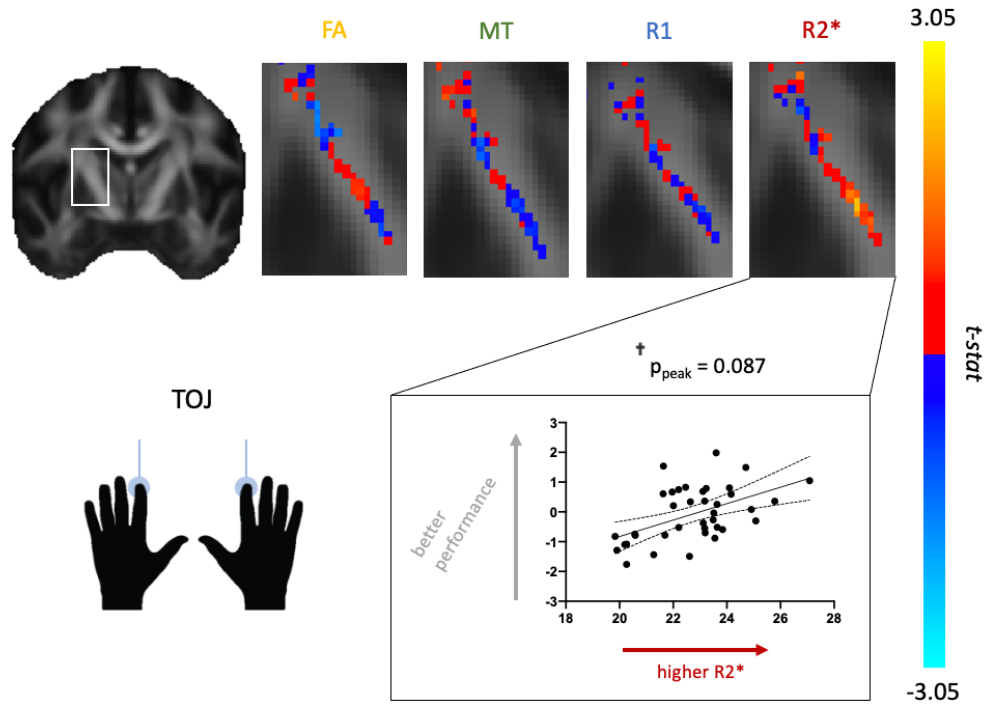
Figure 7: **Correlation between TOJ performance and CST microstructure, reported as univariate results.** For each modality, unthresholded t-statistics are visualized according to the colour bar (right). For R2* only, a cluster of voxels reached p=0.087. Average R2* values within that cluster are shown against performance score in the scatterplot (with line of best fit and 95% confidence bands), which is presented for visualisation and is not used for statistical inference.

Discussion

Our first aim was to assess the robustness of relationships between white matter FA and behaviour across a range of behavioural tasks. We find a unimodal correlation between the structure of the corpus callosum FA and bimanual coordination, in accordance with previous literature (Bathelt et al., 2019; Gooijers and Swinnen, 2014; Johansen-Berg et al., 2007; Metzler-Baddeley et al., 2012; Muetzel et al., 2008; Sullivan et al., 2001). This confirms in a preregistered analysis that individuals with lower callosal FA perform better in tasks requiring bimanual coordination.

However, a robust relationship between FA and behaviour was identified in only one out of three tasks considered here. This can be due to several reasons. One possible explanation is that effect sizes inferred from previous studies might be overinflated due to publication bias (Turner et al., 2008) and under-powered analyses (Button et al., 2013). However, it is worth noting that, of the three tasks considered here, only the FA-AFT experiment, which did successfully identify a FA-behaviour relationship, was a direct replication of a previous testing protocol. The other two tasks were designed as conceptual replications or extensions, but did not precisely replicate experimental conditions and analysis steps. This suggests that relationships between FA and behaviour that have been established may be robust and replicable, but the circumstances in which they occur needs clarification.

A second aim of the present study was to probe whether multimodal MR can provide useful insights on WM-behaviour relationships. We find that

this is the case for at least one of the WM-behaviour relationships we tested: R1 correlates with DSST performance, such that individuals with higher R1 perform better in the DSST task requiring cognitive control. Higher R1 could reflect greater myelin, oligodendrocytes, vasculature or other iron-rich tissue components. In this case, multimodal analysis allowed identification of a WM-behaviour relationship that would have not been detected by an analysis focused on FA in isolation. This confirms that there is value in multimodal imaging, as some modalities may be more sensitive to the presence of a relationship than others.

A third aim was to test whether there are common multimodal microstructural patterns in WM-behaviour relationships, which may provide insights into the underlying biology. We fail to find robust evidence for multimodal effects and cross-modality signatures. Rather, we find that effect sizes and directionality of effect in the relationship between each modality and each behaviour are highly heterogeneous. This means that MR modalities in each tract not only show heterogeneity in how they relate to the same behaviour, but there is also variation as a function of which tract-behaviour correlation is being considered.

A key insight from the study is therefore that the relationship between WM and behaviour is highly varied. Given that each modality has a specific pattern of sensitivity to the underlying biology (Figure 1), the results suggest that different aspects of WM biology may be driving different WM-behaviour correlations. There are two prominent sources of biological heterogeneity in white matter, which are likely relevant to the results in this study.

29

One driver of heterogeneity may be at the level of myelination. We selected metrics that were all sensitive to the amount of myelin in an imaging voxel (Figure 1), predicting that if myelination were responsible for WM-behaviour relationships, a common multimodal pattern across all relationships would be identified. Such patterns were not found, arguing against myelination as a common driver. However, such reasoning might be overly simple-minded. Histological studies have increasingly highlighted the heterogeneity of features in the myelinated axon, which can vary independently of each other (Almeida and Lyons, 2017). For instance, we know that Nodes of Ranvier, myelin sheath thickness, myelin sheath length, and number of myelin sheaths, can all independently affect an axon's physiological properties, which one would expect, in turn, to shape behaviour (Kaller et al., 2017). Varying these features might have differing effects on the overall amount of myelin in a given voxel meaning that the imaging metrics used might not be equally sensitive to all relevant features of the myelinated axon.

A second important driver of heterogeneity is non-myelin features of WM. As exemplified in Figure 1, while all sequences we used are sensitive to myelin, some are also sensitive to fiber orientation and neuronal volume (FA), and some are sensitive to iron and vasculature (R1 and R2*). Therefore, one possible interpretation of the data is that the relationship between AFT performance and the corpus callosum is highly influenced by fiber orientation, whereas the relationship between the DSST performance and the cingulum is shaped by vasculature. Previous studies highlighted that both fiber orientation (Chang et al., 2017; Wedeen et al., 2005) and vasculature (Licht

30

et al., 2011; Rhyu et al., 2010; Thomas et al., 2016) are important for brain function, and our data thus draw further attention to the fact that these factors may be influential in WM-behaviour relationships.

These two factors combined may explain why there is no single aspect of WM that drives behaviour. Rather, our findings confirm that heterogeneity at the cellular level is reflected in variation in the relationship between neuroimaging markers and behaviour. Importantly, this emphasizes that there is no single modality or single combination of modalities which is optimal to study WM-behaviour relationships. In this respect, our study poses practical limits to the possibility of developing a one-size-fits-all approach to the investigation of white matter-behaviour relationships, due to their inherent diversity.

While this heterogeneity means it is not straightforward to predict which MR modality is best suited for each type of WM investigation, it also suggests that multimodal studies of WM should tailor their MR sequence protocols and analyses pipelines to privilege markers and statistical approaches that can test and compare biologically-grounded models. For example, with an appropriate acquisition sequence and a joint multimodal statistical framework, one might be able to test whether a given WM-behaviour correlation is driven by myelination, vasculature (Thomas et al., 2016), or connectivity (Sui et al., 2014). Such approaches are most likely to generate further insights into WM-behaviour relationships in the future.

One key limitation of the study is that the results cannot disentangle to what extent differences between WM tracts contribute to the observed di-

versity of WM-behaviour relationships. One could argue, for example, that our results demonstrate that FA is more important for WM-behaviour relationships involving the corpus callosum, whereas R1 is more important for understanding the cingulum, while MT/R2* are more important in investigations of the corticospinal tract. Because each of the behaviours we selected relates to a different WM tract, this makes it impossible to disentangle whether different kinds of behaviours are most strongly driven by different microstructural patterns, or whether there is neuroanatomical heterogeneity in the importance of different microstructural features of each tract. Although both are likely to matter, further studies relating individual tracts to multiple behaviours are required.

The results also hold useful lessons for statistical aspects of future multimodal studies of WM. WM-behaviour correlations often have small effect sizes, and in our results we find that these effects are sometimes not detected when multiple hypotheses are tested concurrently. Testing for effects across modalities increases the false discovery rate proportionally to the number of modalities tested, and thus needs to be adequately corrected for in order to reach appropriate interpretations (Winkler et al., 2016). However, while multiple comparison correction has long been the gold standard statistical advice, multimodal brain imaging studies often do not report whether, and if so, how, correction for multiple comparisons was carried out (Bezukladova et al., 2020; Winston et al., 2020). Surprisingly, even gold standard guidelines in the field like COBIDAS do not report best practices for statistical reporting in multimodal imaging (Nichols et al., 2017), and many packages

that support multi-modality statistical testing do not allow joint statistical tests, thus leaving room for needless analytic flexibility. Our results suggest there is a need for increased transparency in reporting of multimodal statistics, which statistical guidelines on multimodal imaging might facilitate in the future. In this respect, our results also add weight to previous calls to pre-register the modalities to be used in a given analysis (Picciotto, 2018), and to report all tested modalities in publications.

This aspect of statistics in multimodal studies also needs to be taken into account when assessing the power of a given analysis. When modalities are analysed separately, multimodal studies require multiple statistical tests across modalities. Therefore, for the same effect size, a study analysing multiple modalities may need more subjects to achieve the same power, and it is important to take this into account in power analyses. We thus recommend using larger sample sizes for multimodal compared to unimodal studies. Alternatively, another solution is to use non-parametric multivariate tests (Winkler et al., 2014, 2016) and/or dimensionality reduction techniques (Groves et al., 2011; Sui et al., 2014), in scenarios where multimodal data are available but the data set size is only powered for unimodal tests.

In conclusion, these results highlight a broad heterogeneity in white matter's relationship with behaviour. They also underscore the added value of multimodal imaging approaches, as different neuroimaging modalities might be best suited to detect different WM-behavior relationships. However, this added value needs to be weighed carefully against the need for more power and/or dimensionality reduction approaches in multimodal studies. Finally,

33

the results effectively limit the possibility of developing a one-size-fits-all approach to study white matter, and suggest that different aspects of WM biology may be driving different WM-behaviour correlations.

Acknowledgements

Funding

References

Rafael G Almeida and David A Lyons. On myelinated axon plasticity and neuronal circuit formation and function. *Journal of Neuroscience*, 37(42): 10023–10034, 2017.

Jesper LR Andersson and Stamatios N Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion mr imaging. *Neuroimage*, 125:1063–1078, 2016.

Joe Bathelt, Amy Johnson, Mengya Zhang, and Duncan E Astle. The cingulum as a marker of individual differences in neurocognitive development. *Scientific reports*, 9(1):2281, 2019.

Svetlana Bezukladova, Jouni Tuisku, Markus Matilainen, Anna Vuorimaa, Marjo Nylund, Sarah Smith, Marcus Sucksdorff, Mehrbod Mohammadian, Virva Saunavaara, Sini Laaksonen, et al. Insights into disseminated ms brain pathology with multimodal diffusion tensor and pet imaging. *Neurology-Neuroimmunology Neuroinflammation*, 7(3), 2020.

Wouter Boekel, Eric-Jan Wagenmakers, Luam Belay, Josine Verhagen, Scott wn, and Birte U Forstmann. A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, 66:115–133, 2015.

Katherine S Button, John PA Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma SJ Robinson, and Marcus R Munafò. Power failure:

why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365, 2013.

Eric H Chang, Miklos Argyelan, Manisha Aggarwal, Toni-Shay S Chandon, Katherine H Karlsgodt, Susumu Mori, and Anil K Malhotra. The role of myelination in measures of white matter integrity: combination of diffusion tensor imaging and two-photon microscopy of clarity intact brains. *Neuroimage*, 147:253–261, 2017.

MSA Deloire-Grassin, B Brochet, B Quesson, C Delalande, V Dousset, P Canioni, and KG Petry. In vivo evaluation of remyelination in rat brain by magnetization transfer imaging. *Journal of the neurological sciences*, 178(1):10–16, 2000.

Vincent Dousset, Robert I Grossman, Karen N Ramer, Mitchell D Schnall, Lindon H Young, Francisco Gonzalez-Scarano, Ehud Lavi, and Jeffrey A Cohen. Experimental allergic encephalomyelitis and multiple sclerosis: lesion characterization with magnetization transfer imaging. *Radiology*, 182 (2):483–491, 1992.

Vincent Dousset, Bruno Brochet, Anne Vital, Christian Gross, Amid Benazzouz, Anne Boullerne, Anne-Marie Bidabe, Anne-Marie Gin, and Jean-Marie Caille. Lysolecithin-induced demyelination in primates: preliminary in vivo study with mr and magnetization transfer. *American journal of neuroradiology*, 16(2):225–231, 1995.

Shawna Farquharson, J-Donald Tournier, Fernando Calamante, Gavin

37

Fabinyi, Michal Schneider-Kolsky, Graeme D Jackson, and Alan Connelly. White matter fiber tractography: why we need to move beyond dti. *Journal of neurosurgery*, 118(6):1367–1377, 2013.

Ursula Fünfschilling, Lotti M Supplie, Don Mahad, Susann Boretius, Aiman S Saab, Julia Edgar, Bastian G Brinkmann, Celia M Kassmann, Iva D Tzvetanova, Wiebke Möbius, et al. Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature*, 485(7399):517, 2012.

Jolien Gooijers and SP Swinnen. Interactions between brain structure and behavior: the corpus callosum and bimanual coordination. *Neuroscience & Biobehavioral Reviews*, 43:1–19, 2014.

Adrian R Groves, Christian F Beckmann, Steve M Smith, and Mark W Woolrich. Linked independent component analysis for multimodal data fusion. *Neuroimage*, 54(3):2198–2217, 2011.

Florence Heath, Samuel A Hurley, Heidi Johansen-Berg, and Cassandra Sampaio-Baptista. Advances in noninvasive myelin imaging. *Developmental Neurobiology*, 78(2):136–151, 2018.

Fatima T Husain, Roberto E Medina, Caroline W Davis, Yvonne Szymko-Bennett, Kristina Simonyan, Nathan M Pajor, and Barry Horwitz. Neuroanatomical changes due to hearing loss and chronic tinnitus: a combined vbm and dti study. *Brain research*, 1369:74–88, 2011.

Heidi Johansen-Berg. Behavioural relevance of variation in white matter microstructure. *Current opinion in neurology*, 23(4):351–358, 2010.

Heidi Johansen-Berg, Valeria Della-Maggiore, Timothy EJ Behrens, Stephen M Smith, and Tomáš Paus. Integrity of white matter in the corpus callosum correlates with bimanual co-ordination skills. *Neuroimage*, 36:T16–T21, 2007.

Malte Sebastian Kaller, Alberto Lazari, Cristina Blanco-Duque, Cassandra Sampaio-Baptista, and Heidi Johansen-Berg. Myelin plasticity and behaviour?connecting the dots. *Current opinion in neurobiology*, 47:86–92, 2017.

James Kolasinski, Tamar R Makin, John P Logan, Saad Jbabdi, Stuart Clare, Charlotte J Stagg, and Heidi Johansen-Berg. Perceptually relevant remapping of human somatotopy in 24 hours. *Elife*, 5:e17280, 2016.

Christian Langkammer, Nikolaus Krebs, Walter Goessler, Eva Scheurer, Franz Ebner, Kathrin Yen, Franz Fazekas, and Stefan Ropele. Quantitative mr imaging of brain iron: a postmortem validation study. *Radiology*, 257(2):455–462, 2010.

Alberto Lazari and Ilona Lipp. Can mri measure myelin? systematic review, qualitative assessment, and meta-analysis of studies validating microstructural imaging with myelin histology. *bioRxiv*, 2020.

Alberto Lazari, Sigrid Koudelka, and Cassandra Sampaio-Baptista.

Experience-related reductions of myelin and axon diameter in adulthood. *Journal of neurophysiology*, 120(4):1772–1775, 2018.

Tamar Licht, Inbal Goshen, Avi Avital, Tirzah Kreisel, Salman Zubedat, Ronen Eavri, Menahem Segal, Raz Yirmiya, and Eli Keshet. Reversible modulations of neuronal plasticity by vegf. *Proceedings of the National Academy of Sciences*, 108(12):5081–5086, 2011.

YuChun Lin, JiunJie Wang, CheMing Wu, YauYau Wai, JenFang Yu, and ShuHang Ng. Diffusion tensor imaging of the auditory pathway in sensorineural hearing loss: changes in radial diffusivity and diffusion anisotropy. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 28(3): 598–603, 2008.

Antoine Lutti, Frederic Dick, Martin I Sereno, and Nikolaus Weiskopf. Using high-resolution quantitative mapping of r1 as an index of cortical myelination. *Neuroimage*, 93:176–188, 2014.

Claudia Metzler-Baddeley, Derek K Jones, Jessica Steventon, Laura Westacott, John P Aggleton, and Michael J O'Sullivan. Cingulum microstructure predicts cognitive control in older age and mild cognitive impairment. *Journal of Neuroscience*, 32(49):17612–17619, 2012.

Ryan L Muetzel, Paul F Collins, Bryon A Mueller, Ann M Schissel, Kelvin O Lim, and Monica Luciana. The development of corpus callosum microstruc-

ture and associations with bimanual task performance in healthy adolescents. *Neuroimage*, 39(4):1918–1925, 2008.

Marcus R Munafò, Brian A Nosek, Dorothy VM Bishop, Katherine S Button, Christopher D Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J Ware, and John PA Ioannidis. A manifesto for reproducible science. *Nature Human Behaviour*, 1(1):0021, 2017.

Klaus-Armin Nave. Myelination and support of axonal integrity by glia. *Nature*, 468(7321):244, 2010.

Thomas E Nichols, Samir Das, Simon B Eickhoff, Alan C Evans, Tristan Glatard, Michael Hanke, Nikolaus Kriegeskorte, Michael P Milham, Russell A Poldrack, Jean-Baptiste Poline, et al. Best practices in data analysis and sharing in neuroimaging using mri. *Nature neuroscience*, 20(3):299, 2017.

Brian A Nosek, Charles R Ebersole, Alexander C DeHaven, and David T Mellor. The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11):2600–2606, 2018.

Richard C Oldfield. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113, 1971.

Daniel Papp, Martina F Callaghan, Heiko Meyer, Craig Buckley, and Nikolaus Weiskopf. Correction of inter-scan motion artifacts in quantitative r1 mapping by accounting for receive coil sensitivity effects. *Magnetic resonance in medicine*, 76(5):1478–1485, 2016.

Jean Pelletier, Michel Habib, Olivier Lyon-Caen, Georges Salamon, Michel Poncet, and Richard Khalil. Functional and magnetic resonance imaging correlates of callosal involvement in multiple sclerosis. *Archives of Neurology*, 50(10):1077–1082, 1993.

Marina Picciotto. Analytical transparency and reproducibility in human neuroimaging studies, 2018.

IJ Rhyu, JA Bytheway, SJ Kohler, H Lange, KJ Lee, J Boklewski, K McCormick, NI Williams, GB Stanton, WT Greenough, et al. Effects of aerobic exercise training on cognitive function and cortical vascularity in monkeys. *Neuroscience*, 167(4):1239–1248, 2010.

R Edward Roberts, Elaine J Anderson, and Masud Husain. White matter microstructure and cognitive function. *The Neuroscientist*, 19(1):8–15, 2013.

Melissa Saenz and Ione Fine. Topographic organization of v1 projections through the corpus callosum in humans. *Neuroimage*, 52(4):1224–1229, 2010.

Cassandra Sampaio-Baptista and Heidi Johansen-Berg. White matter plasticity in the adult brain. *Neuron*, 96(6):1239–1251, 2017.

David I Shore, Kellie Gray, Emily Spry, and Charles Spence. Spatial modulation of tactile temporal-order judgments. *Perception*, 34(10):1251–1262, 2005.

Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

Stephen M Smith, Mark Jenkinson, Heidi Johansen-Berg, Daniel Rueckert, Thomas E Nichols, Clare E Mackay, Kate E Watkins, Olga Ciccarelli, M Zaheer Cader, Paul M Matthews, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505, 2006.

Patrick E Steadman, Frances Xia, Moriam Ahmed, Andrew J Mocle, Amber RA Penning, Anna C Geraghty, Hendrik W Steenland, Michelle Monje, Sheena A Josselyn, and Paul W Frankland. Disruption of oligodendrogenesis impairs memory consolidation in adult mice. *Neuron*, 2019.

HB Stolp, G Ball, P-W So, J-D Tournier, M Jones, C Thornton, and AD Edwards. Voxel-wise comparisons of cellular microstructure and diffusion-mri in mouse hippocampus using 3d bridging of optically-clear histology with neuroimaging data (3d-bond). *Scientific reports*, 8(1):1–12, 2018.

Carsten Stüber, Markus Morawski, Andreas Schäfer, Christian Labadie, Miriam Wähnert, Christoph Leuze, Markus Streicher, Nirav Barapatre, Katja Reimann, Stefan Geyer, et al. Myelin and iron concentration in

43

the human brain: a quantitative study of mri contrast. *Neuroimage*, 93: 95–106, 2014.

Jing Sui, Rene Huster, Qingbao Yu, Judith M Segall, and Vince D Calhoun. Function–structure associations of the brain: evidence from multimodal connectivity and covariance studies. *Neuroimage*, 102:11–23, 2014.

Edith V Sullivan, Margaret J Rosenbloom, John E Desmond, and Adolf Pfefferbaum. Sex differences in corpus callosum size: relationship to age and intracranial size. *Neurobiology of aging*, 22(4):603–611, 2001.

Hongfu Sun, Andrew J Walsh, R Marc Lebel, Gregg Blevins, Ingrid Catz, Jian-Qiang Lu, Edward S Johnson, Derek J Emery, Kenneth G Warren, and Alan H Wilman. Validation of quantitative susceptibility mapping with perls' iron staining for subcortical gray matter. *Neuroimage*, 105: 486–492, 2015.

Karsten Tabelow, Evelyne Balteau, John Ashburner, Martina F Callaghan, Bogdan Draganski, Gunther Helms, Ferath Kherif, Tobias Leutritz, Antoine Lutti, Christophe Phillips, et al. hmri–a toolbox for quantitative mri in neuroscience and clinical research. *NeuroImage*, 2019.

Adam G Thomas, Andrea Dennis, Nancy B Rawlings, Charlotte J Stagg, Lucy Matthews, Martyn Morris, Shannon H Kolind, Sean Foxley, Mark Jenkinson, Thomas E Nichols, et al. Multi-modal characterization of rapid anterior hippocampal volume increase associated with aerobic exercise. *Neuroimage*, 131:162–170, 2016.

Michelle Todorow, Joseph F DeSouza, Brenda L Banwell, and Christine Till. Interhemispheric cooperation in global–local visual processing in pediatric multiple sclerosis. *Journal of clinical and experimental neuropsychology*, 36(2):111–126, 2014.

Erick H Turner, Annette M Matthews, Eftihia Linardatos, Robert A Tell, and Robert Rosenthal. Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, 358 (3):252–260, 2008.

Hanlin Wang, Yi Liang, Wenhui Fan, Xia Zhou, Mingming Huang, Guojun Shi, Hui Yu, and Guiquan Shen. Dti study on rehabilitation of the congenital deafness auditory pathway and speech center by cochlear implantation. *European Archives of Oto-Rhino-Laryngology*, 276(9):2411–2417, 2019.

Van J Wedeen, Patric Hagmann, Wen-Yih Isaac Tseng, Timothy G Reese, and Robert M Weisskoff. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic resonance in medicine*, 54(6):1377–1386, 2005.

Nikolaus Weiskopf, John Suckling, Guy Williams, Marta Morgado Correia, Becky Inkster, Roger Tait, Cinly Ooi, Edward T Bullmore, and Antoine Lutti. Quantitative multi-parameter mapping of r1, pd*, mt, and r2* at 3t: a multi-center validation. *Frontiers in neuroscience*, 7:95, 2013.

Nikolaus Weiskopf, Martina F Callaghan, Oliver Josephs, Antoine Lutti, and Siawoosh Mohammadi. Estimating the apparent transverse relaxation time

(r2*) from images with different contrasts (estatics) reduces motion arti-facts. *Frontiers in neuroscience*, 8:278, 2014.

Anderson M Winkler, Gerard R Ridgway, Matthew A Webster, Stephen M Smith, and Thomas E Nichols. Permutation inference for the general linear model. *Neuroimage*, 92:381–397, 2014.

Anderson M Winkler, Matthew A Webster, Jonathan C Brooks, Irene Tracey, Stephen M Smith, and Thomas E Nichols. Non-parametric combination and related permutation tests for neuroimaging. *Human brain mapping*, 37(4):1486–1511, 2016.

Gavin P Winston, Sjoerd B Vos, Benoit Caldairou, Seok-Jun Hong, Monika Czech, Tobias C Wood, Stephen J Wastling, Gareth J Barker, Boris C Bernhardt, Neda Bernasconi, et al. Microstructural imaging in temporal lobe epilepsy: Diffusion imaging changes relate to reduced neurite density. *NeuroImage: Clinical*, page 102231, 2020.

Robert J Zatorre, R Douglas Fields, and Heidi Johansen-Berg. Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature neuroscience*, 15(4):528, 2012.