Tractography-based parcellation does not provide strong evidence of anatomical organisation within the thalamus

Jonathan D. Clayden[a,*], David L. Thomas[b,c], Alexander Kraskov[d]

[a]*Developmental Imaging and Biophysics Section, UCL Great Ormond Street Institute of Child Health, University College London, London, United Kingdom*
[b]*Neuroradiological Academic Unit, UCL Institute of Neurology, University College London, London, United Kingdom*
[c]*Leonard Wolfson Experimental Neurology Centre, UCL Institute of Neurology, Queen Square, London, United Kingdom*
[d]*Department of Clinical and Movement Neurosciences, UCL Institute of Neurology, University College London, London, United Kingdom*

Abstract

Connectivity-based parcellation of subcortical structures using diffusion tractography is now a common paradigm in neuroscience. These analyses often imply voxel-level specificity of connectivity, and the formation of compact, spatially coherent clusters is often taken as strong imaging-based evidence for anatomically distinct subnuclei in an individual. In this study, we demonstrate that internal structure in diffusion anisotropy is not necessary for a plausible parcellation to be obtained, by spatially permuting diffusion parameters within the thalami and repeating the parcellation. Moreover, we show that, in a winner-takes-all paradigm, most voxels receive the same label before and after this shuffling process—a finding that is stable across image acquisitions and tractography algorithms. We therefore suggest that such parcellations should be interpreted with caution.

Keywords: magnetic resonance imaging, diffusion, connectivity, tractography, parcellation

Introduction

Diffusion tractography uses proxy information about white matter structure to reconstruct the paths of neural tracts in the living brain. Being based on magnetic resonance imaging (MRI), it is a noninvasive technique with broad applicability in neuroscience and the clinic. An increasingly common application of tractography is connectivity-based parcellation, a paradigm in which a contiguous anatomical region—typically cortical or subcortical grey matter—is parcellated into subregions based on the inferred projections from each imaging voxel contained within it. This is established by running diffusion tractography a large number of times and assessing the pattern of connections to a set of target regions.

The steadily growing popularity of the technique has seen it being applied to the parcellation of a wide range of brain structures across many studies. The thalamus was the canonical early example, due to its extensive connectivity to different parts of the cortex, its functional relevance to a range of important neurological disorders and the well-established histological evidence of its nuclear structure (Guillery & Sherman, 2002; Krauth et al., 2010; Morel et al., 1997). Tractography-based *in vivo* thalamic parcellations were demonstrated in a seminal paper by Behrens et al. (2003), and similar principles have since been applied to the amygdala and basal ganglia (Bach et al., 2011; Draganski et al., 2008; Lambert et al., 2012; Saygin et al., 2011). Cortical regions have been explored too: Johansen-Berg et al. (2004) showed a marked distinction between the connectivity of the supplementary motor area (SMA) and the adjoining pre-SMA, Anwander et al. (2007) demonstrated a connectivity-based subdivision of Broca's area and Jakab et al. (2012) clustered the insula, amongst many other studies. Other work still has taken a simultaneous global parcellation approach, dividing up the whole brain at once based on its connectivity (e.g., Gallardo et al., 2018; Moreno-Dominguez et al., 2014).

Use of information from MRI as the basis for such parcellations predates the use of tractography. Magnotta et al. (2000) described a cortex-attenuated MRI sequence able to render thalamic nuclei more visible than on conventional structural scans—although segmentation of the nuclei in such images would require additional processing. Tuch (2002, ch. 6) clustered voxels directly from diffusion MRI data, using a combination of proximity and fibre orientation information. But tractography-based parcellation rapidly superseded such approaches in the literature, and improvements to tractography methods—particularly the ability to resolve fibre crossings within image voxels—led to greater stability in the parcellations (Behrens et al., 2007).

In addition to their value in primary neuroscientific investigations, such parcellations are relevant in the clinic. Patient-specific maps of subcortical nuclei provide potentially valuable navigation information prior to electrode implantation for deep brain stimulation (DBS), a therapeutic neurosurgical intervention used to treat movement disorders such as Parkinson's disease and dystonia. Recently, da Silva et al. (2017) have demonstrated a connectivity-based parcellation of the globus pallidus internus, a frequent target for DBS, while Akram et al. (2017) identified clusters in the subthalamic nuclei of patients

---

that were associated with alleviation of various Parkinsonian symptoms.

Although the tractography-based approach has been shown to be reproducible, and in many cases to broadly match expectations from neurophysiology and anatomy (Klein et al., 2007), the approach has received some criticism. Eickhoff et al. (2015) reviewed several challenges, including inconsistencies between parcellations derived from different clustering methods and imaging modalities, difficulties with statistical inference in this context, and the role of functional gradients as opposed to clearly defined nuclear boundaries. Nevertheless, the mechanics of the technique are rarely scrutinised in detail. In particular, while tractography itself faces a number of outstanding issues, such as a preponderance of "false positive" connections (Maier-Hein et al., 2017), the impact of these and other matters of procedure on the parcellations has barely been explored *in vivo*.

In this work we examine the role of fibre orientation information within the parcellated region of interest, by deliberately scrambling the voxels within it and considering the effect on the pattern of projected connections. We demonstrate that coherent structure within the human thalamus in diffusion MRI data is not necessary for plausible delineation of connected subregions. We therefore conclude that such parcellations are potentially prone to over-interpretation, and must be treated with caution.

Methods

Ethical approval for all imaging was granted by the Research Ethics Committee at University College London.

Diffusion-weighted spin-echo echo-planar (DW-SE-EPI) images were acquired from 17 healthy adult volunteers (six female; mean age at scan 32.84 yr, standard deviation 8.13 yr) on a Siemens Avanto 1.5 T scanner with a 32-channel head coil. Diffusion weighting was applied along 30 noncollinear directions at $b = 800$ s mm^{-2} and 60 directions at $b = 2400$ s mm^{-2}, and nine volumes were acquired with $b = 0$. The echo time was 98 ms and the volume repetition time was 8.30 s. 60 slices were acquired, with a matrix size of 96×96 and a voxel size of 2.5 mm in each dimension. The data were converted from DICOM to NIfTI-1 format using TractoR (Clayden et al., 2011), corrected for susceptibility and eddy current induced distortions using `topup` and `eddy` from FSL version 5.0.11 (Andersson et al., 2003; Andersson & Sotiropoulos, 2016; Smith et al., 2004), and masked using FSL's brain extraction tool (Smith, 2002). A T_1-weighted structural scan was also acquired (3D Fast Low-Angle SHot; flip angle 15°, echo time 4.94 ms, repetition time 11 ms, resolution $1 \times 1 \times 1$ mm), and automatically parcellated using FreeSurfer (Desikan et al., 2006). The raw data is available online (https://osf.io/94c5t/; Deligianni et al., 2016).

To explore the generalisability of our findings across acquisitions, an additional data set was acquired from a 36 year-old male volunteer on a Siemens Prisma 3 T scanner with 64-channel head coil. In this case, diffusion data were acquired along 60 directions at $b = 1000$ s mm^{-2} and 60 directions at $b = 2200$ s mm^{-2}, along with 14 interspersed $b = 0$ volumes.

A multi-band DW-SE-EPI sequence was used with multi-band factor 2 (Setsompop et al., 2012); the echo time was 60 ms and the volume repetition time was 3.05 s. 66 slices were acquired with a matrix size of 110×110 and a 0.2 mm slice gap; the voxel size was 2 mm in each dimension. The T_1-weighted scan in this case used a Magnetisation-Prepared Rapid Acquisition Gradient Echo sequence with a flip angle of 8°, echo time of 2.74 ms, repetition time of 2300 ms, inversion time of 909 ms and reconstructed resolution of $1 \times 1 \times 1$ mm. All image pre-processing was equivalent.

The left and right thalami, and non-overlapping masks covering the frontal cortex, precentral gyrus, postcentral gyrus, temporal cortex and parietal–occipital cortex, were extracted from the FreeSurfer parcellations (see Fig. 1a). These masks were transformed to diffusion space using NiftyReg (Modat et al., 2010), based on a nonlinear registration between the T_1-weighted image and a $b = 0$ volume from the diffusion acquisition.

"Ball-and-sticks" models with one, two and three sticks were fitted separately to each dataset using FSL-BEDPOSTX (Behrens et al., 2007). A copy of each dataset was made, and the fitted ball-and-sticks model parameters for each voxel were randomly repositioned in space within the left and right thalamus in turn (see Fig. 2). The code for this process, in the form of a TractoR script, is available alongside this paper.

In every unshuffled and shuffled dataset, 5000 streamlines were seeded from each voxel within the thalamus masks, and streamlines were propagated using TractoR in one direction only. Streamlines reaching one of the target masks were counted as connecting the source voxel to that target, and then immediately terminated (see Fig. 1b). Each thalamus voxel was coloured to indicate the target at which the greatest number of streamlines arrived, following the "winner-takes-all" scheme introduced by Behrens et al. (2003). No colour was assigned where none of the streamlines from a voxel reached any of the target regions.

Constrained spherical deconvolution (CSD) was also separately performed on the preprocessed 3 T dataset, using MRtrix 3.0 RC3 (Tournier et al., 2007). An iterative algorithm was used to fit the response function (i.e., deconvolution kernel) from the $b = 2200$ s mm^{-2} shell in the diffusion-weighted data, using a maximum spherical harmonic order of eight (Tournier et al., 2013). The default tracking parameters in MRtrix were used to generate 5000 streamlines per seed voxel, except that streamlines were propagated in one direction only, to match the process used with the ball-and-sticks models. Maps of the endpoints of streamlines reaching each of the target regions were obtained, and a winner-takes-all parcellation was derived from these maps.

In every case outlined above, a series of summary statistics were compiled, capturing the numbers of streamlines reaching the targets, the numbers of labelled voxels and the degree of consistency in the labels before and after shuffling. To provide a meaningful baseline, the degree of agreement under a direct, random permutation of voxel labels was calculated as described in the Appendix. Analysis was performed with R version 3.5.0 (R Core Team, 2018), and graphics were created with TractoR and the `ggplot2` package (Wickham, 2009).
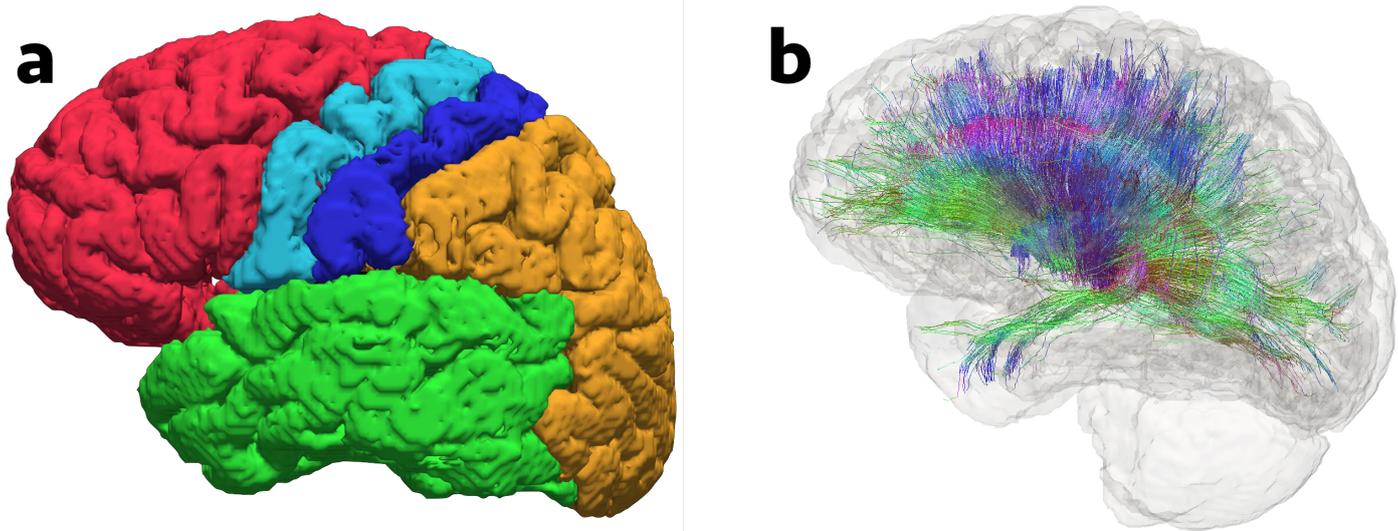
2

Figure 1: Rendered three-dimensional map of the cortical targets used in this study (a), along with a full set of streamlines projecting from the thalamus (b). Targets are the frontal cortex (red), precentral gyrus (light blue), postcentral gyrus (dark blue), temporal cortex (green) and parietal–occipital cortex (orange). Streamlines are coloured according to their local orientation, using the conventional colour scheme with red for left–right, green for anterior–posterior and blue for superior–inferior.
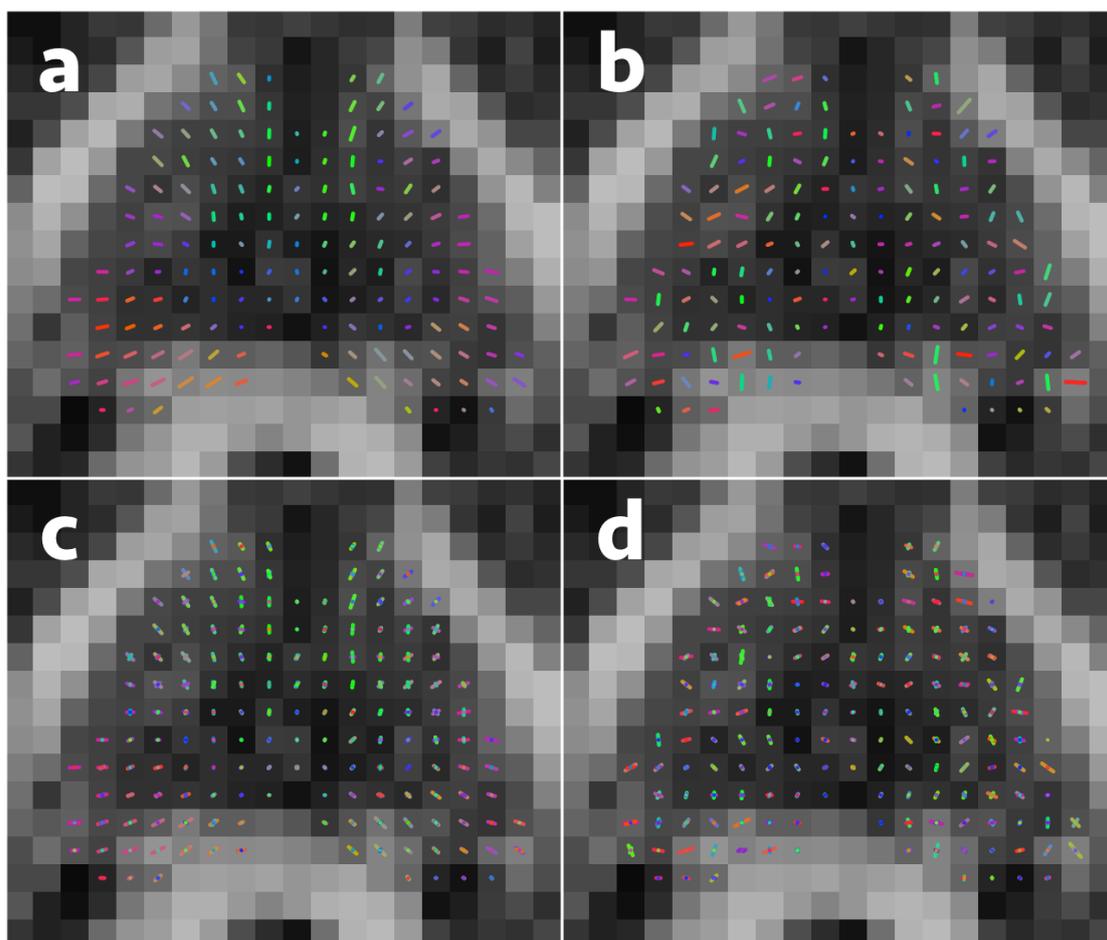

Figure 2: Illustration of the effect of parameter shuffling within the thalamus in one subject. Fibre direction information is shown as fitted by BEDPOSTX for the one stick (a) and three stick (c) cases, alongside the shuffled equivalents (b, d). All parameters are moved wholesale to a randomly chosen other voxel, separately for each hemisphere. The axial slice containing the most thalamus voxels is shown in each case.

3

Results

In every dataset and for every analysis, our shuffling process drastically reduced the number of streamlines reaching any of the target regions, due to the disorderly nature of the orientation information within the thalamus. This resulted in several voxels being unlabelled, especially towards the centre of the thalamus. Nevertheless, the parcellations were still remarkably coherent, with a level of agreement in voxel labels which was well above chance in every case. Neither switching acquisitions nor changing tractography algorithms altered the pattern of the results.

For each subject and diffusion model, the real fitted diffusion parameter maps were shuffled within each thalamus as described, the tractography and parcellation was repeated, and the total number of voxels labelled equivalently before and after shuffling was counted. These values were then averaged over the one, two and three stick diffusion models, giving one aggregate value per subject. The subjects showing the minimum, median and maximum agreement by this aggregated measure were then identified, and are shown in Fig. 3, so as to illustrate a range of outcomes. The figure shows the axial slice containing the largest number of thalamus voxels in each subject, parcellated as described above, coloured as shown in Fig. 1a and overlaid on a fractional anisotropy map. We observe that parcellations in the unshuffled data generally become a little cleaner as the number of fibre directions modelled per voxel increases (from left to right within each subject)—but the same is also true for the shuffled variants. Although several voxels are uncoloured, clear and coherent structure remains visible in the shuffled parcellations, and there remains evidence of compact nuclei, despite the absence of internal orientational structure within the thalami in these cases.

Fig. 4 shows the distributions of various numerical indicators of the effects of shuffling, across the core dataset of 17 subjects. Averaged across this dataset, just 14–17% as many streamlines seeded from the thalamus reached one of the cortical targets after shuffling, compared to before, but 61% (for the one-stick model), 83% (two sticks) and 85% (three sticks) of voxels received a label—compared to almost 100% before shuffling. Out of these connected voxels, 63–66% received the same label (from the five available) after shuffling, compared to a theoretical random baseline averaging 34–35% (see the Appendix for details of the calculation behind this). Clearly, therefore, the shuffling process has not fully disrupted the parcellation as might be expected.

The mean volumes of all parcellated voxels in the thalamus, and each cluster individually, are shown in Table 1 along with Dice coefficients averaged across subjects, indicating the degree of overlap in a way that takes these different volumes into account. The Dice coefficients thus conflate two effects which Fig. 4 shows separately: the reduction in the proportion of thalamic voxels which receive a label at all after shuffling, and the extent to which those labels match the equivalents assigned before shuffling. Nevertheless, some interesting patterns emerge. The postcentral gyrus cluster is the smallest and least consistent, while the frontal cluster is the largest and has the highest Dice coefficients. The frontal and temporal clusters shrink appreciably after shuffling, and the postcentral cluster stays about the same size, but in the 2 and 3 stick models, in particular, the precentral and parietal–occipital clusters become somewhat larger.

Figures for the single 3 T dataset are shown in Fig. 4 as coloured points circled in white, and the shuffled and unshuffled parcellations are shown in Fig. 5, including for the CSD pipeline. The results are generally similar to those described above, with the numerical values mostly within the range observed in the main dataset. The exceptions are the proportions of labelled voxels in the two and three stick models, which are somewhat higher than in the other cases, perhaps due to the higher image resolution. Nevertheless, the degree of label agreement before and after shuffling is in line with the broader pattern. CSD does not have a fixed bound on the number of fibre populations in the same way as BEDPOSTX, and the numerical results were a little higher throughout, with 25% as many streamlines reaching targets after shuffling, 100% of voxels connected and 75% overlap in labels against a random baseline of 39%. Interestingly, the label overlap between the parcellation derived from CSD and the one from BEDPOSTX (with three sticks) was also 75% (cf. Fig. 5).

Finally, Fig. 6 shows variants of the maps from Fig. 3, but focussing on the three-stick model and the subject with the median level of agreement between the labels before and after shuffling. The left column of Fig. 6 shows a variant of the winner-takes-all parcellation where opacity at each voxel is determined by the proportion of connected streamlines that reach the winning region; in other words, it indicates the margin by which the winning target wins that voxel. In practice, it looks very similar to the standard parcellation in Fig. 3, both before and after shuffling. The remaining columns of Fig. 6 show maps pertaining to individual target regions, each of which uses a fixed colour for every voxel, but with opacity normalised across voxels such that the voxel with the greatest number of streamlines reaching that target is opaque. Given that the number of streamlines seeded from each voxel is the same, these maps indicate the relative "strengths" of the connections between individual voxels and the target, and in this case the shuffled maps look quite different, being visibly limited to a few isolated voxels in each case, generally at the boundary with white matter.

Discussion

In this study we have demonstrated that artificially destroying internal structure in diffusion anisotropy within the human thalamus does not fully inhibit the creation of plausible streamline-based parcellations, as might be expected. It can be seen directly from Figs 3 and 5 that, while several voxels lose their label completely, the remaining voxels maintain a largely compact, clustered structure, while Fig. 4 shows that consistency in the parcellation before and after shuffling is well above the level expected by chance. Moreover, we have demonstrated that all of these outcomes persist between different image acquisitions and tractography pipelines. Whilst the robustness of these parcellations might be considered to be a positive finding, it casts significant doubt on the implied spatial specificity of voxelwise
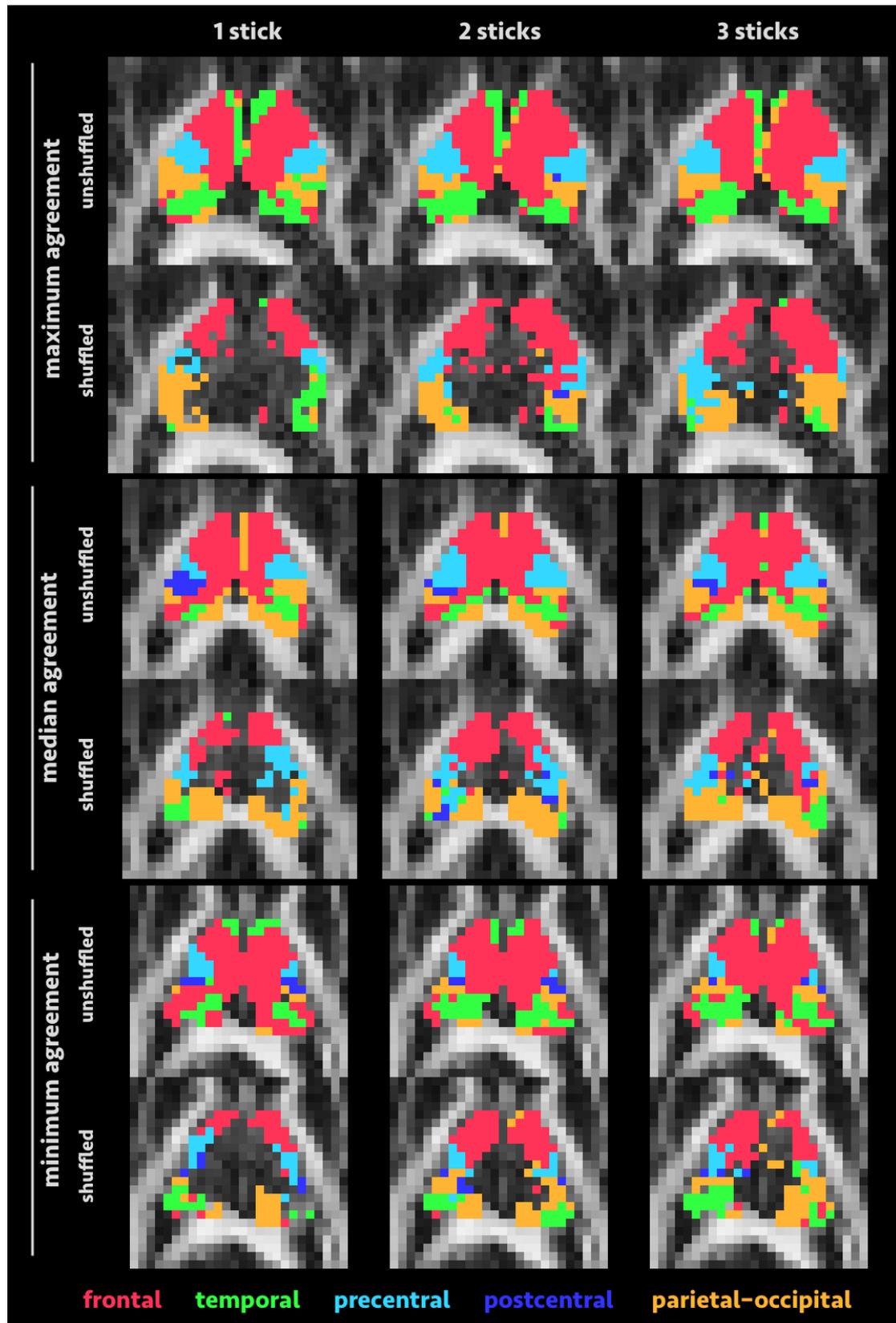
Figure 3: "Winner-takes-all" parcellations before and after spatial shuffling for three subjects, specifically those with the greatest (top), median (middle) and least (bottom) number of voxels sharing a label before and after shuffling. The one-stick case is in the left column, two-stick case in the middle and three-stick case on the right. The axial slice containing the most thalamus voxels is shown in each case, and colours match the cortical areas shown in Fig. 1a. The underlying greyscale map shows fractional anisotropy.
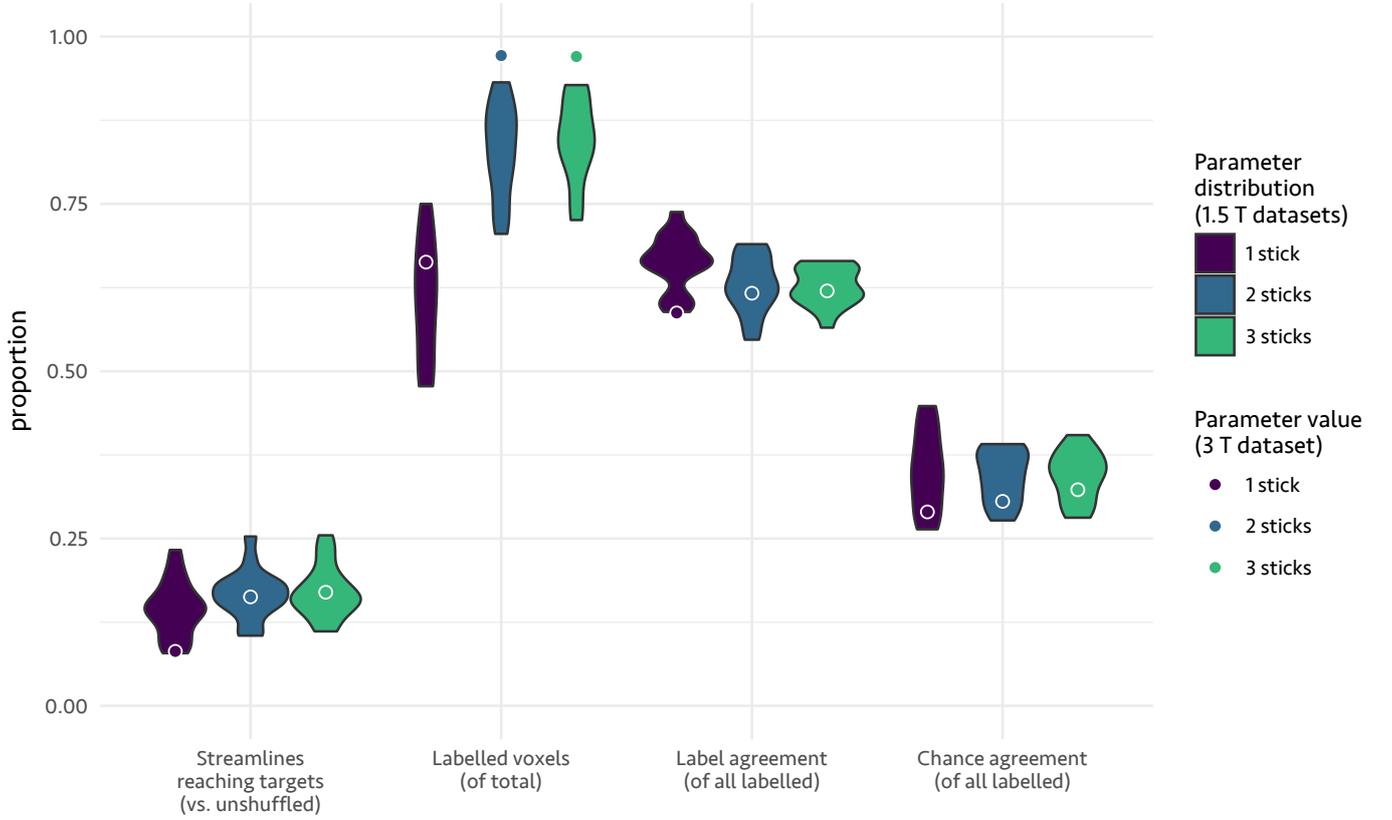
Figure 4: Violin plots quantifying the effects of shuffling across the datasets. Four measures are shown, as proportions: the number of streamlines reaching any of the target regions after shuffling, compared to before; the fraction of voxels receiving a colour; and the proportions of those labelled voxels which receive the same label before and after shuffling, in practice and at chance level (see the Appendix). Each violin shows the distribution of values for a particular number of BEDPOSTX stick compartments across the 1.5 T datasets, while the point circled in white shows the position of the 3 T dataset. Note that the chance level varies between datasets, because of the different numbers of thalamic voxels in each case.

Target region(s)	Unshuffled volume, ml			Shuffled volume, ml			Dice coefficient		
	1 stick	2 sticks	3 sticks	1 stick	2 sticks	3 sticks	1 stick	2 sticks	3 sticks
all	17.58	17.64	17.64	11.83	15.16	15.36	0.50	0.58	0.56
frontal	8.88	8.91	8.89	4.78	5.91	6.06	0.51	0.62	0.64
temporal	3.36	3.34	3.52	1.11	1.44	1.56	0.29	0.32	0.32
precentral	1.66	1.72	1.50	1.62	1.89	2.09	0.57	0.57	0.55
postcentral	0.16	0.06	0.06	0.14	0.08	0.12	0.20	0.19	0.00
parietal–occipital	3.86	3.97	4.14	3.34	4.92	5.00	0.55	0.60	0.59

Table 1: Median volumes, in millilitres, of the regions parcellated for each target in each BEDPOSTX model, before and after shuffling, and Dice coefficients indicating the degree of overlap in the whole parcellation and its constituent parts before and after shuffling. This standard overlap measure captures both the change in volume after shuffling and the level of agreement in the remaining labels.
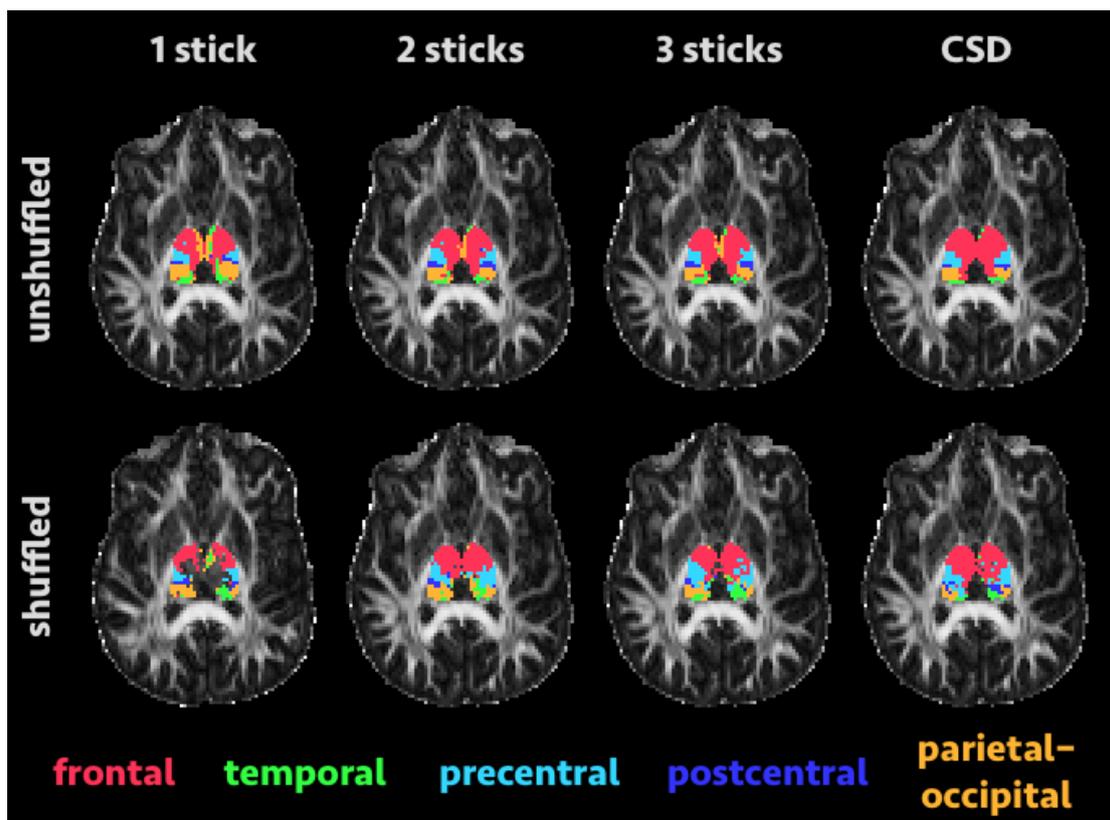
Figure 5: "Winner-takes-all" parcellations in the 3 T dataset, based on ball-and-sticks-based (first three columns) and CSD-based (fourth column) tractography. The axial slice containing the most thalamus voxels is shown in each case, and colours match the cortical areas shown in Fig. 1a.
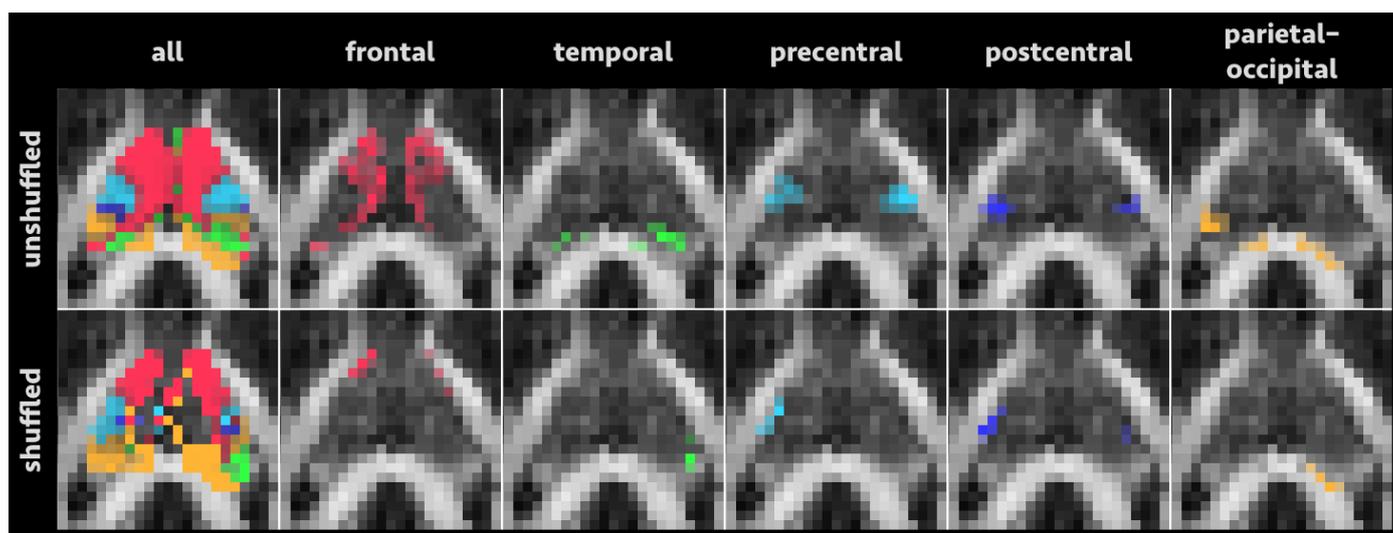

Figure 6: Use of transparency to show the degree of confidence in each voxel label. The first column matches the three-stick subfigure from the median subject in Fig. 3, but with the opacity of each voxel set to indicate the proportion of streamlines from that voxel that reach the winning target region. The remaining columns show maps specific to each target, where the opacity is normalised across voxels such that the voxel with the greatest number of streamlines reaching that target is opaque.

connectivity-based parcellations, and suggests that plausibility in such parcellations should not be interpreted as strong evidence for a particular subject-specific organisation of subnuclei, or for specific axonal pathways projecting to or from each voxel.

Our shuffling process, illustrated in Fig. 2, is a relatively extreme corruption of the data, which is performed after fitting ball-and-stick model parameters or estimating spherical harmonic coefficients. The advantage of performing this kind of shuffling, rather than using synthetic data, is that all of the parameters are calculated from measurements taken within the thalamus in the same hemisphere, and so they are unquestionably within the realm of general plausibility; they are simply misplaced within the region. Since nothing like this level of corruption would be expected in an unmodified dataset, it is reasonable to expect that very plausible results could be obtained even in the presence of substantial numbers of tractographic false positives and false negatives. Examining the streamline paths themselves, as in Fig. 1b, is highly advisable.

Since the thalamus contains a fairly large number of voxels and is orientationally inhomogeneous (cf. Fig. 2), we would not expect the specifics of any given shuffle of the voxels to be important. We checked this explicitly by repeating the shuffling and parcellation in the 3 T dataset 100 times for each diffusion model, and for every value shown in Fig. 4, we found that the mean of these 100 repetitions was very close to the value obtained from a single shuffle, with the standard error of the mean being less than 0.5 percentage points in every case. This is because the statistic relevant to the parcellation—the target through which the largest number of streamlines pass—is heavily overdetermined when there are 5000 streamlines and only five targets, and repeating the shuffling process does not substantially change its functional effect.

After shuffling, fibre tracking within the thalamus essentially becomes a random walk (cf. Morris et al., 2008), a series of short steps in approximately arbitrary directions. Since tractography algorithms generally impose curvature constraints, this results in a very substantial reduction in the number of streamlines that propagate out of the shuffled region, with the shortest arc lengths being most likely. This explains why most of the unlabelled voxels after shuffling are towards the centre of the thalamus, furthest from the surrounding, unshuffled white matter. Indeed, the relative proximity of each voxel to neighbouring white matter tracts is then the primary determinant of the point of exit from the thalamus, and hence the subsequently inferred connectivity. The changes in cluster volumes reported in Table 1 may be due to a transfer between clusters where paths out of the thalamus are relatively tortuous in the unshuffled data, as with the "tail" of the frontal cluster, most readily seen in Fig. 6, which may be largely transferred to the parietal–occipital cluster, which corresponds to closer white matter.

There are fewer disconnected voxels in models with more complex fibre orientation distributions—two and three stick BEDPOSTX models and CSD—because the fibre tracking process is more likely to be able to identify a viable orientation at each step which does not violate its curvature constraints. In the higher-resolution 3 T dataset, there were fewer disconnected voxels still, perhaps due to the fact that the step distance—which is fixed to 0.5 mm in the TractoR tractography—is larger relative to the voxel size. Finally, the default "iFOD2" algorithm used by MRtrix for tractography is a second-order technique that considers plausible streamline arcs rather than just the immediate local fibre orientation information (Tournier et al., 2010), which may explain why it produces somewhat greater label agreement before and after shuffling.

The consequences of our findings, and the appropriate mitigations for future studies, will depend on the application. A sensible approach in basic neuroscience and clinical studies that treat the parcellation as a primary output would be to seek convergent evidence from other imaging or electrophysiological sources (Glasser et al., 2016; Kelly et al., 2012; Wang et al., 2015). Applying a threshold to the number of connected streamlines needed for a voxel to receive a colour, as demonstrated by Behrens et al. (2003), would also help to avoid tenuous labelling—although such thresholds are inevitably arbitrary. By contrast, in certain clinical contexts it may be that identifying the core locus of connectivity to a particular target region is more important than the parcellation as a whole, and in such cases representations such as those in Fig. 6 would be of more direct relevance, as well as helping to give a greater sense of confidence in the putative subnucleus of interest.

Such estimates of confidence can also be incorporated into subsequent analysis. For example, we calculate that the confidence-weighted centre of mass of the precentral or motor cluster, a potential target for therapeutic stimulation in patients with tremor (Akram et al., 2018), is shifted from the unweighted centroid by an average of 1.2–2.1 mm (depending on the diffusion model used), or 9–16% of the cluster diameter, which was typically about 13 mm. This shift is directed towards the edge of the thalamus, nearer the white matter where anisotropy is higher and orientational uncertainty lower. In general, even in the unshuffled data, there is lesser confidence as one moves further into the thalamus. This trend is in addition to the substantial variation across diffusion models directly visible in Figs 3 and 5, and overall we may therefore conclude that the optimal target for an implantation is spatially uncertain. Even assuming the tractography is broadly accurate, there is therefore a need for caution if the risks of ineffective treatment or a need for further surgical implantation are to be minimised.

Although we cannot directly measure them in our experiments, tractography algorithms' in-built biases undoubtedly influence the parcellations obtained, and validation work has made it clear that tractography is far from flawless as a tract-tracing technique. Recently, Donahue et al. (2016) directly and systematically compared tractography in postmortem monkey brains to data from invasive tracer work, and concluded that tractography was only "modestly informative in predicting connection presence and weight". This is another reason to seek concordant evidence from other modalities where accuracy is important.

Of course, this study has a number of limitations. We have looked only at the thalamus—although, as we have noted, this is a canonical and well-studied region of interest for parcellation studies. Moreover, due to its relatively large size, issues

such as the disconnection of voxels toward the centre of the structure are more visible than might be the case elsewhere. For smaller structures completely surrounded by white matter, such as the subthalamic nucleus, a substantially higher proportion of connected voxels would be expected in the shuffled data, and this may also translate into greater agreement in the parcellation, when compared to the unshuffled original. Additionally, we have considered only the winner-takes-all approach to parcellation, rather than any of the various clustering-based methods. This is partly for simplicity, and partly because winner-takes-all is relatively unambiguous, both methodologically and in terms of its interpretation, whereas more complex multi-target methods (e.g., as used in Draganski et al., 2008) involve more nuance. For example, approaches based on clustering the patterns of connectivity to multiple targets would vary due to the clustering algorithm as well as the streamline distribution, rendering interpretation more difficult. Finally, we have focussed on approaches that seed within deep grey matter itself, whereas several parcellation studies propagate streamlines only from the boundary between grey matter and white matter, where anisotropy will generally be higher and orientational uncertainty therefore lower (Gallardo et al., 2018; Moreno-Dominguez et al., 2014; St-Onge et al., 2018). A wholly different category of experiments would be needed to investigate the interpretability of those approaches.

In conclusion, we advocate a more cautious interpretation of connectivity-based parcellations than has been prevalent in the literature, and careful scrutiny of the data in future work. We have shown that applying a substantial corruption to diffusion datasets has relatively little effect on such parcellations—indeed, no bigger than switching tractography algorithms. Particularly with modern, multi-compartment diffusion models and tractographic biases such as the tendency to favour directional continuity, algorithms are able to string together paths through even unreliable fibre-orientation fields and build a parcellation by brute force. Nevertheless, our findings have no direct bearing on the correctness of the parcellation—indeed, a parcellation of the thalamus with frontal connectivity towards the anterior margin and occipital connectivity posterior is logical from the point of view of economical wiring in the brain (Laughlin & Sejnowski, 2003; Van Essen, 1997). But the lack of sensitivity to internal structure undermines the principle of mapping spatially specific, voxel-to-region connectivity *in vivo*, which is the technique's core idea.

Acknowledgments

Appendix: Random shuffling

Any given classification of N voxels separates them into n groups containing m_i elements each, with $i \in \{1, 2, \ldots, n\}$, such that $\sum_i m_i = N$. A comparison of the overlap between two such classifications should consider as a baseline the degree of overlap that would be expected by chance.

The number of possible permutations of the labels is given by the multinomial coefficient,

$$\binom{N}{m_1, m_2, \ldots, m_n} = \frac{N!}{m_1! \, m_2! \, \cdots \, m_n!} \,.$$

After a random permutation of the labels, the number of cases in which a voxel from group i retains that label is

$$\frac{(N-1)!}{(m_i - 1)! \, \prod_{j \neq i} m_j!} \,,$$

the number of combinations of all elements except the one staying the same.

Taking X_l, with $l \in \{1, 2, \ldots, N\}$ indexing over all voxels, to be the random variable with value 1 if voxel l from group i retains its label, and 0 otherwise, we divide the two previous equations to obtain

$$\begin{aligned}
\mathbb{E}(X_l) &= \left(\frac{(N-1)!}{(m_i - 1)! \, \prod_{j \neq i} m_j!} \right) \left(\frac{m_1! \, m_2! \, \cdots \, m_n!}{N!} \right) \\
&= \frac{(N-1)! \, m_i!}{(m_i - 1)! \, N!} \\
&= \frac{m_i}{N} \,.
\end{aligned}$$

Taking $X = X_1 + X_2 + \ldots + X_N$, we can make use of the linearity of expectations to quickly determine that

$$\mathbb{E}(X) = \sum_l \mathbb{E}(X_l) = \sum_{i=1}^{n} m_i \frac{m_i}{N} = \frac{1}{N} \sum_{i=1}^{n} m_i^2 \,.$$

This process is a permutation, and therefore assumes that the proportions of each label remain exactly equal. However, if instead we draw voxel labels at random such that the probability of each voxel having label i is m_i/N, independently of all other voxels, the expectation does not change. Clearly, $\Pr(X_l = 1) = m_i/N$, since voxel l originally had label i, and so we can proceed using linearity of expectations as above. Assuming *fixed probabilities* therefore has the same expected outcome as assuming *fixed proportions*.

References

Akram H., Dayal V., Mahlknecht P., Georgiev D., Hyam J., Foltynie T., Limousin P., De Vita E., Jahanshahi M., Ashburner J., Behrens T., Hariz M. & Zrinzo L. (2018). Connectivity derived thalamic segmentation in deep brain stimulation for tremor. *NeuroImage: Clinical* **18**:130–142.

Akram H., Sotiropoulos S.N., Jbabdi S., Georgiev D., Mahlknecht P., Hyam J., Foltynie T., Limousin P., De Vita E., Jahanshahi M., Hariz M., Ashburner J., Behrens T. & Zrinzo L. (2017). Subthalamic deep brain stimulation sweet spots and hyperdirect cortical connectivity in Parkinson's disease. *NeuroImage* **158**:332–345.

Andersson J.L.R., Skare S. & Ashburner J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage* **20**(2):870–888.

Andersson J.L.R. & Sotiropoulos S.N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* **125**:1063–1078.

Anwander A., Tittgemeyer M., von Cramon D.Y., Friederici A.D. & Knösche T.R. (2007). Connectivity-based parcellation of Broca's area. *Cerebral Cortex* **17**(4):816–825.

Bach D.R., Behrens T.E., Garrido L., Weiskopf N. & Dolan R.J. (2011). Deep and superficial amygdala nuclei projections revealed in vivo by probabilistic tractography. *Journal of Neuroscience* **31**(2):618–623.

Behrens T.E.J., Johansen-Berg H., Jbabdi S., Rushworth M.F.S. & Woolrich M.W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* **34**(1):144–155.

Behrens T.E.J., Johansen-Berg H., Woolrich M.W., Smith S.M., Wheeler-Kingshott C.A.M., Boulby P.A., Barker G.J., Sillery E.L., Sheehan K., Ciccarelli O., Thompson A.J., Brady J.M. & Matthews P.M. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience* **6**(7):750–757.

Clayden J.D., Muñoz Maniega S., Storkey A.J., King M.D., Bastin M.E. & Clark C.A. (2011). TractoR: Magnetic resonance imaging and tractography with R. *Journal of Statistical Software* **44**(8):1–18.

da Silva N.M., Ahmadi S.A., Tafula S.N., Cunha J.P.S., Bötzel K., Vollmar C. & Rozanski V.E. (2017). A diffusion-based connectivity map of the GPi for optimised stereotactic targeting in DBS. *NeuroImage* **144**(Pt A):83–91.

Deligianni F., Carmichael D.W., Zhang G.H., Clark C.A. & Clayden J.D. (2016). NODDI and tensor-based microstructural indices as predictors of functional connectivity. *PLoS ONE* **11**(4):e0153,404.

Desikan R.S., Ségonne F., Fischl B., Quinn B.T., Dickerson B.C., Blacker D., Buckner R.L., Dale A.M., Maguire R.P., Hyman B.T., Albert M.S. & Killiany R.J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**(3):968–980.

Donahue C.J., Sotiropoulos S.N., Jbabdi S., Hernandez-Fernandez M., Behrens T.E., Dyrby T.B., Coalson T., Kennedy H., Knoblauch K., Van Essen D.C. & Glasser M.F. (2016). Using diffusion tractography to predict cortical connection strength and distance: A quantitative comparison with tracers in the monkey. *Journal of Neuroscience* **36**(25):6758–6770.

Draganski B., Kherif F., Klöppel S., Cook P.A., Alexander D.C., Parker G.J.M., Deichmann R., Ashburner J. & Frackowiak R.S.J. (2008). Evidence for segregated and integrative connectivity patterns in the human basal ganglia. *Journal of Neuroscience* **28**(28):7143–7152.

Eickhoff S.B., Thirion B., Varoquaux G. & Bzdok D. (2015). Connectivity-based parcellation: Critique and implications. *Human Brain Mapping* **36**(12):4771–4792.

Gallardo G., Wells W., Deriche R. & Wassermann D. (2018). Groupwise structural parcellation of the whole cortex: A logistic random effects model based approach. *NeuroImage* **170**:307–320.

Glasser M.F., Coalson T.S., Robinson E.C., Hacker C.D., Harwell J., Yacoub E., Ugurbil K., Andersson J., Beckmann C.F., Jenkinson M., Smith S.M. & Van Essen D.C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* **536**(7615):171–178.

Guillery R. & Sherman S. (2002). Thalamic relay functions and their role in corticocortical communication. *Neuron* **33**(2):163–175.

Jakab A., Molnár P.P., Bogner P., Béres M. & Berényi E.L. (2012). Connectivity-based parcellation reveals interhemispheric differences in the insula. *Brain Topography* **25**(3):264–271.

Johansen-Berg H., Behrens T.E.J., Robson M.D., Drobnjak I., Rushworth M.F.S., Brady J.M., Smith S.M., Higham D.J. & Matthews P.M. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences of the USA* **101**(36):13,335–13,340.

Kelly C., Toro R., Di Martino A., Cox C.L., Bellec P., Castellanos F.X. & Milham M.P. (2012). A convergent functional architecture of the insula emerges across imaging modalities. *NeuroImage* **61**(4):1129–1142.

Klein J.C., Behrens T.E.J., Robson M.D., Mackay C.E., Higham D.J. & Johansen-Berg H. (2007). Connectivity-based parcellation of human cortex using diffusion MRI: Establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *NeuroImage* **34**(1):204–211.

Krauth A., Blanc R., Poveda A., Jeanmonod D., Morel A. & Székely G. (2010). A mean three-dimensional atlas of the human thalamus: Generation from multiple histological data. *NeuroImage* **49**(3):2053–2062.

Lambert C., Zrinzo L., Nagy Z., Lutti A., Hariz M., Foltynie T., Draganski B., Ashburner J. & Frackowiak R. (2012). Confirmation of functional zones within the human subthalamic nucleus: Patterns of connectivity and subparcellation using diffusion weighted imaging. *NeuroImage* **60**(1):83–94.

Laughlin S.B. & Sejnowski T.J. (2003). Communication in neuronal networks. *Science* **301**(5641):1870–1874.

Magnotta V.A., Gold S., Andreasen N.C., Ehrhardt J.C. & Yuh W.T. (2000). Visualization of subthalamic nuclei with cortex attenuated inversion recovery MR imaging. *NeuroImage* **11**(4):341–346.

Maier-Hein K.H., Neher P.F., Houde J.C., Côté M.A., Garyfallidis E., Zhong J., Chamberland M., Yeh F.C., Lin Y.C., Ji Q., Reddick W.E., Glass J.O., Chen D.Q., Feng Y., Gao C., Wu Y., Ma J., Renjie H., Li Q., Westin C.F., Deslauriers-Gauthier S., González J.O.O., Paquette M., St-Jean S., Girard G., Rheault F., Sidhu J., Tax C.M.W., Guo F., Mesri H.Y., Dávid S., Froeling M., Heemskerk A.M., Leemans A., Boré A., Pinsard B., Bedetti C., Desrosiers M., Brambati S., Doyon J., Sarica A., Vasta R., Cerasa A., Quattrone A., Yeatman J., Khan A.R., Hodges W., Alexander S., Romascano D., Barakovic M., Auría A., Esteban O., Lemkaddem A., Thiran J.P., Cetingul H.E., Odry B.L., Mailhe B., Nadar M.S., Pizzagalli F., Prasad G., Villalon-Reina J.E., Galvis J., Thompson P.M., Requejo F.D.S., Laguna P.L., Lacerda L.M., Barrett R., Dell'Acqua F., Catani M., Petit L., Caruyer E., Daducci A., Dyrby T.B., Holland-Letz T., Hilgetag C.C., Stieltjes B. & Descoteaux M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications* **8**(1):1349.

Modat M., Ridgway G.R., Taylor Z.A., Lehmann M., Barnes J., Hawkes D.J., Fox N.C. & Ourselin (2010). Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine* **98**(3):278–284.

Morel A., Magnin M. & Jeanmonod D. (1997). Multiarchitectonic and stereotactic atlas of the human thalamus. *Journal of Comparative Neurology* **387**(4):588–630.

Moreno-Dominguez D., Anwander A. & Knösche T.R. (2014). A hierarchical method for whole-brain connectivity-based parcellation. *Human Brain Mapping* **35**(10):5000–5025.

Morris D.M., Embleton K.V. & Parker G.J.M. (2008). Probabilistic fibre tracking: Differentiation of connections from chance events. *NeuroImage* **42**(4):1329–1339.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Saygin Z.M., Osher D.E., Augustinack J., Fischl B. & Gabrieli J.D.E. (2011). Connectivity-based segmentation of human amygdala nuclei using probabilistic tractography. *NeuroImage* **56**(3):1353–1361.

Setsompop K., Cohen-Adad J., Gagoski B.A., Raij T., Yendiki A., Keil B., Wedeen V.J. & Wald L.L. (2012). Improving diffusion MRI using simultaneous multi-slice echo planar imaging. *NeuroImage* **63**(1):569–580.

Smith S.M. (2002). Fast robust automated brain extraction. *Human Brain Mapping* **17**(3):143–155.

Smith S.M., Jenkinson M., Woolrich M.W., Beckmann C.F., Behrens T.E.J., Johansen-Berg H., Bannister P.R., De Luca M., Drobnjak I., Flitney D.E., Niazy R.K., Saunders J., Vickers J., Zhang Y., De Stefano N., Brady J.M. & Matthews P.M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23 Suppl 1**:S208–19.

St-Onge E., Daducci A., Girard G. & Descoteaux M. (2018). Surface-enhanced tractography (SET). *NeuroImage* **169**:524–539.

Tournier J.D., Calamante F. & Connelly A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* **35**(4):1459–1472.

Tournier J.D., Calamante F. & Connelly A. (2010). Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. In *Proceedings of the International Society for Magnetic Resonance in Medicine*. ISMRM.

Tournier J.D., Calamante F. & Connelly A. (2013). Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. *NMR in Biomedicine* **26**(12):1775–1786.

Tuch D.S. (2002). *Diffusion MRI of complex tissue structure*. Ph.D. thesis, Massachusetts Institute of Technology.

Van Essen D.C. (1997). A tension-based theory of morphogenesis and compact

wiring in the central nervous system. *Nature* **385**(6614):313–318.

Wang J., Yang Y., Fan L., Xu J., Li C., Liu Y., Fox P.T., Eickhoff S.B., Yu C. & Jiang T. (2015). Convergent functional architecture of the superior parietal lobule unraveled with multimodal neuroimaging approaches. *Human Brain Mapping* **36**(1):238–257.

Wickham H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer, New York.