TECHNICAL NOTE

The Precision of Anatomical Normalization in the Medial Temporal Lobe Using Spatial Basis Functions

C. H. Salmond,*,† J. Ashburner,‡ F. Vargha-Khadem,* A. Connelly,† D. G. Gadian,† and K. J. Friston‡

*Developmental Cognitive Neuroscience Unit and †Radiology and Physics Unit, Institute of Child Health, University College London,
30 Guilford Street, London WC1N 1EM, United Kingdom; and ‡The Wellcome Department of Imaging Neuroscience,
Institute of Neurology, University College London, Queen Square, London WC1N 1EM, United Kingdom

We investigated the accuracy of spatial basis function normalization using anatomical landmarks to determine how precisely homologous regions are colocalized. We examined precision in terms of: (1) the number of nonlinear basis functions used by the normalization procedure; (2) the degree of (Bayesian) regularization; and (3) the effect of substituting different templates and how this interacted with the number of basis functions. The face validity of spatial normalization was assessed as a function of these parameters, using the colocalization of homologous landmarks in a test sample of 20 normally developing children and 5 children with bilateral hippocampal pathology. Our results suggest that when optimal normalization parameters are used, anatomical landmarks in the medial temporal lobes are colocalized to within a standard deviation of about 1 mm. When suboptimal parameters are used this standard deviation can increase up to 3 mm. Interestingly the optimal parameters are those that provide a rather constrained normalization as opposed to those that optimize intensity matching at the expense of rendering the warps "unlikely." The implications of our results, for users of voxel-based morphometry, are discussed. © 2002 Elsevier Science (USA)

INTRODUCTION

This technical note deals with the precision of spatial normalization in terms of its ability to colocalize a number of medial temporal lobe anatomical landmarks, from subject to subject. The motivation for the work reported came from a series of neurodevelopmental studies looking at bilateral medial temporal lobe abnormalities using voxel-based morphometry. In particular, we were interested in assessing the construct validity of nonlinear warping in terms of landmark coregistration. The results do not represent a comprehensive validation of the spatial normalization procedures we used but do point to the interesting fact that precision, in terms of landmark coregistration, is not necessarily achieved by increasing the degrees of freedom of the nonlinear transformations used to normalize brain images.

Voxel-based morphometry was developed to characterize cerebral gray and white matter differences in structural MRI scans. In contrast to methods that frame the search in terms of regions of interest, voxel-based morphometry can detect structural differences with uniform sensitivity throughout the brain. Voxel-based morphometry is essentially a technique that compares images of gray matter (or white matter) (obtained from segmented MR images). This comparison uses statistical parametric mapping to identify, and make inferences about, regionally specific differences.

Voxel-based morphometry depends on spatially normalizing all the images into the same stereotactic space, extracting the gray (or white) matter from the normalized images, smoothing, and finally performing a statistical analysis to localize and make inferences about group differences. The output of the method is a statistical parametric map (SPM) showing regions where gray (or white) matter density differs significantly among groups.

Normalization is required to ensure that homologous regions are compared across subjects. It is not intended to be an exact process (since this would remove all anatomical differences among the groups, rendering the differences in gray matter segments negligible). The normalization should therefore remove only anatomical differences down to a specified spatial scale so that structures are registered but their relative tissue composition is preserved. The accuracy of colocalization of anatomical structures across subjects is a crit-

ical factor in ensuring that the results can be interpreted in terms of gray matter changes per se.

Spatial Normalization

Normalization is essentially the process of warping MR images from different subjects into a standard space as defined by a template image (Friston *et al.,* 1995). One way of normalizing data involves identifying certain brain "landmarks" by inspection and then manipulating the images so that these landmarks are brought into register. However, to overcome the inherent subjectivity of this method, voxel-based analyses generally use non-label-based techniques that are fully automated.

Non-label-based normalization techniques minimize some index of the difference between the source image and the template image (Ashburner and Friston, 1999a,b). SPM99 minimizes the sum of the squared differences between the image to be normalized and the template while maximizing the prior probability of the transformation (also known as regularization). The maximum a posteriori solution is found iteratively: the algorithm starts with an initial parameter estimate and searches from there. The algorithm stops when criterion is achieved (when the weighted sum of square differences no longer decreases or after a finite number of iterations).

Normalization can be divided into two components: affine (or linear) and nonlinear transformations. Affine transformations are generally carried out first and account for differences in position, orientation, and overall brain size. Then nonlinear normalization is used to account for low spatial frequency global variability in head shape. In SPM99, the nonlinear transformations are restricted to linear combinations of three-dimensional discrete cosine functions (Ashburner and Friston, 1999b).

The normalization is therefore shaped by a number of constraints: hard constraints (such as the number of discrete cosine functions employed) and soft constraints (such as the degree of regularization). The construct validity of the normalization procedure has been addressed by Ashburner and Friston (1999c). These investigations focused on ensuring that the normalization procedure minimizes various measures of template and data differences (such as membrane energy). In the present work, we assess face validity in terms of the precision of the normalization procedure, by examining the variation in the coordinate location of various anatomical landmarks across a group of individuals. The effects of changing the constraints on the nonlinear transformations as well as the dependency of the results on the template used are investigated.

Voxel-Based Morphometry

Although the issues addressed in this note are relevant to all forms of neuroimaging (that encompass multisubject studies), we have focused on their relevance for voxel-based morphometry (VBM) in the medial temporal region. Voxel-based morphometry is just one of a number of computational techniques that can be used to assess anatomical differences among different cohorts.

There are essentially three different approaches to this sort of computational anatomy. The first is voxel-based morphometry in which the warpings used to normalize each subject's brain are constrained to be relatively smooth. Critically, this smoothness must be greater than the kernel used to smooth the gray matter partitions when creating gray matter density maps. The idea behind voxel-based morphometry is to place anatomical regions in register but not change their relative local tissue composition. This is ensured by using smooth warps that, to a first approximation, simply move structures around without dilating or compressing them at the spatial scale at which inferences are made.

The alternative approach is to use warps that have many more degrees of freedom, allowing them to remove both position and size or shape differences down to the finest anatomical scale. Following normalization, with these high-dimensional techniques, all the information pertaining to anatomical differences is encoded by the deformation fields. These differences can then be assessed by comparing the deformation fields directly (deformation field-based morphometry) or using tensor-based morphometry. Tensor-based morphometry retains the regional specificity of voxel-based morphometry by computing a sensible scalar metric at each point in the brain from the deformation field. The most common metric used is the Jacobian, which can be thought of as a measure of local volume change (see Christensen *et al.,* 1995; Davatzikos, 2001; Thompson *et al.,* 2000).

In this work we were concerned exclusively with voxel-based morphometry and smooth deformations. This smoothness is enforced by expressing the deformation field as a linear combination of smooth basis functions. The degree of smoothness is controlled by how many basis functions are used and this is a key parameter we manipulated to assess precision in terms of landmark coregistration.

CHANGING THE NUMBER OF BASIS FUNCTIONS (HARD CONSTRAINTS)

Materials and Methods

All subjects (20 children, mean age 13 years, 11 males, 9 females, with no known neurological or psychiatric history) were scanned, unsedated, on a 1.5-T

TABLE 1

Anatomical Landmark Locations

Label	Category	z coordinate	Localization
Front	Basic landmarks	In-plane $z = 0$	Maximal y value
Back	Basic landmarks	In-plane $z = 0$	Minimal y value
Left	Basic landmarks	In-plane $z = 0$	Minimal x value
Right	Basic landmarks	In-plane $z = 0$	Maximal x value
Left white matter	Medial temporal landmarks	In coronal plane through superior pons–brain-stem join (chosen in the sagittal plane $x = 0$)	Most medial white matter on lateral border of left hippocampus
Right white matter	Medial temporal landmarks	In coronal plane through superior pons–brain-stem join (chosen in the sagittal plane $x = 0$)	Most medial white matter on lateral border of right hippocampus
Left pons	Medial temporal landmarks	In coronal plane through superior pons–brain-stem join (chosen in the sagittal plane $x = 0$)	Widest point of the pons brain stem on left
Right pons	Medial temporal landmarks	In coronal plane through superior pons–brain-stem join (chosen in the sagittal plane $x = 0$)	Widest point of the pons brain stem on right

Siemens Vision scanner, using a T1-weighted 3D MPRAGE sequence (Mugler and Brookeman, 1990) with the following parameters: TR 9.7 ms, TE 4 ms, TI 300 ms, flip angle 12°, matrix size $256 \times 256 \times 128$, field of view $250 \times 250 \times 160$ mm.

The 3D data sets were analyzed in SPM99 (Wellcome Department of Imaging Neuroscience, London, UK). Each scan was normalized (Friston *et al.,* 1995; Ashburner and Friston, 1999a) to the T1 template supplied with SPM99. This template is constructed from 152 T1-weighted scans from the MNI [supplied by Alan Evans, Montreal Neurological Institute, Canada (ICBM, NIH P-20 project, Principal Investigator John Mazziotta)]. Normalization was performed using different numbers of nonlinear basis functions in three orthogonal directions to give four levels of constraint:

- Level 1—No nonlinear transformations
- Level 2—$4 \times 5 \times 4$ nonlinear basis functions
- Level 3—$7 \times 8 \times 7$ nonlinear basis functions
- Level 4—$10 \times 11 \times 10$ nonlinear basis functions

All other normalization parameters were constant across the groups (medium regularization, 16 iterations). To assess face validity, eight anatomical landmarks were identified as described in Table 1. These comprised "bounding box"-like landmarks and specific landmarks in the medial temporal region (see Fig. 1). We were particularly interested in medial temporal precision, given our analyses below of data from children with hippocampal atrophy.

The coordinates of these landmarks were measured on each normalized scan, with the assessor blinded to the level of constraint. The mean position for each anatomical landmark was calculated within each level and the standard deviation about this position was calculated. Measurements were carried out twice on Level 3 to assess test–retest reliability. The precision

data (i.e., the standard deviation of the eight landmark locations) were analyzed in SPSS using repeated measures and Pearson's correlation coefficients.

Results and Discussion

The reliability of anatomical landmark identification was confirmed by Pearson's correlation coefficients for all eight landmarks. All correlation coefficients were significant at $P < 0.05$ ($r > 0.5$). Figure 2 shows that the four landmarks in the plane $z = 0$ were colocalized to within ~1 mm (<1 voxel) regardless of how many nonlinear basis functions were used. This is a reflection of the fact that these "basic" aspects of colocalization are well accommodated by linear transformations. The four landmarks in the coronal plane through the pons intersection were significantly less well localized than these basic measures [$F(3.7, 71.1) = 35$; $P < 0.001$]. In addition, from Fig. 2, it can be seen that $4 \times 5 \times 4$ nonlinear basis functions give significantly better colocalization (significantly smaller standard deviation from mean anatomical location) for all four medial temporal landmarks.

These results suggest that a smaller number of nonlinear basis functions can be "better" than a larger number. This may appear counterintuitive, since more basis functions endow the transformation with more degrees of freedom, which in turn should make the normalization more accurate. However, this intensity matching is not spatial matching and does not necessarily give the most spatially precise registration. In other words, it is possible that while the algorithm minimizes the difference between the template and the image, the amount of regularization fails to prevent anatomically unlikely transformations, i.e., the algorithm prioritizes reducing the difference between the template and the image at the expense of anatomical
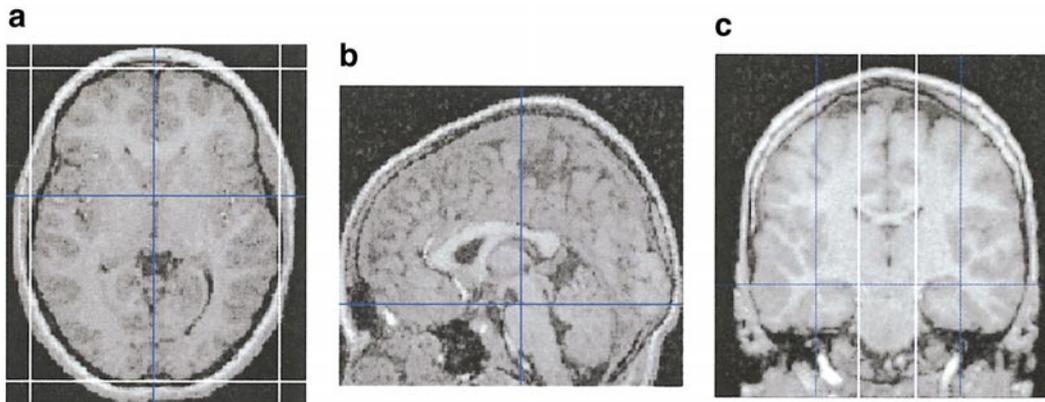
FIG. 1. Anatomical landmarks, (a) Landmarks in $z = 0$. (b) The coronal slice for white matter and pons landmarks. (c) Medial temporal landmarks.

validity. This possibility was investigated in the next analyses, in which we looked for evidence of an interaction between the basis functions and regularization.

CHANGING THE DEGREES OF REGULARIZATION (SOFT CONSTRAINTS)

Regularization refers to the inclusion of a penalty term for unlikely warps (based on specified priors) that enter into the minimization. This log likelihood term penalizes rough, quickly changing warps. Operationally this implies penalizing the use of high spatial frequency basis functions more than lower frequencies (see Ashburner and Friston, 1999b).

Materials and Methods

The 3D data sets (as described above) were normalized with two different sets of nonlinear basis functions with high regularization to form two additional normalization levels:
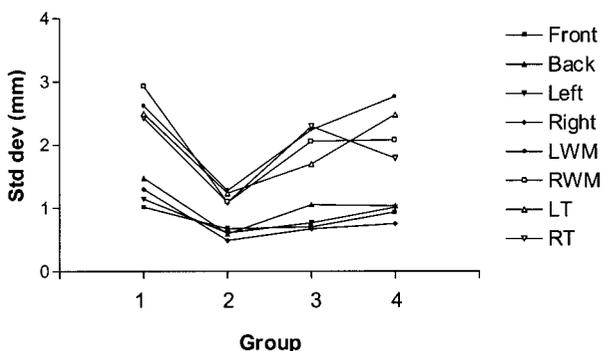
• Level 2A—$4 \times 5 \times 4$ nonlinear basis functions

• Level 3A—$7 \times 8 \times 7$ nonlinear basis functions

The data were otherwise normalized exactly as described above. The coordinates of landmarks were measured and analyzed as in the previous section.

Results and Discussion

As Fig. 3 shows, the normalization precision was unchanged by increasing regularization with $7 \times 8 \times 7$ nonlinear basis functions. With $4 \times 5 \times 4$ nonlinear basis functions, the accuracy was reduced by increasing the regularization. These results suggest that reducing the importance of minimizing the template and image difference relative to the probability of the anatomical warping interacts with the number of basis functions in terms of face validity. Regularizing the larger basis set has little effect. However, regularizing the smaller (optimal for these eight landmarks) set overly constrains the warp and reduces the precision. These findings suggest that having a greater number of basis functions does not necessarily improve normalization accuracy and the hard constraints offered by basis functions can actually lead to better spatial coregistration.

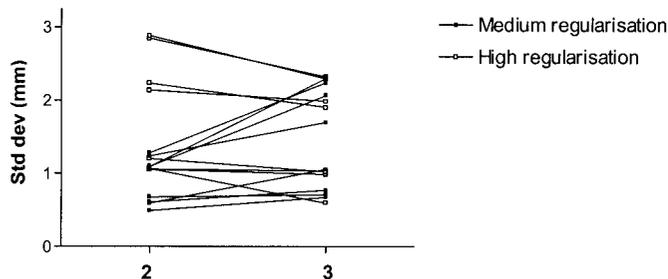It should be noted that while $4 \times 5 \times 4$ nonlinear basis functions gave the most accurate normalization

FIG. 2. Standard deviation from mean position of landmarks using medium regularization. Group 1, no nonlinear transformations; Group 2, $4 \times 5 \times 4$ nonlinear transformations; Group 3, $7 \times 8 \times 7$ nonlinear transformations; Group 4, $10 \times 11 \times 10$ nonlinear transformations.

FIG. 3. Standard deviation from mean position of landmarks using varying levels of regularization. Group 2, $4 \times 5 \times 4$ nonlinear transformations; Group 3, $7 \times 8 \times 7$ nonlinear transformations.
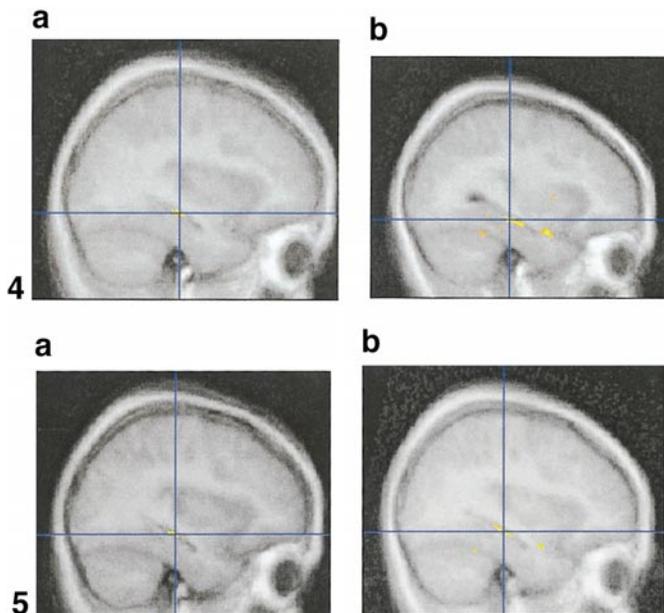
FIG. 4. Significant areas of decreased gray matter in the amnesic children versus controls in left hippocampal formation ($7 \times 8 \times 7$ nonlinear basis functions). Results are superimposed on the mean normalized images and thresholded at uncorrected $P < 0.001$. (a) GOS template. (b) T1 template.

FIG. 5. Significant areas of decreased gray matter in the amnesic children versus controls in left hippocampal formation ($4 \times 5 \times 4$ nonlinear basis functions). Results are superimposed on the mean normalized images and thresholded at uncorrected $P < 0.001$ (a) GOS template. (b) T1 template.

in the range tested, it may not be the optimal number for all landmarks. However since $4 \times 5 \times 4$ nonlinear basis functions colocalized all eight landmarks to within ~1.5 mm and this is the size of the voxels in the normalized images, it is likely that this accuracy is adequate for most applications.

CHANGING THE TEMPLATE

Normalization procedures are intended to provide good anatomical colocalization in relation to the template the data are normalized to. The more basis functions used in the nonlinear normalization, the greater the possibility that the data are overfitted, rendering the results sensitive to the choice of template. In this section we investigated whether the colocalization accuracy was template dependent. Having established registration accuracy in the medial temporal lobe to within 3 mm using the landmark approach, we proceeded to examine the impact of changing templates on the anatomical precision of VBM per se. To do this we used a test sample consisting of patients with bilateral hippocampal atrophy. To extend the analysis of the previous sections, we also assessed the effect of template on landmark colocalization using $7 \times 8 \times 7$ basis functions.

Materials and Methods

Five patients (mean age 12.4, four males, one female) who had developmental amnesia associated with early hypoxic–ischemic episodes comprised the test sample. All these patients have been shown to have bilateral hippocampal atrophy using volumetric methods in addition to VBM (Gadian *et al.,* 2000). Eight controls (mean age 13.9, three males, five females) were selected for comparison. All subjects were scanned unsedated on a 1.5-T Siemens Vision scanner, using a T1-weighted 3D MPRAGE sequence as described before.

The 3D data sets were analyzed in SPM99 (Wellcome Department of Imaging Neuroscience). Normalization was carried out using two different templates: the MNI template (supplied with SPM99, 152 adult T1 scans) and an inhouse template (GOS template; 27 children scanned on the same scanner with the same image acquisition sequence as for the subjects used in this analysis; mean age 14). Normalization was carried out using two different sets of nonlinear functions: $4 \times 5 \times 4$ and $7 \times 8 \times 7$ with medium regularization. The data were otherwise normalized as described above. The data were then segmented (Ashburner and Friston, 1997), smoothed with a Gaussian isotropic kernel of FWHM 4 mm, and analyzed to look for decreases in gray matter density in the amnesic patients versus the controls. Gray matter density is simply the average volume of MRI-classified gray matter per unit volume of the brain.

The extent and location of the abnormalities within the hippocampal formation were then compared between the two templates at both $4 \times 5 \times 4$ and $7 \times 8 \times 7$ nonlinear basis function levels.

In addition, the scans of the 20 controls used in previous sections were normalized to the GOS template with $7 \times 8 \times 7$ nonlinear basis functions to determine whether the template chosen affected the colocalization precision.

Results and Discussion

Figure 4 shows the results of the analysis (with $7 \times 8 \times 7$ nonlinear basis functions) of the left hippocampal formation. The extent of the abnormality appeared greater using the GOS template relative to the MNI template analysis. In contrast, Fig. 5 shows that using $4 \times 5 \times 4$ nonlinear basis functions results in similar extent and location abnormalities within the left hippocampal formation [this pattern is also present in the right hippocampal formation (not shown)].

Figures 4 and 5 suggest that when $4 \times 5 \times 4$ nonlinear basis functions are used, the results of the analyses are minimally dependent on the template used. This suggests that use of the default template even comparing children is suitable when using the optimal number of nonlinear basis functions in the medial temporal lobe. However, with larger basis function sets,

the template does appear to affect the results of the analysis. Investigators using a larger number of basis functions need to consider which template might be most appropriate for their study and allow for the possibility of overfitting. Template dependencies such as those seen in Fig. 4 remain when the smoothing was raised to 8 mm from 4 mm FWHM (data not shown).

There was no significant difference between the normalization precision of any of the eight landmarks regardless of which template was used. It is important to note that the precision of the normalization, as reflected in the landmark colocalization, does not depend on the template. This is to be expected as long as the warps required to match an image to the template are not too "unlikely." This finding moderates enthusiasm for "custom" templates unless one is dealing with very different brains (e.g., patients with gross pathology).

Finally, it should be noted that the stereotactic space, defined operationally by the template used for intensity matching during spatial normalization, is not necessarily the best coordinate system to minimize registration error.

CONCLUSION

In conclusion, our results suggest that when optimum normalization parameters are used, anatomical landmarks in the medial temporal lobe are colocalized to within a standard deviation of about 1 mm. When suboptimal parameters are used this standard deviation can increase up to 3 mm. Interestingly the optimal parameters are those that provide a more constrained normalization. Although less constrained normalizations allow for a more precise intensity matching they do so at the expense of making the warps more unlikely. The face validity of the normalization was not improved by the use of "custom" templates, even for children.

Restricting the number of basis functions, which produced empirically better results, effectively meant that warps above a certain frequency were penalized infinitely. This could imply that the form of regularization used by the warping algorithm may need to impose much higher penalties against high-frequency deformations. Currently the model minimizes the membrane energy of the warps, which is based on the sum of squares of the first derivatives of the deformations. It is possible that the regularization should minimize the sum of squares of a higher derivative. Regularization based on higher order derivatives has the effect of increasing the penalty against higher frequency deformations relative to those of lower frequencies. These forms may turn out to be better models of the true distribution of warps likely to be needed for registering brain images.

REFERENCES

Ashburner, J., and Friston, K. 1997. Multimodal image coregistration and partitioning—A unified framework. *NeuroImage* **6:** 209–217.

Ashburner, J., and Friston, K. J. 1999a. Spatial normalisation. In *Brain Warping* (A. W. Toga, Ed.). Academic Press, San Diego.

Ashburner, J., and Friston, K. J. 1999b. Nonlinear spatial normalisation using basis functions. *Hum. Brain Mapping* **7:** 254–266.

Ashburner, J., and Friston, K. 1999c. Voxel based morphometry—The methods. *NeuroImage* **11:** 205–218.

Christensen, G. E., Rabbitt, R. D., Miller, M. I., Joshi, S. C., Grenander, U., Coogan, T. A., and Van Essen, D. C. 1995. Topological properties of smooth anatomic maps. In *Proc. Information Processing in Medical Imaging* (Y. Bizais., C. Barillot, and R. Di Paola, Eds), pp. 101–12. Kluwer Academic, Dordrecht.

Davatzikos, C., Genc, A., Xu, D., and Resnick, S. M. 2001. Voxel-based morphometry using the RAVENS maps: Methods and validation using simulated longitudinal atrophy. *NeuroImage* **14:** 1361–1369.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., and Frackowiak, R. S. J. 1995. Spatial registration and normalisation of images. *Hum. Brain Mapping* **2:** 165–189.

Gadian, D. G., Aicardi, J., Watkins, K. E., Porter, D. A., Mishkin, M., and Vargha-Khadem, F. 2000. Developmental amnesia associated with early hypoxic–ischaemic injury. *Brain* **123:** 499–507.

Mugler, J. P., and Brookeman, J. R. 1990. Three dimensional magnetisation-prepared rapid gradient echo imaging. *Magn. Reson. Med.* **15:** 152–157.

Thompson, P. M., Giedd, J. N., Woods, R. P., MacDonald, D., Evans, A. C., and Toga, A. W. 2000. Growth patterns in the developing brain detected by using continuum mechanical tensor maps. *Nature* **404:** 190–193.