Mindboggle: Automated brain labeling with multiple atlases
Arno Klein*1,2, Brett Mensh3, Satrajit Ghosh4, Jason Tourville5 and Joy Hirsch1

Address: 1fMRI Research Center, Columbia University, New York, USA, 2Parsons Institute for Information Mapping, The New School, New York, USA, 3New York State Psychiatric Institute, Columbia University, New York, USA, 4Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, USA and 5Department of Cognitive and Neural Systems, Boston University, Boston, USA

Email: Arno Klein* - arno@binarybottle.com; Brett Mensh - brett@mensh.com; Satrajit Ghosh - satra@speech.mit.edu; Jason Tourville - jtour@bu.edu; Joy Hirsch - jh2155@columbia.edu

* Corresponding author

Abstract

Background: To make inferences about brain structures or activity across multiple individuals, one first needs to determine the structural correspondences across their image data. We have recently developed Mindboggle as a fully automated, feature-matching approach to assign anatomical labels to cortical structures and activity in human brain MRI data. Label assignment is based on structural correspondences between labeled atlases and unlabeled image data, where an atlas consists of a set of labels manually assigned to a single brain image. In the present work, we study the influence of using variable numbers of individual atlases to nonlinearly label human brain image data.

Methods: Each brain image voxel of each of 20 human subjects is assigned a label by each of the remaining 19 atlases using Mindboggle. The most common label is selected and is given a confidence rating based on the number of atlases that assigned that label. The automatically assigned labels for each subject brain are compared with the manual labels for that subject (its atlas). Unlike recent approaches that transform subject data to a labeled, probabilistic atlas space (constructed from a database of atlases), Mindboggle labels a subject by each atlas in a database independently.

Results: When Mindboggle labels a human subject's brain image with at least four atlases, the resulting label agreement with coregistered manual labels is significantly higher than when only a single atlas is used. Different numbers of atlases provide significantly higher label agreements for individual brain regions.

Conclusion: Increasing the number of reference brains used to automatically label a human subject brain improves labeling accuracy with respect to manually assigned labels. Mindboggle software can provide confidence measures for labels based on probabilistic assignment of labels and could be applied to large databases of brain images.

Background

When comparing structures or functions across brains, it is common to label the gross anatomy of brain image data and to compare the structures or functions that lie within anatomically labeled regions. Since brains differ in their anatomy [1-10], it would seem reasonable to refer to the
anatomy of many brains when labeling an individual subject's brain image. Atlases are manually labeled brains used as references. Using every atlas from a group of atlases independent of each other was found to give labeling results superior to those obtained by selecting the closest matching single atlas from the group, the average atlas, or an individual atlas, for the case of confocal microscopy images of bee brains [11]. However, labeling a subject's brain image with many different brains presents unreasonable demands on human labelers, who may not be consistent in their label assignments [12-15]. Fully automated labeling would facilitate large-scale labeling efforts while adding efficiency and consistency.

Image registration software (reviewed in [16-18]) may be used to coregister subject and atlas brain images, thereby labeling the subject images with superimposed atlas labels. There exist many different nonlinear image registration and feature-matching approaches to this problem [19-62]. Mindboggle software (see below) offers certain advantages over most of these approaches: it does not make the same assumptions about preserving topography from brain to brain, is relatively fast, and it performed well in comparison tests with standard image registration software packages (AIR, SPM2, ANIMAL, and linear registration with FLIRT) and in artificial lesion tests [63].

Having an automated registration or feature-matching program and a database of atlases introduces the problem of how to reconcile the multiple atlas label sets when labeling a single subject's brain. Labels could be assigned based on the selection or construction of similar or representative anatomy from these atlases. It is becoming more common to label subject brain image data with a single, composite atlas representing some average of multiple brain atlases (an average brain atlas) or retaining information about the differences between the atlases or between the atlases and the subject brain image (a probabilistic brain atlas).

Average brain atlases attempt to assign to each voxel (volume element) a representative value associated with image intensity or anatomical label. An intensity-based average brain atlas is the voxelwise mean intensity across individual brain images after linear [64-66] or nonlinear [67] coregistration. Additionally or alternatively, an average brain atlas may represent average sulcus shapes and positions computed in the original brain image space [8,68,69] or in an alternative space such as on a sphere [70]. An example of a label-based average brain atlas was constructed by Hammers, et. al. [71], where the majority label was computed for each voxel across 20 manually labeled brains after nonlinear registration to the MNI152 [64] template using SPM99 [27]. The use of an average atlas presupposes that there is such a thing as a representative brain and does not usually account for variability across brains.

Probabilistic brain atlases, on the other hand, do provide additional statistical information across the population used to construct the atlas [62,72-82]. This information may be related to the variance of landmark positions [73], probability of anatomical labels [44,79,83,84], probability of tissue classes [80], or multiple anatomical dimensions, for example characteristics of surface geometry and Bayesian priors associated with neighborhood relations between labels [62], and the multi-dimensional atlases under development by Mazziotta and Zilles and their colleagues [72,77,78,81]. An abstract representation of a database of manually labeled brains can also serve as a probabilistic atlas; for example, expert neural networks trained on a learning database of such brains [48] or graphs relating parametric surfaces [36]. However, there are only two examples known by the authors in which a complete cortical atlas is constructed from multiple label sets where each label set was assigned manually [62,71], rather than by automating the labeling of many brains without independent validation of the labeling technique. As with average brain atlases, probabilistic atlases have primarily been used as templates to which a subject brain is transformed and compared. This comparison presupposes that the single transform will account for differences between the subject brain image and each of the multiple brain images that were used to construct the atlas.
In this paper, we have chosen to extend the use of an individual atlas to multiple atlases in a recently introduced, fully automated, feature-based nonlinear labeling method called Mindboggle (freely downloadable, open source Matlab code) [63,85]. Rather than use a single (average or probabilistic) atlas, Mindboggle employs each atlas in a database independently to label the cortical voxels of a subject brain image, and for each voxel chooses the majority label assigned by the different atlases. We explore the effects of using two different labeling schemes and variable numbers of atlases on labeling accuracy and on the numbers of labels assigned per voxel.

**Methods**

**Image acquisition**

We used two sets of T1-weighted MRI data from a total of 20 young, healthy adult subjects. The first group of 10 subjects was scanned at the MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging using a 3T Siemens scanner and standard head coil (TE: 2.9 ms, TR: 6.6 ms, flip angle: 8°). The in-plane resolution was approximately 1 x 1 mm, the slice thickness was 1.33 mm, and the dimensions and field of view were 256 x 256 voxels. These subjects consist of four men and six women between the ages of 22 and 29 years old ($\mu = 25.3$). All are right-handed. The data were bias-corrected, affine-regis-

The second group of 10 subjects was scanned at Columbia University on a 1.5T GE scanner (TE: 5 ms, TR: 34 ms, flip angle: 45°). Slice thickness was 1.5-mm axial, in-plane resolution was 0.86 mm. Images were resliced coronally to a slice thickness of 3 mm, rotated into cardinal orientation, then segmented and parcellated using Cardviews software from MGH. These subjects consist of five men and five women between the ages of 26 and 41 years old ($\mu = 32.7$).

**Image processing before applying Mindboggle algorithm**

Mindboggle calls on third-party software to perform three preliminary steps on a subject brain image: (1) cropping non-brain matter, (2) linear coregistration with the MNI152 template [64], and (3) segmentation into gray matter, white matter, and cerebrospinal fluid. For this study, these steps were performed by (1) BET [90], (2) FLIRT [91] set to correlation ratios, 12-parameter affine transforms and trilinear interpolation, and (3) SPM2 [27] for the first group of 10 brains and FAST [92] for the second group of 10 brains.

**Mindboggle algorithm**

Mindboggle is a freely downloadable, open source software package written in Matlab (version 6, release 13, with the Image Processing Toolbox, The Mathworks Inc., USA) and has been tested on different models of desktop and laptop computers running different distributions of Linux, as well as MacOSX and Windows. The general system requirements are the basic requirements of the Matlab environment. The system used to conduct the following tests consists of a 2.2 GHz Pentium IV processor running Redhat Linux 9.0 on a PC with 1 GB memory. Mindboggle was selected as the nonlinear method because it was created by one of the authors (AK) and performed favorably in comparisons with the popular nonlinear methods AIR, ANIMAL, and SPM2 [63].

Mindboggle's general strategy is to fill a subject's cortical gray matter mask with atlas labels, based on correspondences found between structures in a subject image and in one or more atlases (see Figure 1). Details of the original algorithm may be read in [63], and consist of the following five steps performed on a subject's brain image data:

1. extract cerebral cortical sulci,
2. prepare hundreds of pieces from image-processed versions of these sulci,
3. match each piece from an atlas with a combination of pieces from the subject,
(4) translate local atlas label boundaries according to the difference in position between each match, and

(5) warp the atlas label volume to the transformed boundaries and propagate these labels to fill a subject mask. Mindboggle optionally resets planar boundaries for frontal and temporal poles as well as the occipital lobes, if the atlas itself is labeled using these planar boundaries.

Mindboggle extracts cerebral cortical sulci in the following manner (see Note 1 in Appendix). First, Mindboggle crops exposed brain surface by eroding the segmented cortex three voxels deep. Mindboggle also crops subcortex and cerebellum with a mask constructed from a union of two of the Montreal Neurological Institute's atlases: the single-subject atlas [93] and the MNI152 template [64]. All registration and labeling by Mindboggle is performed in MNI152 space (resolution of 1 × 1 × 1 mm and dimensions of 181 × 217 × 181 voxels).

Sulcus pieces are constructed as follows (see Figure 2). The segmented gray matter with cerebrospinal fluid is thinned to a pixel-wide skeleton for each slice (Matlab's bwmorph.m function). All of the skeletonized slices are stacked to create a 3-D skeleton. This skeleton is split by an interhemispheric plane formed by warping a vertical plane to the medial slab of the skeleton using a modified Self-Organizing Map algorithm (see Note 2 in Appendix). The skeleton is then broken up into pieces as follows. Starting from the top slice of the skeleton, each set of connected pixels is considered a separate piece. Each pixel in the slice below is assigned membership to the nearest piece in the above slice. The latter operation is repeated from top to bottom, as well as from bottom to top, resulting in two independent sets of candidate pieces, with each pixel having two assignments, one for each set. A single set of pieces is obtained by identifying the unique set of pairs of assignments. The 3-D pieces are then fragmented using a k-means algorithm and regrouped together if they share extensive borders. This last regrouping step is conducted so that compact structures with a low surface-to-volume ratio such as a ball do not get broken up in arbitrary ways by the k-means algorithm. "Extensive borders" is defined as a ratio of border to surface voxels equal to at least one-tenth, where a border voxel has at least one other piece in its immediate neighborhood of six voxels, and a surface voxel has fewer than six occupied voxels in its neighborhood.
Finding similar pieces in an atlas helps to determine how to transform atlas label boundaries, and therefore how to distribute atlas labels in the subject brain. Matching each piece from an atlas with a combination of (up to three) similar pieces from the subject is performed by minimizing a cost function. The cost function consists of a sum of normalized quantities derived from differences in: mean position, number of voxels, number of subvolumes, and non-overlap. Differences in mean position and number of voxels are measures of the differences in location and size between the atlas and subject pieces. The number of subvolumes for a given piece is the number of $5 \times 5 \times 5$-voxel boxes dividing the image volume that contain the piece. This measure is useful for distinguishing between pieces that have different spatial distributions, such as between a tight ball and an extensive sheet. Non-overlap of two pieces, $P_1$ and $P_2$, is equal to the fraction of subvolumes of $P_1$ that do not overlap $P_2$ added to the fraction of subvolumes of $P_2$ that do not overlap $P_1$. This measure is useful for distinguishing between differently shaped pieces that may otherwise be similar according to the other three measures.

Atlas label boundaries are locally translated according to the difference in position between nearby atlas and matching subject pieces. The translation is the difference of the mean of the local boundary from the mean of the subject piece(s), plus the difference between the mean of the atlas piece from the mean of the local boundary (after scaling by the ratio of the atlas and subject piece bounding boxes).

The atlas label volume is then warped to the transformed atlas label boundaries as follows (see Note 3 in Appendix). The atlas label that was closest to each original boundary point moves to the transformed boundary point, carrying along its neighboring labels as a function of their distance from the point (according to a Gaussian distribution function). After warping, each unlabeled voxel within the segmented gray matter mask is assigned the majority label in its $5 \times 5 \times 5$-voxel neighborhood; this last step is repeated several times.

**Evaluation**

We evaluated labels assigned by Mindboggle to a brain image (in MNI152 space) by comparing them with the manual labels for that brain (linearly registered to MNI152 space). The manual labels used for evaluation were also used to construct Mindboggle atlases. They were assigned by a single human labeler to each of the 20 subject brains (before linear registration to the MNI152 space), according to one of two different parcellation schemes. The first group of 10 subjects was labeled by Jason Tourville according to a scheme that is a modified version of Cardviews (see below) and implemented in a software tool developed by Satrajit Ghosh at the Department of Cognitive and Neural Systems, Boston University [94]. The second group of 10 subjects was labeled by Olga Kambalov according to the Cardviews parcellation scheme, created at the Center for Morphometric Analysis, Massachusetts General Hospital, and implemented in Cardviews software [12]. The labeler for each group of subjects is an expert in Cardviews.
For both parcellation schemes, 74 cortical labels were selected from the original 96 labels and merged to give 36 labels (18 per hemisphere): superior, middle, and inferior frontal and temporal gyrii, frontal and temporal poles, pre- and postcentral gyrii, superior and inferior parietal lobules, occipital lobe, fusiform, lingual/parahippocampal, and orbital (frontal) gyrii, insula, and cingulate gyrus. The anatomical divisions are coarser than those of Cardviews primarily because regions divided by planes in the Cardviews approach are combined.

Figure 3 presents an isosurface representation of a single manually labeled subject brain. To determine whether increasing the number of atlases would improve the accuracy of Mindboggle labeling, we compared the manual label for each voxel of a subject image with the majority of all Mindboggle labels for that voxel, for an increasing number of atlases used to assign labels. When determining the majority label, ties were broken by random selection. Each subject is automatically labeled by a random selection of atlases for each number of atlases. For comparisons up to nine atlases, the atlases were randomly selected from within the same subject pool; for comparisons up to 19 atlases, atlases were randomly selected from either subject pool.

The primary evaluation measure we employ is percent label agreement between atlas labels and manual labels assigned to a subject's segmented gray matter mask, with each gray matter voxel having one manual and one automated (Mindboggle) label. The agreement between atlas label set $A_i$ and manual label set $M_i$ is defined as the volume of intersection divided by the volume of the manually labeled region, computed in voxels and summed over a set of multiple labeled regions each with index $i$, where $|.|$ indicates number of voxels:

$$\text{Label agreement} = \sum_i \frac{|M_i \cap A_i|}{|M_i|}$$

Our type I error, a measure of how many incorrect labels are found in a given manually labeled region, is simply equal to one minus the label agreement for that region. We define a type II error for a given manually labeled region as the number of automatically labeled voxels outside the region that have been assigned that region's label, divided by the total number of automatically labeled voxels with that label. This is equal to one minus the fraction of voxels automatically assigned a given label that lies within the corresponding region:

$$\text{Type II error} = 1 - \sum_i \frac{|M_i \cap A_i|}{|A_i|}$$

These error measures assume that the manual labels are correct, and they can range from zero to one; a value of zero is achieved when automated and manual labels perfectly overlap for each label.

Another evaluation measure we employ is percent label accord [12], the intersection between two similarly labeled regions divided by the mean volume of the two regions:
The above voxel-based measures ignore misregistration within a labeled region. Any conclusions based on them must therefore be restricted to the labeled volumes and may not be applicable to finer resolutions.

**Results and discussion**

We found there to be greater disagreement between atlases as the number of atlases increases, as one would expect. This is clearly demonstrated in Figures 4 and 5. Figure 4 displays the anatomical distribution of the number of different labels assigned by the atlases to each voxel. Figure 5 plots the total number of voxels with a given number of different labels per voxel. Both figures present their data as a function of the number of atlases. If we compare Figure 4 with Figure 3 (it is the same subject), the disagreements are clustered about anatomical boundaries, with the highest numbers of labels per voxel at the boundaries between multiple anatomical regions, as one would expect.

Figure 6A demonstrates the variability in labeling errors when different single atlases are used to label one subject. Figure 6B demonstrates the effect of the use of multiple atlases on labeling errors for the same subject. Figure 6B indicates that increasing the number of atlases reduces labeling errors in Mindboggle. Figure 7 demonstrates that the two subject populations, manually labeled with slightly different parcellation schemes, give clearly separable labeling results, and that the label agreement between

\[
\text{Percent label accord} = \frac{\sum_{i} |M_i \cap A_i|}{\sum_{i} (|M_i| + |A_i|)/2}
\]

The above voxel-based measures ignore misregistration within a labeled region. Any conclusions based on them must therefore be restricted to the labeled volumes and may not be applicable to finer resolutions.

**Results and discussion**

We found there to be greater disagreement between atlases as the number of atlases increases, as one would expect. This is clearly demonstrated in Figures 4 and 5. Figure 4 displays the anatomical distribution of the number of different labels assigned by the atlases to each voxel. Figure 5 plots the total number of voxels with a given number of different labels per voxel. Both figures present their data as a function of the number of atlases. If we compare Figure 4 with Figure 3 (it is the same subject), the disagreements are clustered about anatomical boundaries, with the highest numbers of labels per voxel at the boundaries between multiple anatomical regions, as one would expect.

Figure 6A demonstrates the variability in labeling errors when different single atlases are used to label one subject. Figure 6B demonstrates the effect of the use of multiple atlases on labeling errors for the same subject. Figure 6B indicates that increasing the number of atlases reduces labeling errors in Mindboggle. Figure 7 demonstrates that the two subject populations, manually labeled with slightly different parcellation schemes, give clearly separable labeling results, and that the label agreement between
manual labels and (voxelwise majority) Mindboggle labels remains distinct between the two subject groups even as the number of atlases increases. Each member from the first group of 10 subjects was labeled with one atlas from the same group, then two, three, up to nine atlases, with each atlas selected at random from the remaining unselected atlases. The same procedure was repeated for each member from the second group of 10 subjects. For each member of the combined subject population, one to 19 atlases are selected at random from either subject group, are in black. Table 1 contains the data used in this figure.

If we divide the voxels into groups, by the number of different labels per voxel, as in Figure 10, we may see that there is an inverse relationship between the number of different labels and the label agreement with manual labels. Therefore, the number of labels per voxel provides a rough confidence measure for the majority label assigned to each voxel.

Therefore, Mindboggle is sensitive to variance in the subject population and to the parcellation scheme used to manually label the atlases, in particular to the vertical planes that are used to define boundaries to large regions (occipital lobe, frontal and temporal poles). These planes are not positioned by the sulcus piece matching stage but by an automated identification and matching of specific anatomical landmarks. The definitions of these landmarks may be different between parcellation schemes and may not be as consistently or as accurately determined manually or automatically in one scheme versus another. Some of the differences between the results obtained by the two subject groups (see Methods: Image Acquisition) may be attributed to the broader sampling in the second group of subjects (three races versus one, unknown vs. right-handed, and much wider age range). We can expect even greater deviations from brains that are very young, very old, or inflicted with a pathological condition, something we are presently investigating.

Even with this dependence of absolute results on parcellation scheme, we may determine whether there is a relative improvement of results across all subjects as a function of the number of atlases used to obtain the voxelwise majority labels. From Table 1 and the accompanying graph in Figure 8, we may see that increasing the number of atlases asymptotically increases mean label agreement with man-
A one-way ANOVA was performed to test if the means are the same for the label agreements obtained by the different numbers of atlases. A multiple comparison test was then performed using Tukey's honestly significantly difference criterion to determine which pairs of means are significantly different. We see from Figure 9 that simply increasing the number of atlases from one to at least four results in a statistically significant increase in label agreement ($p < 10^{-6}$ for all comparisons), and further increasing the number of atlases to at least nine (or at least seven for the first set of subjects) results in a statistically significant increase in label agreement compared with using three atlases. However, the increase in label agreement from four to five or more atlases is not statistically significant for the mixed subject group.

One should not conclude based on these data that atlas databases need only contain four or five atlases to be representative. The standard deviations for our subject pool were high enough to warrant further investigation into sources of error. These sources include morphological dissimilarities between subject subpopulations, different parcellation schemes, and limitations of the Mindboggle algorithm. Interestingly, Kittler et al. [86] found that the classification performance of the voting rule applied to face and voice biometric data also peaked at four to five experts (atlases).

The majority voting rule is probably not the optimal way to decide on a voxel's label [86-89], especially if the selected atlases deviate considerably from the subject brain to be labeled. A missing or unusual structure in a subject brain represented in only a minority of the atlases would most likely result in an inappropriate label. Rather than simply weighting the contribution of each of the atlases equally, each atlas vote for each subject voxel could be weighted by a function of the matching cost for the structure containing that voxel, since Mindboggle's matching cost function is intended to determine degree of correspondence between structures across brains.

We further separated the results by labeled region, to compare label agreement and type II errors between manual and Mindboggle labels for each label. As may be seen in Table 2, different numbers of atlases provide significantly higher label agreements for specific brain regions. Caviness et al. [12] found the percent label accord between two expert human labelers, the manual inter-rater reliability, to be 80.23% ($\sigma = 8.08\%$) averaged across all 96 labels in four brains. Since we found, as did Caviness et al., a weak correlation between percent accord and region size, we should expect that a manual inter-rater reliability for our

| A     | All subjects | $\sigma$ | II | Group I | $\sigma$ | II | Group 2 | $\sigma$ | II |
|-------|--------------|---------|----|---------|---------|----|---------|---------|----|
| 1     | 74.39        | (3.90)  | 0.20| 79.05   | (1.06)  | 0.17| 73.55   | (2.43)  | 0.21|
| 2     | 73.75        | (2.81)  | 0.21| 78.38   | (1.32)  | 0.18| 72.68   | (2.07)  | 0.21|
| 3     | 76.54        | (2.93)  | 0.19| 80.19   | (1.42)  | 0.17| 75.30   | (2.00)  | 0.20|
| 4     | 77.73        | (2.78)  | 0.18| 80.95   | (1.43)  | 0.16| 76.31   | (1.40)  | 0.19|
| 5     | 77.88        | (2.67)  | 0.18| 81.98   | (1.39)  | 0.15| 77.24   | (1.83)  | 0.19|
| 6     | 78.30        | (2.52)  | 0.18| 81.58   | (1.42)  | 0.16| 77.45   | (1.66)  | 0.18|
| 7     | 78.56        | (2.75)  | 0.18| 82.09   | (1.37)  | 0.15| 77.76   | (1.78)  | 0.18|
| 8     | 78.67        | (2.86)  | 0.18| 82.21   | (1.28)  | 0.15| 77.87   | (1.93)  | 0.18|
| 9     | 79.18        | (2.59)  | 0.17| 82.50   | (1.20)  | 0.15| 78.15   | (1.94)  | 0.18|
| 10    | 78.91        | (2.60)  | 0.17|         |         |     |         |         |     |
| 11    | 79.25        | (2.75)  | 0.17|         |         |     |         |         |     |
| 12    | 79.47        | (2.78)  | 0.17|         |         |     |         |         |     |
| 13    | 79.47        | (2.86)  | 0.17|         |         |     |         |         |     |
| 14    | 79.41        | (2.84)  | 0.17|         |         |     |         |         |     |
| 15    | 79.41        | (2.83)  | 0.17|         |         |     |         |         |     |
| 16    | 79.48        | (2.57)  | 0.17|         |         |     |         |         |     |
| 17    | 79.59        | (2.74)  | 0.17|         |         |     |         |         |     |
| 18    | 79.61        | (2.76)  | 0.17|         |         |     |         |         |     |
| 19    | 79.62        | (2.74)  | 0.17|         |         |     |         |         |     |

Percent label agreements and type II errors (II) are given for each number of atlases (A) used to label subject group 1, group 2, and all subjects. The accompanying graph is in Figure 6. Standard deviations are in parentheses (for type II errors, $\sigma$ equals 0.02 for all subjects and 0.01 for groups 1 and 2).
parcellation's fewer and larger regions to be somewhat higher than 80%. The problem with making a direct comparison between the same number and sizes of parcellation units is that Mindboggle relies solely on structural features to define anatomical boundaries whereas the Caviness approach also uses planes that extend far from the structural features used to construct the planes. We are presently evaluating Mindboggle on the entire set of 96 labels. The percent label accords obtained by Mindboggle in this study range in value across the different labeled regions, and average to 79.86% ($a = 4.18\%$) for subject group 1 and 76.23% ($\sigma = 5.17\%$) for subject group 2 (9 atlases for each subject), with the highest accords (> 90%) for the largest regions, the frontal and temporal poles and occipital lobes, and the lowest accords (< 70%) for the postcentral gyri. The fact that the Mindboggle vs. manual accuracy is comparable to the reported inter-rater reliability is very encouraging.

For a single atlas to label a single subject, Mindboggle presently takes less than 17 minutes after linear registration and gray matter segmentation on a 2.2 GHz Pentium IV processor running Redhat Linux 9.0 on a PC with 1 GB memory: 1.3 minutes to construct a sulcus skeleton, 2.5 minutes to divide the skeleton with an interhemispheric plane, 3.4 minutes to construct and tally data on sulcus pieces, 2.5 minutes to find matching pieces in the atlas and to transform them from the atlas to the subject brain, and the remaining 7 minutes to warp and propagate labels through the gray matter mask. For each additional atlas, matching, warping, and labeling takes under 10 minutes if performed sequentially. For example, labeling a subject using five atlases would take 17 minutes if conducted in parallel, or an hour if conducted sequentially. The run time would reduce significantly not only by running Mindboggle for each atlas in parallel, but also by implementing faster preprocessing algorithms and optimized code rewritten in a lower-level language such as C as opposed to Matlab.

We conclude that by using multiple atlases, the overall label agreement between manual labels and the majority labels assigned by these atlases significantly improves when using a nonlinear procedure such as Mindboggle.

We are now in the process of applying this multiple atlas extension of Mindboggle to anatomically label functional activity data. Combining a confidence measure for anatomical boundaries derived from multiple atlases with

---

**Table 2: Regions whose label agreement improves with multiple atlases**

| Labels                | L | R |
|-----------------------|---|---|
| frontal pole          | 11|   |
| sup. frontal          | 13|  8|
| mid. frontal          | 11|   |
| inf. frontal          | 16|  6|
| orbital               | 13|  8|
| precentral            | 11|   |
| postcentral           | 11|   |
| sup. parietal         |  8|  9|
| inf. parietal         |  8| 11|
| temporal pole         |  4|   |
| sup. temporal         |  6| 11|
| mid. temporal         | 17| 12|
| inf. temporal         | 17| 11|
| fusiform              |  7| 11|
| lingual               |   |   |
| occipital lobe        |  9|   |
| cingulate             |  9|   |
| insula                |  9|  3|

The effect of multiple atlases on label agreement varies by manually labeled region. Here we compare label agreements obtained by Mindboggle for each region by different numbers of atlases randomly selected from either of the two subject groups (up to 19 atlases for each of the 20 subjects). Numerical entries denote the minimum number of atlases that result in significantly higher label agreements than for single atlas data, by region ($p << 0.0001$ for all regions except inferior temporal gyms: $p < 0.01$ and right temporal pole: $p < 0.0002$). For example, an entry of “9” means that significantly higher results were obtained using 9, 10, 11, ... to 19 atlases versus using one atlas. The significance test is the same as that applied to the whole brain label agreement data.
statistical maps of functional activity data across subjects should help to establish our level of confidence in reported functional findings.

Appendix

Note 1
Since this study was conducted, Mindboggle no longer crops any part of the brain's gray matter; sulci are instead extracted by creating a mask by morphologically closing white matter (using Matlab's imclose.m function). This was visually determined to result in better sulcus extraction.

Note 2
Mindboggle now splits the skeleton much more quickly and accurately with a surface constructed by selecting a medial slab of the skeleton, flattening the slab into a mean surface along the x-axis, and applying a median filter to the surface x values (Matlab's medfilt2.m).

Note 3
Mindboggle no longer warps the atlas label volume or fills unlabeled regions by the majority label in its neighborhood. Instead, it simply fills the transformed boundaries with nearby labels according to a distance function (Matlab's bwdist.m).

Competing interests
The author(s) declare that they have no competing interests.

Authors’ contributions
AK invented Mindboggle, conceived of and executed the study, and drafted the manuscript. BM, SG, and JT contributed manually labeled brain images. JH sponsored and supported the research. All authors read and approved the final manuscript.

Acknowledgements
Arno Klein would like to thank Jack Grinband for helpful discussions. Satrajit Ghosh and Jason Tourville were supported by NIH grant R01 DC02852 (Frank Guenther, PI).

References
1. Jouandet ML, Tramo MJ, Herron DM, Hermann A, Loftus WC, Bazell J, Gazzaniga MS: Brainprints: computer-generated two-dimensional maps of the human cerebral cortex in vivo. Journal of Cognitive Neuroscience 1989, 1:88-117.
2. Ono M, Kubik S, Abernathy CD: Atlas of the cerebral sulci Stuttgart: Georg Thieme Verlag; 1990.
3. Steinmetz H, Seitz R: Functional anatomy of language processing: neuroimaging and the problem of individual variability. Neuropsychologia 1991, 29:1149-1161.
4. Rademacher J, Caviness VS Jr, Steinmetz H, Galaburda AM: Topographic variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. Cerebral Cortex 1993, 3:313-329.
5. Steinmetz H, Herzog A, Huang Y, Hacklander T: Discordant brain-surface anatomy in monozygotic twins [correspondence]. New England Journal of Medicine 1994, 331:952-953.
6. Zilles K, Shleicher A, Langemann C, Amunts K, Morosan P, Palomo-Gallagher N, Schormann T, Mohnberg H, Bürgel U, Steinmetz H, Schlaug G, Roland PE: Quantitative analysis of sulci in the human cerebral cortex: development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. Human Brain Mapping 1997, 5:218-221.
7. Le Goualher G, Argenti AM, Duyme M, Bärré WFC, Hulshoff Pol HE, Boomsma DI, Zouaoui A, Bariliot C, Evans A: Statistical sulcal shape comparisons: application to the detection of genetic encoding of the central sulcus shape. Neuroimage 2000, 11:564-574.
8. Thompson PM, Woods RP, Mega MS, Toga AW: Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. Human Brain Mapping 2000, 9:81-92.
9. Zilles K, Kawashima R, Dabringhaus A, Fukuda H, Schormann T: Hemispheric shape of European and Japanese brains: 3-D MRI analysis of intersubject variability, ethnic, and gender differences. Neuroimage 2001, 13:262-271.
10. Wright IC, Sham P, Murray RM, Weinberger DR, Bullmore ET: Genetic contributions to regional variability in human brain structure: methods and preliminary results. Neuroimage 2002, 17:256-271.
11. Rohlfing T, Brandt R, Menzel R, Maurer CR Jr: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 2004, 21:1428-1442.
12. Caviness VS Jr, Meyer J, Makris N, Kennedy DN: MRI-based topographical parcellation of human neocortex: an anatomically specified method with estimate of reliability. Journal of Cognitive Neuroscience 1996, 8:366-387.
13. Feix JA, Damasio H, Grabowski TJ: Lesion segmentation and manual warping to a reference brain: intra- and interobserver reliability. Human Brain Mapping 2000, 9:192-211.
14. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT: Automated Talairach Atlas labels for functional brain mapping. Human Brain Mapping 2000, 10:120-131.
15. Towle VL, Khorasani L, Uftring S, Pelizzi C, Erickson RK, Spire JP, Hoffmann K, Chu D, Scherg M: Noninvasive identification of human central sulcus: a comparison of gyral morphology, functional MRI, dipole localization, and direct cortical mapping. Neuroimage 2003, 19:684-697.
16. Maintz JBA: An overview of medical image registration methods. Symposium of the Belgian hospital physicists association (SBPH/ BVZF) 1996, 12:1-22.
17. Maintz JBA, Viergever MA: A survey of medical image registration. Medical Image Analysis 1998, 2:1-36.
18. Toga EAW: Brain Warping San Diego: Academic Press; 1999.
19. Broit C: Optimal registration of deformed images. In PhD thesis Department of Computer and Information Science, University of Pennsylvania, Philadelphia; 1981.
20. Bajcsy R, Kovacic S: Multiresolution elastic matching. Computer Vision Graphics Image Processing 1989, 46:1-21.
21. Bookstein FL: Principal warps: thin plate splines and their decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 1989, 11:567-585.
22. Gee JC, Reivich M, Bajcsy R: Elastically deforming 3D atlas to match anatomical brain images. Journal of Computer Assisted Tomography 1993, 17:225-236.
23. Miller MI, Christensen GE, Amit Y, Grenander U: Mathematical textbook of deformable neuroanatomies. Proceedings of the National Academy of Sciences 1993, 90:11944-11948.
24. Christensen GE, Labott RD, Miller MI: 3D brain mapping using a deformable neuroanatomy. Phys Med Biol 1994, 39:609-618.
25. Collins DL, Holmes CJ, Peters TM, Evans AC: Automatic 3-D model-based neuroanatomical segmentation. Human Brain Mapping 1995, 3:190-208.
26. Declerck J, Subsol G, Thirion JP, Atye N: Automatic retrieval of anatomical structures in 3D medical images. In Proc 1st Computer Vision, Virtual Reality and Robotics in Medicine, Nice, France Edited by: Atye N, Springer-Verlag, Berlin; 1995:153-162.
27. Friston KJ, Ashburner J, Poline JB, Frith CD, Heath J, Frackowiak RSJ: Spatial Registration and Normalization of Images. Human Brain Mapping 1995, 2:165-189.
28. Gee JC, LeBriquer L, Barillot C, Haynor DR, Bajcsy R: Bayesian approach to the brain image matching problem. Medical Imaging 1995: Image Processing, Proc. SPIE, Bellingham, WA 1995, 2434:145-156.

29. Davatzikos C: Spatial normalization of 3D brain images using deformable models. Journal of Computer Assisted Tomography 1996, 20:656-665.

30. Collin PM, Toga AW: A surface-based technique for warping three-dimensional images of the brain. IEEE Transactions on Medical Imaging 1996, 15:402-417.

31. Christensen GE, Rabbits RD, Miller MI: Deformable templates using large deformation kinematics. IEEE Transactions on Image Processing 1996, 5:1435-1447.

32. Ashburner J, Neelin P, Collins DL, Evans A, Friston K: Incorporating prior knowledge into image registration. NeuroImage 1997, 6:344-352.

33. Sandor S, Leahy R: Surface-based labeling of cortical anatomy using deformable atlas. IEEE Transactions on Medical Imaging 1997, 16:41-54.

34. Davatzikos C: Mapping image data to stereotaxic spaces: applications to brain mapping. Human Brain Mapping 1998, 6:334-338.

35. Gaunts T, Maes F, Vandermeulen D, Suetens P: Non-rigid multimodal image registration using mutual information. In Proc. 1st Medical Image Computing and Computer-Assisted Intervention, Cambridge, MA, USA, Lecture Notes in Computer Science Volume 1496. Edited by: Wells WM, Colchester A, Delp S. Springer-Verlag, Heidelberg; 1999:1099-1108.

36. Le Goualher G, Collins L, Barillot C, Evans A: Multisubject non-rigid registration of brain MRI using intensity and geometric features. In Proc. 4th Medical Image Computing and Computer-Assisted Intervention, Utrecht, Netherlands, Lecture Notes in Computer Science Volume 2208. Edited by: Niesen WJ, Viergever MA. Springer-Verlag, Heidelberg; 2001:734-742.

37. Lohmann G, Yves von Cramon D: Anatomical topography of the human temporal lobe. Human Brain Mapping 1999, 11:11-22.

38. Rivière D, Mangin JF, Papadopoulos Orfanos D, Régis J, Ayache N: Multisubject non-rigid registration of brain MRI using intensity and geometric features. In 4th Medical Image Computing and Computer-Assisted Intervention, Utrecht, Netherlands, Lecture Notes in Computer Science Volume 2208. Edited by: Niesen WJ, Viergever MA. Springer-Verlag, Heidelberg; 2001:300-314.

39. Woods RP, Grafton ST, Watson JDG, Sicotte NL, Mazziotta JC: Improved cortical structure segmentation. Human Brain Mapping 1998, 6:339-347.

40. Woods RP, Grafton ST, Watson JDG, Sicotte NL, Mazziotta JC: Automated image registration: II. Intersubject validation of linear and nonlinear models. Journal of Computer Assisted Tomography 1998, 22:153-165.

41. Woods RP, Grafton ST, Watson JDG, Sicotte NL, Mazziotta JC: Automated image registration: II. Intersubject validation of linear and nonlinear models. Journal of Computer Assisted Tomography 1998, 22:153-165.

42. Jaume S, Macq B, Warfield SK: Labeling the brain surface using a deformable multiresolution mesh. In Proc. 5th Medical Image Computing and Computer-Assisted Intervention, Tokyo, Japan, Lecture Notes in Computer Science Volume 2489. Edited by: Doi T, Kikinis R. Springer-Verlag, Heidelberg; 2002:541-548.

43. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klavensen S, Montillo A, Makris N, Rosen B, Dale AM: Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuroon 2002, 15:341-355.

44. Meier D, Fisher E: Parameter space warping: shape-based correspondence between morphologically different objects. IEEE Transactions on Medical Imaging 2002, 21:31-47.

45. Pettman ME, Han X, Xu C, Prince JL: Automated sulcal segmentation using watersheds on the cortical surface. Neuroimage 2002, 15:329-344.

46. Shera D, Davatzikos C: HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration. Transactions on Medical Imaging 2002, 21:1421-1439.

47. Cachia A, Maggini F, Rivière D, Papadopoulos-Orfanos D, Kherif F, Bloch I, Régis J: A generic framework for parcellation of the cerebral surface into gyri using geodesic Voronoi diagrams. Medical Image Analysis 2003, 7:403-416.

48. D'Agostino E, Petsas F, Vandermeulen D, Suetens P: An information theoretic approach for non-rigid image registration using voxel class probabilities. In Biomedical Image Registration (Second International Workshop), Philadelphia, PA, USA, Lecture Notes in Computer Science Volume 2717. Springer-Verlag, Heidelberg; 2003:122-131.

49. D'Agostino E, Madersbacher J, Petsas F, Vandermeulen D, Fischer B, Suetens P: Free-form registration using mutual information and curvature regularization. In Biomedical Image Registration (Second International Workshop), Philadelphia, PA, USA, Lecture Notes in Computer Science Volume 2717. Springer-Verlag, Heidelberg; 2003:11-20.

50. Liu T, Shera D, Davatzikos C: Deformable registration of cortical structures via hybrid volumetric and surface warping. In Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science Volume 2879. Edited by: Goos G, Hartmans J, van Leeuwen J. Springer-Verlag, Heidelberg; 2003:780-787.

51. Klein A, Hirsch J: Mindboggle: a scatterbrained approach to automatic brain labeling. Neuroimage 2005, 24:261-280.

52. Evans AC, Collins DL, Miller B: An MRI-based stereotactic brain atlas from 300 young normal subjects. In Proc. 22nd Symposium of the Society for Neuroscience, Anaheim 1992, 408.
65. Andreasen NC, Arndt S, Swayze V 2nd, Cizadlo T, Flbaum M, O'Leary DS, Ehrhardt JC, Yuh WTC: Thalamic abnormalities in schizophrenic humans: observed through magnetic resonance image averaging. Science 1994, 264:294-298.

66. Collins DL, Neelin P, Peters TM, Evans AC: Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. Journal of Computer Assisted Tomography 1994, 18(3):202-205.

67. Bookstein FL: Thin-Plate Splines and the Atlas Problem for Biomedical Images. In Proc. 12th Information Processing in Medical Imaging, Wye, UK, Lecture Notes in Computer Science Volume 511. Edited by: Colchester ACF, Hawkes DJ. Springer-Verlag, Heidelberg; 1991:326-342.

68. Roland PE, Graufelds CJ, Wåhlin J, Ingelman L, Andersson M, Ledberg A, Pedersen J, ÅAkerman S, Dabringhaus A, Zilles K: Human Brain Atlas: for high-resolution functional and anatomical mapping. Human Brain Mapping 1994, 1:173-184.

69. Fischl B, Sereno MI, Tootell RBH: A high-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapping 1999, 8:272-284.

70. Hammers A, Allon R, Koepp MJ, Free SL, Myers R, Lemieux L, Mitchell TN, Brooks DJ, Duncan JS: Three-dimensional maximum probabilistic atlas of the human brain, with particular reference to the temporal lobe. Human Brain Mapping 2003, 19:224-247.

71. Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J: A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). Neuroimage 1995, 2:89-101.

72. Bookstein FL: Visualizing group differences in outline shape: methods from biometrics of landmark points. In Proc. 4th Visualization in Biomedical Computing, Hamburg, Germany, Lecture Notes in Computer Science Volume 1131. Edited by: Holzne KH, Kikinis R. Springer-Verlag, Heidelberg; 1996:405-410.

73. Thompson PM, Schwartz C, Toga AW: High-resolution random mesh algorithms for creating a probabilistic 3D surface atlas of the human brain. Neuroimage 1996, 21:297-314.

74. Thompson PM, Toga AW: Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations. Neuroimage 1996, 2:291-214.

75. Le Bihan D, Breton J, Lefévre J-C. Design of a statistical model of brain shape. In Proc. 1st Information Processing in Medical Imaging, Vermont, USA, Lecture Notes in Computer Science Volume 1230. Edited by: Duncan JS, Gindi G. Springer-Verlag, Heidelberg; 1997:477-482.

76. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paats T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, Mori D, Iacoboni M, Schormann T, Amunts K, Palomo-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B: A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 2001, 356:1293-1322.

77. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paats T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, Mori D, Iacoboni M, Schormann T, Amunts K, Palomo-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Feidler J, Smith K, Boomsma D, Hulshof Pol H, Cannon T, Kawashima R, Mazoyer B: A four-dimensional probabilistic atlas of the human brain through mass tissue matching. Journal of the American Medical Informatics Association 2001, 8:401-430.

78. Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH: An integrated software suite for surface-based analyses of cerebral cortex. Journal of the American Medical Informatics Association 2001, 8:443-459.

79. Hill DLG, Hajnal JV, Rueckert D, Smith SM, Hawkes T, McLeish K: A dynamic brain atlas. In Proc. 5th Medical Image Computing and Computer-Assisted Intervention, Tokyo, Japan, Lecture Notes in Computer Science Volume 2488. Edited by: Dohi T, Kikinis R. Springer-Verlag, Heidelberg; 2002:532-539.

80. Abbott A: A new atlas of the brain. Nature (News Feature) 2003, 424:249-250.

81. Koivu A, Lee M, Kim JS, Lee JS, Kim J, Lee DS, Kwon JS, Kim SI: A probabilistic and statistical approach for Korean standard brain atlas [abstract]. 9th Annual Meeting of the Organization for Human Brain Mapping, New York City 2003.

82. Evans AC, Collins DL, Neelin P, MacDonald D, Kamber M, Marrett TS: Three-dimensional correlative imaging: applications in human brain mapping. In Functional Neuroimaging: Technical Foundations. Edited by: Thatcher RW, Hallett M, Zeffiro T, John ER, Huerta M San Diego: Academic Press; 1994:145-162.

83. Dinov ID, Mega MS, Thompson PM, Lee L, Woods RP, Holmes CJ, Summers DW, Toga AW: Analyzing functional brain images in a probabilistic atlas: a validation of subvolume thresholding. Journal of Computer Assisted Tomography 2000, 24:128-138.

84. Klein A: Mindboggle. [http://www.binarybottle.com/mindboggle.html].

85. Protect your work. BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime.” Sir Paul Nurse, Cancer Research UK

Your research papers will be:
- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2342/5/7/prepub
