Mindboggling morphometry of human brains

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Klein, Arno; Ghosh, Satrajit S.; Bao, Forrest S.; Giard, Joachim; H?me, Yrj?: Stavsky, Eliezer; Lee, Noah et al. “Mindboggling Morphometry of Human Brains.” Edited by Dina Schneidman. PLOS Computational Biology 13, no. 2 (February 2017): e1005350 © 2017 Klein et al</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1371/journal.pcbi.1005350">http://dx.doi.org/10.1371/journal.pcbi.1005350</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Public Library of Science</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Tue Mar 12 03:49:32 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/109947">http://hdl.handle.net/1721.1/109947</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution 4.0 International License</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a></td>
</tr>
</tbody>
</table>
RESEARCH ARTICLE

Mindboggling morphometry of human brains

Arno Klein1*, Satrajit S. Ghosh2,3, Forrest S. Bao4, Joachim Giard5ed, Yrjö Hämé6eb, Eliezer Stavsky6ed, Noah Lee6ed, Brian Rossa7ee, Martin Reuter8, Elias Chaibub Neto9, Anisha Keshavan10

1 Child Mind Institute, New York, New York, United States of America, 2 McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 3 Department of Otolaryngology, Harvard Medical School, Boston, Massachusetts, United States of America, 4 Department of Electrical and Computer Engineering, University of Akron, Akron, Ohio, United States of America, 5 University of Louvain, Louvain, Belgium, 6 Columbia University, New York, New York, United States of America, 7 TankThink Labs, Boston, Massachusetts, United States of America, 8 Harvard Medical School, Cambridge, Massachusetts, United States of America, 9 Sage Bionetworks, Seattle, Washington, United States of America, 10 University of California San Francisco, San Francisco, California, United States of America

¤ Current Address: Axinesis, Wavre, Belgium
© Current Address: Philips, Vantaa, Finland
¢ Current Address: Blend Labs, Inc., San Francisco, California, United States of America
© Current Address: 1010data, New York, New York, United States of America
¬ Current Address: FØCAL, Cambridge, Massachusetts, United States of America
* arno@mindboggle.info

Abstract

Mindboggle (http://mindboggle.info) is an open source brain morphometry platform that takes in preprocessed T1-weighted MRI data and outputs volume, surface, and tabular data containing label, feature, and shape information for further analysis. In this article, we document the software and demonstrate its use in studies of shape variation in healthy and diseased humans. The number of different shape measures and the size of the populations make this the largest and most detailed shape analysis of human brains ever conducted. Brain image morphometry shows great potential for providing much-needed biological markers for diagnosing, tracking, and predicting progression of mental health disorders. Very few software algorithms provide more than measures of volume and cortical thickness, while more subtle shape measures may provide more sensitive and specific biomarkers. Mindboggle computes a variety of (primarily surface-based) shapes: area, volume, thickness, curvature, depth, Laplace-Beltrami spectra, Zernike moments, etc. We evaluate Mindboggle’s algorithms using the largest set of manually labeled, publicly available brain images in the world and compare them against state-of-the-art algorithms where they exist. All data, code, and results of these evaluations are publicly available.
Introduction

This article summarizes years of work on the Mindboggle project (http://mindboggle.info), including development and application of software that automates the extraction, identification, and shape analysis of features from human brain magnetic resonance imaging (MRI) data. The principal original contributions of the Mindboggle software include (1) a hybrid approach to combine different software packages’ gray/white matter segmentations, (2) new algorithms for volume and surface shape measures devoted to brain images, including travel depth and cortical thickness, and (3) new shape-based feature extraction algorithms for brain structures such as folds, sulci, and fundi. Further contributions described in this article include (1) evaluations of Mindboggle volume and surface shape measurement algorithms against other software algorithms, (2) evaluation of Mindboggle’s fundus extraction algorithm against other software algorithms, (3) Python implementations of algorithms for general-purpose shape measures such as Laplace-Beltrami spectra and Zernike moments, and (4) application of Mindboggle to provide the most detailed shape measures computed on human brain image data. This Introduction provides background and motivation for the project as well as a history of the project, the Design and Implementation section outlines the software’s processing steps, the Results section describes evaluations and applications of the software, and the Availability and Future Directions section provides commentary and future directions.

The promise of brain imaging for finding biological markers of mental illness

Brain images have been used to derive biological markers of mental illness and disease for years, most notably to predict prognoses among patients with behavioral disorders, often more accurately than current behavioral instruments such as widely used scales and structured interviews. For example, brain images have been used to predict relapse in methamphetamine dependence [1], onset of psychosis in at-risk individuals [2,3], recovery from depression eight months later [4], response to drug treatment for depression [5,6], anxiety [7], and for cognitive behavioral therapy in schizophrenia [8] and social anxiety disorder [9,10] (see [11] for a more extensive review). Despite the above promising experimental results, there is still a dearth of reliable biomarkers [12]. The importance of identifying new biomarkers is reflected in the National Institute of Mental Health’s Strategic Objectives: “Currently, very few biomarkers have been identified for mental disorders due in part to their complexity and an incomplete understanding of the neurobiological basis of mental disorders. . .”

Variation in human brains and the “correspondence problem”

A significant impediment to our understanding of mental health is variation in human brain anatomy, physiology, function, connectivity, response to treatment, and so on. The normal range of variation must first be established to determine what is outside of this range, and only then can we hope to address neuropsychiatric assessment, diagnosis, prognosis, treatment, or prevention. An effective biomarker traditionally consists of one or more measures that maximize the separability between groups while minimizing the variance within each group. Brain images provide many ways of measuring different aspects of the brain, but it is not always clear how to compare these measures over time or across individuals. Comparing brains presumes that a brain-to-brain correspondence or mapping has been solved. To do this, scientists ubiquitously co-register images to each other, either individually or in groups, commonly with the use of a standard template brain or labeled atlas. However, registration alone does not guarantee correspondence [13] and templates are often not representative of the group being
studied [14,15]. Additional factors that affect the quality of registration are often ignored. For example, we have empirically demonstrated that registration algorithms vary widely in their accuracy [16], that even the best require removal of non-brain matter to perform adequately [17,18], and conventional registration is less robust to missing regions than feature-based registration methods [19]. Despite this, many brain imaging studies co-register brains based on image similarity, assume alignment of corresponding anatomy [20], and compare the brains at the level of a small extent such as a sphere or rectilinear volume, which can be on the order of 1/100,000th the volume of the image.

Anatomical feature-based correspondence

Neuroanatomists rely instead on high-level “features” such as distinctive cortical folding patterns and relative positions of subcortical structures to consistently identify anatomical structures or label brain regions ([21,22] and communications with neuroanatomists [23]). Such morphological features may also be identified by using multimodal imaging data and classifiers trained on such data [24]. In addition to whole (gyrus and sulcus) folds, components such as sulcal pits and sulcal fundi hold promise for establishing correspondence across brains. Sulcal pits, points of maximal depth or curvature in sulci, are interesting because they may be well conserved structures formed early in development [25–27] and have been used to characterize conditions such as polymicrogyria [28]. Sulcal fundi are defined as curves that run along the depths of sulci. They form branching skeletons that simplify the complex pattern of folds of the brain, may be measured for morphometry studies, and are used to help define the boundaries between gyri [22]. Like pits, fundi are thought to characterize early stages of morphological development, and therefore may exhibit abnormalities in neurodevelopmental and heritable disorders.

Shape measures as biomarkers

To compare features across individuals we need to quantify them. One quantification method is to characterize the quantities and distributions of grayscale values within a volume, but this does not work well for features of limited extent, such as a point, line, or surface patch. Another method is to coregister a given brain or brain feature with a reference and to define similarity with the reference based on the registration itself (deformation-based morphometry). Yet another method is to directly measure shape, where shape is defined as the geometrical information that remains when location, scale and rotation are removed from an object [29]. Publicly available brain image datasets that include any shape measures usually provide only a few shape measures per anatomical region: volume (such as the Internet Brain Volume Database, http://www.nitrc.org/projects/ibvd), surface area, and/or cortical thickness. These measures are useful for studies of neurogenesis or atrophy in morphological development, degeneration, injury, and disease progression. Volume measurement is almost ubiquitous in such studies, and cortical thickness measures derived from structural MRI data have been reported to help characterize a variety of disorders [30] such as mild cognitive impairment and Alzheimer’s disease [31–33], multiple sclerosis [34], schizophrenia [35], autism spectrum disorder [36], and alcohol dependence [37], and to predict onset or progression of, for example, Alzheimer’s disease [38–44], major depressive disorder [45], and attention-deficit/hyperactivity disorder [46].

More subtle shape measures may provide more sensitive and specific biomarkers, and combining shape measures in a multivariate analysis can improve results over any single measure [47]. The lack of shape measures may be attributable to the paucity of software programs such as BrainVISA [48,49] (https://www.nitrc.org/projects/brainvisa_ext) that compute more...
nuanced measures. Sulcal width has been used to differentiate between groups with mild cognitive impairment [50] and global and local gyrification indices computed from sulci have been used to characterize schizophrenia [51] and early-onset vs. intermediate-onset bipolar disorder as well as bipolar and unipolar depression [52–54]. More abstract shape measures such as Zernike moments (see below) have been used in patient classification, such as to distinguish cases of dementia [55].

History of the Mindboggle project

The Mindboggle project (http://mindboggle.info) has generated anatomically labeled brain image data and open source software for extracting and measuring the shapes of anatomical brain structures, and is therefore well positioned to provide the shape-based biomarkers mentioned above. This section provides a historical background of the project.

2005: The initial version of the Mindboggle software (https://osf.io/gfwnr/) was written in Matlab (Mathworks, Inc., Natick, MA) as part of a doctoral dissertation [58]. It introduced a feature-driven approach to label human brain MRI data using one atlas [19] or multiple atlases [59].

2009: With generous funding from the National Institute of Mental Health, we began to write Mindboggle from scratch in Python with some surface mesh measurements programmed in C++.

2010: To ensure that the most consistent and accurate anatomical labels are assigned to brain image data, we introduced a new cortical labeling protocol with 62 labels (Fig 1) called the Desikan-Killiany-Tourville (DKT) protocol [22,23] (http://mindboggle.info/labels.html). We applied this protocol to manually edit the anatomical labels for 101 individuals (20 individuals also include CMA non-cortical labels [60]). The resulting Mindboggle-101 dataset [22,61] (http://mindboggle.info/data.html, https://osf.io/nhtur/) is still the largest publicly available set of manually edited human brain labels in the world. These brains were used to construct multiple templates [62] and atlases [63], including the joint fusion [64] volume atlas used by the Mindboggle software for volume-based segmentation and labeling, and the DKT-40 and DKT-100 surface atlases [63] used for labeling cortical surfaces by the FreeSurfer software package [65–67] (https://surfer.nmr.mgh.harvard.edu/). The DKT-100 is used as the default atlas for labeling brains in FreeSurfer (version 6). The Mindboggle-101 brains are used for evaluations and shape analyses described in the Results section.

2013: A prototype for online, interactive visualization of Mindboggle shape analysis data won a hackathon challenge at the Human Brain Mapping (HBM 2013) conference. After use of the XTK (https://github.com/xtk/X) WebGL JavaScript library [68,69], we used the threejs (http://threejs.org/) and D3 JavaScript libraries in a second (HBM 2015 [70]) and third (HBM 2016) hackathon to create the ROYGBIV online interactive brain image viewer (Fig 1; http://roygbiv.mindboggle.info), which is under active development (https://github.com/binarybottle/roygbiv).

2015: Mindboggle processed Alzheimer’s Disease Neuroimaging Initiative (ADNI; adni.loni.usc.edu; [71]) and AddNeuroMed [72] data for an international Alzheimer’s disease challenge [73] (https://www.synapse.org/Synapse:syn2290704/wiki/60828). Teams performed statistical analyses on Mindboggle shape measures to try and determine which brains had Alzheimer’s disease, mild cognitive impairment, or were healthy, and to try and estimate a cognitive measure (mini-mental state exam score). The Results section presents an analysis of some of these data.

2016: Mindboggle was launched for broader public use after making the following improvements:
Fig 1. Cortical labels displayed in the ROYGBIV interactive online brain image viewer. The anatomical labels included in the DKT cortical labeling protocol [22] used to label the Mindboggle-101 data are displayed on a left cortical surface. These two panels show the current state of our prototype for a browser-based interactive visualization of the left hemisphere of a human brain [70] and accompanying plot of some of Mindboggle’s shape measures for a selected region (http://roygbiv.mindboggle.info).

doi:10.1371/journal.pcbi.1005350.g001
Design and implementation

Mindboggle’s open source brain morphometry platform takes in preprocessed T1-weighted MRI data, and outputs volume, surface, and tabular data containing label, feature, and shape information for further analysis. Mindboggle can be run on the command line as “mindboggle” and also exists as a cross-platform Docker container for convenience and reproducibility of results [56]. The software runs on Linux and is written in Python 3 and Python-wrapped C++ code called within a Nipype pipeline framework (http://nipy.org/nipype, doi: 10.5281/zenodo.50186) to promote a modular, flexible design that captures provenance information [57]. We have tested the software most extensively with Python 3.5.1 on Ubuntu Linux 14.04. Running Mindboggle on the Docker installation on a Macbook Pro (2.6 GHz Intel Core i7 with 16 GB memory; macOS 10.12) took about 100 minutes, of which 20 minutes were spent optionally computing Laplace-Beltrami spectra and Zernike moments. When only the surface shapes of gyrus labels were computed, without Laplace-Beltrami spectra or Zernike moments, Mindboggle took less than 7 minutes. Issues and bugs are tracked on GitHub (https://github.com/nipy/mindboggle/issues) and support questions are posted on NeuroStars (https://neurostars.org/tags/mindboggle/) with the tag “mindboggle”.

The documentation is updated online (https://readthedocs.org/projects/mindboggle) and the tests are updated online (https://circleci.com/gh/nipy/mindboggle) every time a commit is made to the GitHub repository (https://github.com/nipy/mindboggle).

Mindboggle’s flexible, modular, open source pipeline facilitates the addition of functions for computing almost any shape measure in any programming language. We initialized Mindboggle with shape measures that we thought have great potential for describing the shapes of brain structures and that complement shape measures supplied by existing software packages. It is just as easy to include functions in Mindboggle for volume-based as it is for surface-based measures, but we decided to focus primarily on surface-based shape measures to complement the volume-based methods available in standard brain image analysis packages. We also want to emphasize in this work intrinsic shape measures of brain structures rather than shapes inferred by registration-based methods such as voxel-based, tensor-based, and deformation-based morphometry that rely on a reference or canonical template and are sensitive to errors in registration. We also do not consider density values to be intrinsic shape measures, as they do not describe the shape of an object, but quantify values obtained within an object, in an analogous manner as one would quantify an fMRI signal or PET ligand binding within a voxel or region of interest.

Input and output data

For running individual functions on surface meshes, the only inputs to the software are outer cortical surface meshes constructed from T1-weighted MRI data by software such as FreeSurfer, Caret [74] or BrainVISA [48], once converted to an appropriate format (see below). For
this study we used FreeSurfer v5.1-derived labels and meshes, but the recently released FreeSurfer version 6 is recommended because it uses Mindboggle's DKT-100 surface-based atlas (with the DKT31 labeling protocol) by default to generate labels on the cortical surfaces, and generates corresponding labeled cortical and non-cortical volumes (wmparc.mgz) [75]. To preprocess data for use by Mindboggle, run the following FreeSurfer command on a T1-weighted $IMAGE file (e.g., subject1.nii.gz) to output a $SUBJECT folder (e.g., subject1):
   
   ``` shell
   recon-all -all -i $IMAGE -s $SUBJECT
   ```

   The `recon-all` command performs many steps (https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all), but the ones that are most relevant include (1) segmentation of the brain image into different tissue classes (gray/white/cerebrospinal fluid), (2) reconstruction of a triangular surface mesh approximating the pial surface for each brain hemisphere, and (3) anatomical labeling of each surface and each volume.

   To refine segmentation, labeling, and volume shape analysis, Mindboggle optionally takes output from the Advanced Normalization Tools (ANTS, v2.1.0rc3 or higher recommended; http://stnava.github.io/ANTs/), which performs various image processing steps such as brain volume extraction [17,76], tissue-class segmentation [77], and registration-based labeling [16,18,76]. To generate the ANTs transforms and segmentation files used by Mindboggle, run the `antsCorticalThickness.sh` script [76] on the same $IMAGE file, set an output $PREFIX, and provide paths to the OASIS-30 Atropos template files in directory $TEMPLATE (backslash denotes a line return):

   ``` shell
   antsCorticalThickness.sh -d 3 -a IMAGE -o PREFIX \
   -e $TEMPLATE/T_template 0.nii.gz \
   -t $TEMPLATE/T_template 0_BrainCerebellum.nii.gz \
   -m $TEMPLATE/T_template 0_BrainCerebellumProbabilityMask.nii.gz \
   -f $TEMPLATE/T_template 0_BrainCerebellumExtractionMask.nii.gz \
   -p $TEMPLATE/Priors2/priors%d.nii.gz
   ```

   Links to the template and example input and output data can be found on the Mindboggle website. Output formats include NIfTI format for volume files, VTK format for surface meshes, and comma-delimited CSV format for tables. Each file contains integers that correspond to anatomical labels or features (0–24 for sulci or fundi for either hemisphere). All output data are in the original subject’s space, except for additional surfaces and mean coordinates in MNI152 space [113]. The S1 Supplement contains a directory tree with outputs from most of the optional arguments, and does not include interim results stored in a working directory or downloaded files in a cache directory.

**Mindboggle processing steps**

Mindboggle performs the following steps. We provide some details of the algorithms at the end of each step, but for full descriptions, see the relevant software documentation (http://mindboggle.info/software.html). The S2 Supplement is an automatically generated flow diagram of the processing steps:

1. Convert FreeSurfer formats to NIfTI volumes and VTK surfaces.
2. Optionally combine FreeSurfer and ANTs gray/white segmented volumes and fill with labels.
3. Compute volumetric shape measures for each labeled region.
4. Compute shape measures for every cortical surface mesh vertex.
5. Extract cortical surface features.
6. Segment cortical surface features with labels.
7. Compute shape measures for each cortical surface label or sulcus.
8. Compute statistics for each shape measure in Step 4 for collections of vertices.

Step 1: Convert FreeSurfer formats to NIfTI volumes and VTK surfaces

Mindboggle performs all of its processing in two open standard formats: NIfTI (.nii.gz; http://nifti.nimh.nih.gov/) for volume images and VTK (.vtk, Visualization Toolkit; http://www.vtk.org/) for surface meshes. ANTs output already supports NIfTI; given FreeSurfer input, the first step that Mindboggle performs is to convert FreeSurfer volume and surface formats to NIfTI and VTK for further processing. All volume images in this study have a resolution of 1x1x1 mm$^3$ per voxel (volume element). All surface-based shape measures are computed on the "pial surface" (cortical-cerebrospinal fluid boundary) by default, since it is sensitive to differences in cortical thickness.

Step 2: Optionally combine FreeSurfer and ANTs gray/white segmented volumes and fill with labels

This optional step of the pipeline will be skipped in the future when methods for tissue class segmentation of T1-weighted MR brain images into gray and white matter improve. FreeSurfer and ANTs make different kinds of mistakes while performing tissue class segmentation (Fig 2). After visual inspection of the gray/white matter boundaries in over 100 EMBARC (https://clinicaltrials.gov/ct2/show/NCT01407094) brain images processed by FreeSurfer, we found that at least 25 brains had significant overcropping of the brain, particularly in ventral regions such as lateral and medial orbitofrontal cortex and inferior temporal lobe due to poor

![Fig 2. FreeSurfer and ANTs gray/white matter segmentation.](https://example.com/fig2.png)

Left: Coronal slice of a T1-weighted brain MRI. Middle: Cross-section of FreeSurfer inner (magenta) and outer (green) cortical surfaces overlaid on top of the same slice. The red ellipse circumscribes a region where the FreeSurfer surface reconstruction failed to include gray matter on the periphery of the brain. Right: Cross-section of ANTs segmentation. The blue ellipse circumscribes a region where the ANTs segmentation failed to segment white matter within a gyrus that the FreeSurfer correctly segmented (compare with the middle panel). The purple box in the lower right highlights a region outside of the brain that the ANTs segmentation mistakenly includes as gray matter. To reconcile some of these discrepancies, Mindboggle currently includes an optional processing step that combines the segmentations from FreeSurfer and ANTs. This step essentially overlays the white matter volume enclosed by the magenta surface in the middle panel atop the gray/white segmented volume in the right panel.

doi:10.1371/journal.pcbi.1005350.g002
surface mesh reconstruction in those regions. This corroborates Klauschen’s observation that FreeSurfer underestimates gray matter and overestimates white matter [78]. We also found that ANTs tends to include more cortical gray matter than FreeSurfer, but at the expense of losing white matter that extends deep into gyral folds, and sometimes includes non-brain tissue such as transverse sinus, sigmoid sinus, superior sagittal sinus, and bony orbit. Mindboggle attempts to reconcile the differences between FreeSurfer and ANTs segmentations by combining them.

**Hybrid segmentation algorithm.** The relabel_volume function converts the (wmparc.mgz) labeled file generated by FreeSurfer and the (BrainSegmentation.nii.gz) segmented file generated by the ANTs Atropos function [77] to binary files of pseudo-white matter and gray (including deep gray) matter. The combine_2labels_in_2volumes function overlays FreeSurfer white matter atop ANTs cortical gray, by taking the union of cortex voxels from both binary files as gray matter, the union of the non-cortex voxels from the two binary files as white matter, and assigning intersecting cortex and non-cortex voxels as non-cortex. While this strategy often preserves gray matter bordering the outside of the brain, it still suffers from over-inclusion of non-brain matter, and sometimes replaces true gray matter with white matter in areas where surface reconstruction makes mistakes.

The FreeSurfer/ANTS hybrid segmentation introduces new gray-white matter boundaries, so the corresponding anatomical (gyral-sulcal) boundaries generated by FreeSurfer and ANTs need to be updated accordingly. Mindboggle uses ImageMath’s PropagateLabelsThroughMask function in ANTs to propagate both FreeSurfer and ANTs anatomical labels to fill the gray and white matter volumes independently. The FreeSurfer-labeled cerebellum voxels overwrite any intersecting cortex voxels, in case of overlap.

Step 3: Compute volumetric shape measures for each labeled region

- **volume**
- **thickness of cortical labels** *(thickinthehead)*

As mentioned in the Introduction, the most common shape measures computed for brain image data are volume and cortical thickness for a given labeled region of the brain. Volume measurements are influenced by various factors such as cortical thickness, surface area [79], and microstructural tissue properties [80]. Computing the volume per labeled region is straightforward: Mindboggle’s volume_per_brain_region function simply multiplies the volume per voxel by the number of voxels per region. In contrast, cortical thickness can be estimated using a variety of MRI processing algorithms [49,76,81–84]. Since Mindboggle accepts FreeSurfer data as input, we include FreeSurfer cortical thickness [81] estimates with Mindboggle’s shape measures. When surface reconstruction from MRI data produces favorable results (see above), FreeSurfer cortical thickness measures can be highly reliable [82,85,86]. To avoid surface reconstruction-based problems with the cortical thickness measure, we built a function called thickinthehead that computes a simple thickness measure for each cortical region from a brain image volume without relying on surface data (Fig 3). See Results for our evaluation of cortical thickness measures.

**Thickinthehead algorithm.** The thickinthehead function first saves a brain volume that has been segmented into cortex and non-cortex voxels into separate binary files, then resamples these cortex and non-cortex files from, for example, 1mm$^3$ to 0.5mm$^3$ voxel dimensions to better represent the contours of the cortex. Next it extracts outer and inner boundary voxels of the cortex by morphologically eroding the cortex by one (resampled) voxel bordering the outside of the brain and bordering the inside of the brain (non-cortex). Then it estimates
the middle cortical surface area by the average volume of the outer and inner boundary voxels of the cortex. Finally, it estimates the thickness of a labeled cortical region as the volume of the labeled region divided by the middle surface area of that region. The \texttt{thickinthehead} function calls the \texttt{ImageMath}, \texttt{Threshold}, and \texttt{ResampleImageBySpacing} functions in ANTs.

**Step 4: Compute shape measures for every cortical surface mesh vertex**

- surface area
- mean curvature
- geodesic depth
- travel depth
Aside from the convexity and thickness measures inherited from FreeSurfer, shape measures computed for each vertex of a cortical surface triangular mesh are generated by Mindboggle’s open source C++ code (using the Visualization Toolkit, VTK) developed by Joachim Giard: surface area, mean curvature, geodesic depth, and travel depth. Surface area is computed per vertex (as opposed to per face of the mesh to be consistent with all other Mindboggle shape measures) as the area of the Voronoi polygon enclosing the vertex (Fig 4). Area can be used to normalize other values computed within a given region such as a gyrus or sulcus [87].

Curvature is an obvious shape measure for a curved and folded surface like the cerebral cortex and has the potential to help make inferences about other characteristics of the brain, such as sulcus width, atrophy [88,89], structural connections [90] and differential expansion of the cortex [91]. Mindboggle computes both mean and Gaussian curvatures (Fig 5).
Depth is an important measure characterizing the highly folded surface of the human cerebral cortex. Since much of the surface is buried deep within these folds, an accurate measure of depth is useful for defining and extracting deep features, such as sulci [92,93], sulcal fundus curves [94–96], and sulcal pits [26,97–99]. Depth may also serve as an indicator of developmental stage [26].

We are aware of three predominant methods for measuring depth of points on the surface of the cerebral cortex, where depth is the distance between a given point on the brain surface to an outer reference surface of zero depth (the portions of the brain surface in contact with the outer reference surface are gyral crowns or crests). The first is Euclidean depth, the distance along a straight path from the point on the brain to the outer reference surface. A straight path has the undesirable property that it will cross through anything, which can make a highly folded surface indistinguishable from a slightly folded surface that fills the same volume. The second is geodesic depth, the shortest distance along the surface of the brain from the point to where the brain surface makes contact with the outer reference surface. Geodesic paths are very sensitive to slight or gradual changes in depth, resulting in exaggerated distances where the outer reference surface does not wrap the brain closely. Geodesic paths are also greatly affected by cavities, so distances can be exaggerated where there are irregularities, particularly in the bottoms of sulcus folds. The third measure, FreeSurfer software’s “convexity,” while not explicitly referred to as depth, is used to indicate relative depth. It is based on the displacement of surface mesh vertices after inflating the surface mesh [65]. This can result in assigning positive depth to points on the outermost surface of the brain such as on a gyral crest, however, which is not consistent with an intuitive measure of depth.

Travel depth was introduced as a hybrid depth measure for macromolecules, defined as the shortest distance that a solvent molecule would travel from the convex hull of the macromolecule without penetrating the macromolecule surface. It was first defined for surfaces but using a voxel-based algorithm [100] that uses Dijkstra’s algorithm for finding shortest paths, and was later refined to use a much faster and more accurate vertex-based computation [101]. Mindboggle’s travel depth algorithm uses the latter; it assigns a depth value to every vertex in a mesh, assigns more reasonable path distances that are less sensitive to surface irregularities and imaging artifacts than geodesic distances, and is faithful to the topology of the surface. Figure 6 shows an example of geodesic and travel depth values, and the Results section summarizes our comparison of travel depth with geodesic depth and FreeSurfer convexity measures.

Fig 6. Geodesic depth and travel depth per vertex. Mindboggle computes geodesic depth (left) and travel depth (right) for each surface mesh vertex. This medial view of the sulcus folds from the left cortical hemisphere is colored by depth, with the deepest vertices in yellow. Note that the deepest vertices according to geodesic depth reside toward the center of the insula (center fold), whereas the deepest vertices according to travel depth run along the deepest furrows of the insula, as one would expect.

doi:10.1371/journal.pcbi.1005350.g006
**Curvature algorithms.** Mindboggle’s mean and Gaussian curvatures are based on the relative direction of the normal vectors in a small neighborhood, which works best for low resolution or for local peaks, but can be sensitive to the local linear geometry of the mesh. Increasing the radius of the neighborhood mitigates this sensitivity, so a neighborhood parameter corresponding to the radius of a geodesic disk is defined in the unit of the mesh. If coordinates are in millimeters, the default setting of 2 results in an analysis of the normal vectors within a 2mm radius disk. Other options include computing both mean and Gaussian curvatures based on the local ratios between a filtered surface and the original surface area (the filtering is done using Euclidean distances, so it’s best for less accurate but fast visualization), or computing the mean curvature based on the direction of the displacement vectors during a Laplacian filtering (a good approximation based on the Laplacian, but underestimates very large, negative or positive, curvatures due to saturation).

**Travel depth algorithm.** The travel depth algorithm constructs a combination of Euclidean paths outside the cortical surface and estimated geodesic paths along the cortical surface. The principal idea of the algorithm lies in the classification of a surface into “visible” and “hidden” areas (Fig A in the S3 Supplement). A point on the surface is considered “visible” by another point if they can be connected by a straight line without intersecting the volume enclosed by the surface. In other words, there is a “line of sight” between the two points that does not run through the interior of the surface. A point is considered “hidden” from another point if it is not visible and can only be reached by a path running either along the surface or connecting points of the surface without intersecting the enclosed volume. The above implementations of travel depth use a convex hull (Fig B in the S3 Supplement), as do most measures of cortical depth such as the adaptive distance transform [102], while other algorithms do not define a zero-depth reference surface but rely instead on convergence of an algorithm, such as the depth potential map [103]. The shape of the brain is concave in places, resulting in some gyral crowns that do not touch the convex hull. For example, in Fig C of the S3 Supplement, the gyri of the medial temporal lobe are assigned positive depth, resulting in an unreasonably high depth for the folds of that region. Since the convex hull is not suitable for application to brain images, or for surfaces with global concavities, we define and construct a different reference surface that we call the wrapper surface (Fig E in the S3 Supplement). The wrapper surface has to be chosen such that the top of a gyrus has zero depth. We compute a wrapper surface as follows. We create a volume image representing the interior of the mesh, dilate this image with a probe of radius $r$, then erode it with the same probe. This operation is also known as morphological closing, and it is important to carefully set the probe radius. If the radius is too large, the wrapper surface will be similar to the convex hull, and if the radius is too small, the wrapper surface will be too close to the original surface and the travel depth will be close to zero even inside folds. We used an empirically determined radius of 5 mm. The wrapper surface mesh is an isosurface of this morphologically closed image volume, created using the marching cubes algorithm. On a brain mesh with 150,000 vertices and 300,000 triangles, the algorithm takes around 200 seconds on an ordinary computer when the wrapper surface is provided. The generation of the wrapper surface takes an additional 20 seconds for a probe radius of 5 mm.

**Step 5: Extract cortical surface features**

- folds
- fundus per fold

Mindboggle extracts hierarchical structures from cortical surfaces [104,105], including folds and fundus curves running along the depths of the folds. A fold is a group of connected, deep
vertices (left side of Fig 7). When assigned anatomical labels, folds can be broken up into sulci (right side of Fig 7).

Fold extraction algorithm. To extract folds, a depth threshold is used to segment deep vertices of the surface mesh. We have observed in the histograms of travel depth measures of cortical surfaces that there is a rapidly decreasing distribution of low depth values (corresponding to the outer surface, or gyral crowns) with a long tail of higher depth values (corresponding to the folds). Mindboggle’s \texttt{find_depth_threshold} function therefore computes a histogram of travel depth measures, smooths the histogram’s bin values, convolves to compute slopes, and finds the depth value for the first bin with zero slope. The \texttt{extract_folds} function uses this depth value, segments deep vertices, and removes extremely small folds (empirically set at 50 vertices or fewer out of a total mesh size of over 100,000 vertices).

A fundus is a branching curve that runs along the deepest and most highly curved portions of a fold (Fig 8). As mentioned above, fundi can serve as boundaries between anatomical
regions and are interesting for their relationship to morphological development and disorders. But they are too tedious, time-consuming, and difficult to be drawn in a consistent manner on the surface meshes derived from MR images. Mindboggle provides multiple functions for extracting fundi that are optionally generated from the command line.

**Fundus extraction algorithm.** Mindboggle uses its `extract_fundi` function by default, which is evaluated against other fundus extraction methods in the Results section. This function extracts one fundus from each fold by finding the deepest vertices inside the fold, finding endpoints along the edge of the fold, connecting the former to the latter with tracks that run along deep and curved paths, and running a final filtration step. A more detailed description of these four steps follows. In the first step, the deepest vertices are those with values at least two
median absolute deviations above the median (non-zero) value. If two of these deep vertices are within (a default of) 10 edges from each other, the vertex with the higher value is chosen to reduce the number of possible fundus paths as well as to reduce computation time. To find the endpoints in the second step, the find_outer_endpoints function propagates multiple tracks from seed vertices at median depth in the fold through concentric rings toward the fold’s edge, selecting maximal values within each ring, and terminating at candidate endpoints. The final endpoints are those candidates at the end of tracks that have a high median value. If two candidate endpoints are within (a default of) 10 edges from each other, the endpoint with the higher value is chosen; otherwise the resulting fundi can have spurious branching at the fold’s edge. The connect_points_erosion function connects the deepest fold vertices to the endpoints with a skeleton of 1-vertex-thick curves by erosion. It erodes by iteratively removing simple topological points and endpoints in order of lowest to highest values, where a simple topological point is a vertex that when added to or removed from an object on a surface mesh (such as a fundus curve) does not alter the object’s topology.

Step 6: Segment cortical surface features with labels

- sulci from folds
- fundus per sulcus

Since folds are defined as deep, connected areas of a surface, and since folds may be connected to each other in ways that differ across brains, there usually does not exist a one-to-one mapping between folds of one brain and those of another. To address the correspondence problem, we need to find just those portions of the folds that correspond across brains. To accomplish this, Mindboggle segments folds into sulci, which do have a one-to-one correspondence across non-pathological brains (right side of Fig 7). Mindboggle defines a sulcus as a folded portion of cortex whose opposing banks are labeled with one or more sulcus label pairs in the DKT labeling protocol. Each label pair is unique to one sulcus and represents a boundary between two adjacent gyri, and each vertex has one gyrus label.

**Sulcus and fundus extraction algorithms.** The extract_sulci function assigns vertices in a fold to a sulcus in one of two cases. In the first case, if a vertex has a label that is in only one label pair in the fold, it is assigned that label pair’s sulcus if it can be connected through vertices with one of the pair’s labels to the boundary between the two labels. In the second case, the segment_regions function propagates labels from a label boundary to vertices whose labels are in multiple label pairs in the fold. Once sulci are defined, the segment_by_region function uses sulcus labels to segment fold fundi into sulcal fundi, which, like sulci, are features with one-to-one correspondence across non-pathological brains.

Step 7: Compute shape measures for each cortical surface label or sulcus

- surface area
- Laplace-Beltrami spectrum
- Zernike moments

In addition to shape measures computed for each vertex of a surface (Step 4), Mindboggle also computes shape measures that apply to collections of vertices such as gyri and sulci (Step 6): surface area (sum of surface areas across vertices), Laplace-Beltrami spectra, and Zernike moments.
Martin Reuter established important properties of the spectrum that relates to a shape’s intrinsic geometry with his “Shape-DNA” method [106–108]. This approach is specifically valuable for non-rigid shapes, such as anatomical structures: it is insensitive to local bending, as it quantifies only non-isometric deformation, e.g., stretching. The spectrum corresponds to the frequencies of the modes of the shape, and its real-valued components, the eigenvalues, therefore describe different levels of detail (from more global low-frequency features to localized high-frequency details, Fig 9).

Laplace-Beltrami algorithm. The eigen-decomposition of the Laplace-Beltrami operator is computed via a finite element method (FEM). Mindboggle’s Python \texttt{fem_laplacian} function is based on Reuter’s Shape-DNA Matlab implementation, and their eigenvalues agree to the 16th decimal place, attributable to machine precision.

To calculate the distance between the descriptors of two shapes, Reuter describes several approaches, e.g., $L^p$-norm, Hausdorff distance and weighted distances. One of the more prominent and simple distance measures is the Euclidean distance ($L^2$ norm) of the first $N$ smallest (non-zero) eigenvalues, where $N$ is called the truncation parameter. To account for the linearly increasing magnitude of the eigenvalues (Weyl’s law), Reuter recommends dividing each value by its area and its index (done by default in Mindboggle). As an alternative, the Weighted Spectral Distance (WESD) [109] is included in Mindboggle (but not used by default). It computes the $L^p$-norm of a weighted difference between the vectors of the $N$ smallest eigenvalues. This approach forms a pseudo-metric and also avoids domination of higher components on the final distance, making it insensitive to the truncation parameter $N$ (with a decreasing influence as $N$ gets larger). Additionally, the choice of $p$ (for the $L^p$-norm) influences how sensitive the metric is to finer as opposed to coarser differences in the shape; as $p$ increases, WESD becomes less sensitive to differences at finer scales.
Moments can describe the shape of objects, images, or statistical distributions of points, and different types of moments confer different advantages [110]. Geometric moments of 3-D coordinates have been used to construct shape descriptors for human brain morphometry [111] because of desirable characteristics such as invariance to rotation, symmetry, and scale, and they can be computed for any topology. Zernike moments [112] have also been applied to human brain morphometry for classifying dementia patients [55] and confer several advantages over geometric moments. They form a set of orthogonal descriptors, where each descriptor contains independent information about the structure, allowing the original shape to be reconstructed from the moments. They have been extensively characterized for shape retrieval performance and are robust to noise. Zernike moments can also be calculated at different orders (levels of detail): low order moments represent low frequency information while high orders represent high frequency information.

**Zernike moments algorithm.** Mikhno et al. [55] implemented Pozo et al.’s [113] efficient 3-D implementation of Zernike moments in Matlab, and helped us test our Python implementation to ensure they give consistent results. The length of the descriptors exponentially increases with order, so order 20 yields 121 descriptors while order 35 yields 342, for example. Values are generally less than or equal to one, with values much greater than one indicating instability in the calculation, which could be due to the way the mesh is created or due to calculating at an order that is too high given the resolution or size of the object.

**Step 8: Compute statistics for each shape measure in Step 4 for collections of vertices**

- median
- median absolute deviation
- mean
- standard deviation
- skewness
- kurtosis
- lower quartile
- upper quartile

There can be thousands of vertices in a single feature such as a gyrus, sulcus, or fundus, so it makes sense to characterize a feature’s shape as a distribution of per-vertex shape values (Step 4) when the shape measures don’t apply to collections of vertices (Step 7). Mindboggle’s `stats_per_label` function generates tables containing both, with summary statistical measures representing the distributions of per-vertex shape values.

**Results**

Mindboggle has been and continues to be subjected to a variety of evaluations (https://osf.io/x3up7/) and applied in a variety of contexts. In this section, we compare related shape measures, evaluate fundus extraction algorithms, and evaluate the consistency of shape measures between scans. We also demonstrate Mindboggle’s utility in measuring shape differences between left and right hemispheres, and in measuring brain shape variation.
Comparisons between brain shape measures

We compared shape measures with one another in a representative individual from the Mindboggle-101 data set (Fig 10) and for the entire data set (Figs 11, 12 and 13) to emphasize to the reader that shape measures are not independent of one another and that care must be taken when comparing differently defined shape measures or when using one as a proxy for another. Fig 10 plots over 130,000 vertices of one brain hemisphere, where the coordinates are two different shape measures assigned to each vertex: geodesic depth by travel depth (top) and mean curvature by travel depth (bottom). This figure demonstrates that curvature is positively correlated with depth and that geodesic depth produces higher shape values than travel depth, and may exaggerate depth, such as in the insula (also clearly evident in Figs 6 and 11).

While it may be useful to compare the distributions of two different shape measures for each region over a population (as in Figs 11, 12 and 13), we also computed the distance correlation between related shape measures for each cortical region in the Mindboggle-101 subjects (Table 1; https://osf.io/9cn7s/). To compare related (travel and geodesic depth, mean and FreeSurfer curvature) surface shape measures, we computed the distance correlation between each pair of shape measures across all of the vertices per region, and computed the average of the distance correlations per region across the 101 subjects. Distance correlation enabled a comparison of the pattern of values for a given region between two shape measures without regard for their absolute values. Mindboggle’s travel depth and geodesic depth measures were very highly correlated for 60 of the 62 regions, with distance correlations ranging from 0.91 to 1.00 (all but four greater than 0.95). The two outliers were the left and right insula (0.29 and 0.31), which corroborates our earlier assertion that geodesic depth can exaggerate depth values compared to travel depth in regions such as the insula. Mindboggle’s mean curvature and FreeSurfer’s curvature measures had distance correlations ranging from 0.73 (insula) to 0.91 (rostral middle frontal), with the top 10 values all for frontal and parietal regions. Since thickinthehead values are computed per region, not per vertex, to compare thickinthehead with median FreeSurfer thickness values, we constructed a pair of vectors for each region with 101 values, each value corresponding to the shape measure for that region in a subject, and computed the distance correlation between the two vectors. The highest distance correlations (0.8 to 0.7) were obtained by frontal and parietal regions, and the lowest correlations (0.3 to 0.2) by precuneus, parahippocampal, fusiform, and cingulate regions. See the “Comparison between cortical thickness measures” section below for a comparison between absolute cortical thickness measures.

Comparison between travel depth and FreeSurfer’s convexity measure

As described above, travel depth uses a reference wrapper surface that lies closer to the cortical surface than a convex hull would. In particular, the wrapper lies closer to the medial temporal lobe, so the gyri in this area have depth values equal to zero as one would want. FreeSurfer’s convexity measure [81], often used to indicate relative depth, leads to non-zero and even negative values for vertices on these gyri (Fig D in the S3 Supplement). We computed the mean and standard deviation of four statistical measures of travel depth and FreeSurfer’s convexity values for over 130,000 vertices in a representative cortical surface. For this comparison, we consider a point to be close to the wrapper surface if the distance between the two is smaller than 0.1 mm, a depth value is considered small if it is less than 0.1 mm, and a convexity value is considered small if it is less than the smallest convexity value for all the vertices in the mesh. For travel depth, by definition all vertices (and only those vertices) that are close to the wrapper surface have a small depth. For convexity, almost all vertices (97.71%) that have a small convexity value are close to the wrapper surface, but they represent only 6.89% of the vertices close
**Fig 10. Relationships between brain shape measures.** In these plots, we compare a pair of shape measures for each vertex of each right cortical region in a representative individual from the Mindboggle-101 brains, colored arbitrarily by region. Top: In this plot comparing two measures of depth, geodesic depth is higher than travel depth, and may exaggerate depth, such as in the insula (gray dots extending to the upper left). Bottom: In this plot of mean curvature by travel depth, we again see that the shape measures are not independent of one another. As one might expect, we see greater curvature at greater depth.

doi:10.1371/journal.pcbi.1005350.g010
Comparison between cortical depth measures

We are aware of only one study directly comparing FreeSurfer with manual cortical thickness measures, where the manual estimates were made in nine gyral crowns of a post-mortem brain, selected for their low curvature and high probability of having been sampled.
perpendicular to the plane of section [114]. We compared thickinthehead, FreeSurfer, and ANT’s cortical thickness estimates in different populations, including the Mindboggle-101 subjects (Fig 13) and in the 40 EMBARC control subjects (https://osf.io/jwhea/). For 16 cortical regions in the 40 subjects, we measured scan-rescan reliability of cortical thickness measures, and we compared thickness measures with published estimates based on manual delineations of MR images of living brains [115]. Forty percent of FreeSurfer estimates for the 640 labels were in the published ranges of values, whereas almost ninety percent of thickinthehead’s estimates were within these ranges (as mentioned above, Klauschen observed that FreeSurfer underestimates gray matter and overestimates white matter [78]).
ANTs values deviated further from the published estimates and were less reliable (greater inter-scan and inter-subject ranges) than the FreeSurfer or thickinthehead values.

**Evaluation of fundus extraction algorithms**

This section presents the first quantitative comparison of fundus extraction software algorithms. Since there exists no ground truth for fundus curves, we must resort to other means of evaluation. We leave it to future work to determine their utility for practical applications such as diagnosis and prediction of disorders. Since the DKT labeling protocol defines many of its
anatomical label boundaries along approximations of fundus curves, we used the manually edited anatomical label boundaries in the Mindboggle-101 dataset as gold standard data to evaluate the positions of fundi extracted by four different algorithms in 2013. Specifically, for each of the 48 fundi/sulci defined by the DKT protocol, we computed the mean of the minimum Euclidean distances from the label boundary vertices in the sulcus to the fundus vertices in the sulcus, as well as from the fundus vertices in the sulcus to the label boundary vertices in the sulcus. The algorithms included Mindboggle’s default `connect_points_erosion` function described above, Forrest Bao’s pruned minimum spanning tree algorithm [104], Gang Li’s algorithm [116], and an algorithm in the BrainVISA software [96]. That last algorithm was omitted from the results because too few fundi were extracted to make an adequate

### Table 1. Distance correlations between related shape measures.

To compare pairs of related (travel and geodesic depth, mean and FreeSurfer curvature) surface shape measures, we computed the distance correlation between vectors of shape values for all vertices in each cortical region, and averaged the distance correlations across the 101 Mindboggle-101 subjects. For `thickintthehead` and FreeSurfer thickness measures, we computed the distance correlation between vectors of median shape values for all 101 Mindboggle-101 subjects for each cortical region.

<table>
<thead>
<tr>
<th>Cortical region</th>
<th>travel depth vs. geodesic depth</th>
<th>mean curvature vs. FreeSurfer curvature</th>
<th>thickintthehead vs. FreeSurfer thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>left</td>
<td>right</td>
<td>left</td>
</tr>
<tr>
<td>caudal anterior cingulate</td>
<td>0.97</td>
<td>0.96</td>
<td>0.83</td>
</tr>
<tr>
<td>caudal middle frontal</td>
<td>0.99</td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>cuneus</td>
<td>0.99</td>
<td>0.99</td>
<td>0.84</td>
</tr>
<tr>
<td>entorhinal</td>
<td>0.96</td>
<td>0.96</td>
<td>0.77</td>
</tr>
<tr>
<td>fusiform</td>
<td>0.97</td>
<td>0.97</td>
<td>0.83</td>
</tr>
<tr>
<td>inferior parietal</td>
<td>0.99</td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>inferior temporal</td>
<td>0.98</td>
<td>0.98</td>
<td>0.84</td>
</tr>
<tr>
<td>isthmus cingulate</td>
<td>0.91</td>
<td>0.93</td>
<td>0.78</td>
</tr>
<tr>
<td>lateral occipital</td>
<td>0.99</td>
<td>0.99</td>
<td>0.86</td>
</tr>
<tr>
<td>lateral orbitofrontal</td>
<td>0.92</td>
<td>0.92</td>
<td>0.80</td>
</tr>
<tr>
<td>lingual</td>
<td>0.97</td>
<td>0.98</td>
<td>0.83</td>
</tr>
<tr>
<td>medial orbitofrontal</td>
<td>0.97</td>
<td>0.97</td>
<td>0.82</td>
</tr>
<tr>
<td>middle temporal</td>
<td>0.99</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td>parahippocampal</td>
<td>0.96</td>
<td>0.97</td>
<td>0.81</td>
</tr>
<tr>
<td>paracentral</td>
<td>0.99</td>
<td>0.99</td>
<td>0.87</td>
</tr>
<tr>
<td>pars opercularis</td>
<td>0.98</td>
<td>0.98</td>
<td>0.89</td>
</tr>
<tr>
<td>pars orbitals</td>
<td>0.98</td>
<td>0.98</td>
<td>0.90</td>
</tr>
<tr>
<td>pars triangularis</td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
</tr>
<tr>
<td>pericalcarine</td>
<td>0.96</td>
<td>0.97</td>
<td>0.76</td>
</tr>
<tr>
<td>postcentral</td>
<td>1.00</td>
<td>1.00</td>
<td>0.87</td>
</tr>
<tr>
<td>posterior cingulate</td>
<td>0.99</td>
<td>0.99</td>
<td>0.84</td>
</tr>
<tr>
<td>precentral</td>
<td>0.99</td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>precuneus</td>
<td>0.98</td>
<td>0.98</td>
<td>0.86</td>
</tr>
<tr>
<td>rostral anterior cingulate</td>
<td>0.98</td>
<td>0.97</td>
<td>0.80</td>
</tr>
<tr>
<td>rostral middle frontal</td>
<td>0.99</td>
<td>0.99</td>
<td>0.91</td>
</tr>
<tr>
<td>superior frontal</td>
<td>0.99</td>
<td>0.99</td>
<td>0.89</td>
</tr>
<tr>
<td>superior parietal</td>
<td>0.99</td>
<td>0.99</td>
<td>0.89</td>
</tr>
<tr>
<td>superior temporal</td>
<td>0.99</td>
<td>0.99</td>
<td>0.86</td>
</tr>
<tr>
<td>supramarginal</td>
<td>1.00</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td>transverse temporal</td>
<td>0.97</td>
<td>0.98</td>
<td>0.76</td>
</tr>
<tr>
<td>insula</td>
<td>0.29</td>
<td>0.31</td>
<td>0.73</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pcbi.1005350.t001
comparison (BrainVISA extracts 65 sulci per hemisphere, and it is possible that the program did not define some folds as sulci that contain fundi according to the DKT labeling protocol).

All of the fundi, summary statistics, and results are available online (https://osf.io/r95wb/). While there was no clear winner, we can summarize our comparison by computing the mean distance between fundi and label boundaries across all sulci for the three methods and by tallying how many sulci had the smallest mean distance among the methods. When measured from label boundaries to fundi, Gang Li’s and Mindboggle’s fundi were closer than were Forrest Bao’s (mean distances of 2.09mm and 2.38mm vs. 3.65mm, respectively; 25 and 21 vs. 2 closest sulci), whereas when measured from fundi to label boundaries, Forrest Bao’s fundi were closer than were Mindboggle’s or Gang Li’s (mean distances of 3.33mm vs. 4.06mm and 4.65mm, respectively; 41 vs. 5 and 2 closest sulci). When measuring from either direction, the maximum distances averaged across all sulci were higher for Forrest Bao’s fundi (11.65mm and 11.61mm) than for Mindboggle’s (10.84mm and 9.75mm) or Gang Li’s (11.12mm and 6.87mm).

Consistency of shape measures between MRI scans of the same person

For a shape measure to be useful in comparative morphometry, it should be more sensitive to differences in anatomy than to differences in MRI scanning setup or artifacts. To get a sense of the degree of scan/rescan consistency of our shape measures, we ran Mindboggle on 41 Mindboggle-101 subjects with a second MRI scan (OASIS-TRT-20 and MMRR-21 cohorts). We computed the fractional shape difference per cortical region as the absolute value of the difference between the region’s shape values for the two scans divided by the first scan’s shape value. For the volumetric shape measures (volume and thickness of cortical thickness), shape value is computed by region; for the surface-based shape measures (area, travel and geodesic depth, mean and FreeSurfer curvature, and FreeSurfer thickness), shape value is assigned the median value across all vertices within a region. All shape tables, statistical summary tables, and accompanying plots are available online (https://osf.io/mhc37/).

Table A in the S4 Supplement gives the average across the 41 subjects of the fractional shape differences between MRI scans for each of the 31 left cortical regions, and for each shape measure, and Table 2 gives a statistical summary of the differences. In general, the values are low enough to suggest high inter-scan shape consistency, but we will point out values greater than or equal to 0.10. Of the volumetric shape measures (volume and thickness of cortical thickness), only one value exceeded or equaled 0.10: entorhinal volume (0.21). Entorhinal cortex had the second smallest volume of manually labeled MRI cortical regions in 101 healthy human brains [22] (after transverse temporal cortex; see https://osf.io/st7nk/), and low scan/rescan consistency for small brain structures corroborates Jovicich’s observation in 2013 [117]: “We found that the smaller structures (pallidum and amygdala) yielded the highest absolute volume reproducibility errors, approximately 3.8% (average across sites), whereas all other structures had errors in the range 1.8–2.2% (average across sites), with the longitudinal segmentation analysis. Our absolute % errors in test-retest volumetric estimates are comparable to those reported by previous studies (Kruggel et al., 2010; Morey et al., 2010; Reuter et al., 2012).”

Regarding cortical thickness measures, Jovicich observed: “The thickness reproducibility results of the various structures were largely consistent across sites and vendors, with errors in the range 0.8–5.0% for the longitudinal analysis (table 7).” Of the surface shape measures, the following exceeded or equaled 0.10 for three measures (travel depth, geodesic depth, and FreeSurfer curvature): entorhinal, medial orbitofrontal, and (caudal anterior, rostral anterior, and isthmus) cingulate regions; and for at least one of the measures: lateral orbitofrontal, parahippocampal, pericalcarine, and insular regions. The greatest differences were for FreeSurfer
curvature in the pericalcarine (0.34), insula (0.28), and rostral anterior cingulate (0.23), followed by entorhinal volume (0.21) and travel depth (0.20). FreeSurfer curvature had the greatest number of outliers (Table 2) and was the only shape measure that spanned negative to positive values, so regions with very small median curvature values could have inflated these fractions. Future evaluations will assess the impact that differences in scans have on morphometry-based clinical research.

Measuring shape differences between left and right hemispheres

To measure interhemispheric shape differences, we computed the fractional shape difference per cortical region as in the preceding section, replacing inter-scan differences with interhemispheric differences (https://osf.io/dp4zy/), and using all 101 Mindboggle-101 brains. Table B of the S4 Supplement gives the average across the 101 subjects of the fractional shape differences between hemispheres for each of the 31 cortical regions, and for each shape measure, and Table 3 gives a statistical summary of the differences. The values are much higher than the corresponding inter-scan differences in the previous section, suggesting that shape differences between hemispheres are greater than shape differences between MRI scans of the same hemisphere.

Table 2. Summary statistics of shape differences between MRI scans. This table gives a statistical summary of the shape differences between two scans of the same brain for 41 brains. The “mean” column is the average of the mean values in Table A of the S4 Supplement, while the other columns contain averages of their respective values over the 31 regions; for example, the “std” column contains the average of the standard deviations computed for each of the 31 regions. [>0.50 and >0.25 give the number of regions (out of 1,271 = 31 regions times 41 subjects) where the fractional absolute difference was above 0.50 and 0.25, respectively.]

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std</th>
<th>min</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>max</th>
<th>&gt;0.50</th>
<th>&gt;0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>volume</td>
<td>0.047</td>
<td>0.044</td>
<td>0.002</td>
<td>0.019</td>
<td>0.035</td>
<td>0.063</td>
<td>0.213</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>thickinthehead</td>
<td>0.052</td>
<td>0.046</td>
<td>0.001</td>
<td>0.017</td>
<td>0.039</td>
<td>0.078</td>
<td>0.183</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>area</td>
<td>0.039</td>
<td>0.038</td>
<td>0.001</td>
<td>0.014</td>
<td>0.030</td>
<td>0.054</td>
<td>0.182</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>travel depth</td>
<td>0.061</td>
<td>0.050</td>
<td>0.002</td>
<td>0.024</td>
<td>0.050</td>
<td>0.085</td>
<td>0.229</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>geodesic depth</td>
<td>0.059</td>
<td>0.049</td>
<td>0.002</td>
<td>0.022</td>
<td>0.048</td>
<td>0.082</td>
<td>0.222</td>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td>mean curvatures</td>
<td>0.044</td>
<td>0.038</td>
<td>0.001</td>
<td>0.016</td>
<td>0.033</td>
<td>0.059</td>
<td>0.170</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>FreeSurfer curvature</td>
<td>0.094</td>
<td>0.091</td>
<td>0.004</td>
<td>0.033</td>
<td>0.070</td>
<td>0.127</td>
<td>0.433</td>
<td>17</td>
<td>78</td>
</tr>
<tr>
<td>FreeSurfer thickness</td>
<td>0.041</td>
<td>0.036</td>
<td>0.001</td>
<td>0.014</td>
<td>0.032</td>
<td>0.060</td>
<td>0.147</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pcbi.1005350.t002

Table 3. Summary statistics of shape differences between left and right hemispheres. This table gives a statistical summary of the interhemispheric shape differences for the 101 Mindboggle-101 brains. The “mean” column is the average of the mean values in Table B of the S4 Supplement, while the other columns contain averages of their respective values over the 31 regions; for example, the “std” column contains the average of the standard deviations computed for each of the 31 regions. [>0.50 and >0.25 give the number of regions (out of 3,131 = 31 regions times 101 subjects) where the fractional absolute difference was above 0.50 and 0.25, respectively.]

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std</th>
<th>min</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>max</th>
<th>&gt;0.50</th>
<th>&gt;0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>volume</td>
<td>0.129</td>
<td>0.091</td>
<td>0.002</td>
<td>0.058</td>
<td>0.117</td>
<td>0.180</td>
<td>0.448</td>
<td>38</td>
<td>443</td>
</tr>
<tr>
<td>thickinthehead</td>
<td>0.044</td>
<td>0.036</td>
<td>0.001</td>
<td>0.017</td>
<td>0.037</td>
<td>0.064</td>
<td>0.169</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>area</td>
<td>0.183</td>
<td>0.163</td>
<td>0.002</td>
<td>0.074</td>
<td>0.148</td>
<td>0.248</td>
<td>1.025</td>
<td>165</td>
<td>744</td>
</tr>
<tr>
<td>travel depth</td>
<td>0.198</td>
<td>0.199</td>
<td>0.003</td>
<td>0.074</td>
<td>0.150</td>
<td>0.257</td>
<td>1.251</td>
<td>211</td>
<td>764</td>
</tr>
<tr>
<td>geodesic depth</td>
<td>0.173</td>
<td>0.163</td>
<td>0.002</td>
<td>0.067</td>
<td>0.133</td>
<td>0.229</td>
<td>1.009</td>
<td>148</td>
<td>658</td>
</tr>
<tr>
<td>mean curvatures</td>
<td>0.104</td>
<td>0.113</td>
<td>0.002</td>
<td>0.038</td>
<td>0.079</td>
<td>0.141</td>
<td>0.872</td>
<td>59</td>
<td>192</td>
</tr>
<tr>
<td>FreeSurfer curvature</td>
<td>0.190</td>
<td>0.435</td>
<td>0.003</td>
<td>0.074</td>
<td>0.150</td>
<td>0.250</td>
<td>3.890</td>
<td>205</td>
<td>626</td>
</tr>
<tr>
<td>FreeSurfer thickness</td>
<td>0.050</td>
<td>0.077</td>
<td>0.001</td>
<td>0.018</td>
<td>0.036</td>
<td>0.062</td>
<td>0.691</td>
<td>23</td>
<td>28</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pcbi.1005350.t003
Measuring human brain shape variation

To estimate the normal range of variation in the shapes of healthy adult human brains, we applied Mindboggle software in 2015 to compute shape measures for our Mindboggle-101 dataset. The result is the largest set of shape measures computed on healthy human brain data (See the S5 Supplement and https://osf.io/gzshf/ for detailed results) [118,119]. We are treating these as normative data against which anyone can compare similarly processed images of different healthy adult populations as well as patient populations.

The data we analyzed consist of repeated measurements on five distinct real-valued shape measures (mean curvature, geodesic depth, travel depth, FreeSurfer convexity, and FreeSurfer thickness) for each of 31 distinct regions per brain hemisphere in each of the 101 subjects. Each subject was scanned at one of five different laboratories. At the bottom of Fig 14 is one example of the many heatmap tables we have generated from these data (all results are accessible at https://osf.io/d7hx8/). Each table presents one value for each labeled region or sulcus for each of the 101 subjects. The value is either volume or thickinthethicortical thickness for volumetric images, or for one of the five surface shape measures above, one of eight summary statistical measures (mean, median, median absolute deviation, standard deviation, lower and upper quartiles, skewness, and kurtosis) computed across all vertices in the surface mesh of the labeled region or sulcus.

We organized the data in a nested fashion: brain hemisphere is nested within subject, and subject is nested within laboratory. In addition to the five shape measurements and the three nested classification factors, the data also include three covariates: sex (male, female), age (integer variable), and handedness (left, right; we relabeled two ambidextrous subjects as left-handed). Given the grouped nature of the data, we used linear mixed models for the statistical modeling of the data. To assess the importance of each of the covariates and nested classification factors, we fitted 24 distinct linear mixed models for each shape measure and brain region combination to assess the importance of each of the covariates (sex, handedness, and age as fixed effects) and nested classification factors (laboratory, subject, and brain hemisphere as random effects). For each shape measure, we decomposed the total variance into the variance between laboratories, between subjects within a laboratory, between brain hemispheres within a subject, and within brain hemispheres.

For each shape measure and brain region combination, we used the Bayesian Information Criterion (BIC) score to select the best model among the 24 competing models. A BIC score is a goodness of fit measure used to perform model selection among models with different dimensions (number of parameters), and is proportional to the negative log likelihood of the model penalized by the number of parameters in the model. It strikes a balance between model fit (measured by the log-likelihood score) and model complexity (measured by the number of parameters in the model). In the context of linear models, an over-parameterized model will always have a larger log-likelihood score than a more parsimonious model, but it will also likely overfit the data. Nonetheless, by including a penalty proportional to the number of parameters in the model, the BIC score can be used to compare models with different dimensions since over-parameterized models are penalized to a greater extent. The smaller the BIC score, the better the model fits the data.

Two models stood out as the best models for the mean curvature, travel depth, FreeSurfer convexity, and FreeSurfer thickness shape measures across the 31 brain regions (S5 Supplement). Both models include handedness and age as fixed effects. They only differ by the inclusion of the extra “subject within lab” nesting level. For all shape measures and brain regions, the bulk of the variability was concentrated in the residual, not in the hemisphere (“side”), subject, or laboratory (top of Fig 14).
**Fig 14. Brain shape variation in healthy humans.** Top: Overview of the variance results for five shape measures computed on each of 31 manually labeled cortical regions (combined across both hemispheres for this figure) in the 101 Mindboggle-101 healthy human brains. The blue color-coded heatmap shows the relative contributions of subject, hemisphere, and residual to describe the variability for each shape measure, with a greater contribution coded by a darker blue. For all shape measures and brain regions, most of the variability was concentrated in the residual. See the S5 Supplement for a description of the statistical models.
We repeated the same analysis as above on two scans acquired three years apart from hundreds of the ADNI participants (126 with Alzheimer’s, 199 healthy controls) as part of an international Alzheimer’s challenge (see “History of the Mindboggle project” section above) to see if we could find changes in brain shape measures that correlate with changes in ADNI-MEM cognitive scores over the course of three years. This resulted in the most detailed shape analysis of brains with Alzheimer’s disease ever conducted [119] (https://osf.io/d7hx8/). To identify shape measures associated with Alzheimer’s disease, we used the average of the ranks of the following tests in that study: Kolmogorov-Smirnov test to see if there was a difference between distributions at baseline and at three years, and correlation of change in shape and change in ADNI-MEM cognitive scores.

We found that healthy brains and brains with Alzheimer’s disease have similar shape statistical summaries, but changes in the following shape measures after a three-year interval were significantly correlated with changes in ADNI-MEM cognitive score:

- Volume for right caudal anterior cingulate and left: entorhinal, inferior parietal, (middle, superior) temporal, superior frontal, precuneus, and supramarginal gyri
- FreeSurfer thickness for left and right: entorhinal, fusiform, inferior parietal, (inferior, middle, superior) temporal, superior frontal, precuneus, and supramarginal gyri; left: (caudal middle/lateral, orbito/rostral middle) frontal, and pars triangularis gyri; right lingual gyrus
- Mean curvature for left and right rostral middle frontal gyri; left (middle, superior) temporal gyri; right inferior temporal gyri

Availability and future directions

In this article, we have documented the Mindboggle open source brain morphometry platform and demonstrated its use in studies of shape variation in healthy and diseased humans. There are many ways in which the open source software community can extend Mindboggle’s capabilities, and there are many possible applications for Mindboggle to brain and non-brain data. Here we will provide links to the software and data used in this study, briefly summarize the study results, and point toward possible further evaluations and alternative approaches.

Software and data used in this study

Mindboggle home: http://mindboggle.info
Mindboggle software: https://github.com/nipy/mindboggle
- Documentation: https://readthedocs.org/projects/mindboggle
- Issues and bugs: https://github.com/nipy/mindboggle/issues
- Support questions (post with tag “mindboggle”): https://neurostars.org/tags/mindboggle/
- Continuous integration tests: https://circleci.com/gh/nipy/mindboggle
- BIDS-Apps Docker app: https://github.com/BIDS-Apps/mindboggle

doi:10.1371/journal.pcbi.1005350.g014
Summary of results

In this section we summarize the findings of our evaluations in the Results section. The number of different shape measures and the size of the populations make this the largest and most detailed shape analysis of human brains ever conducted. We computed over 8,000 values corresponding to statistical summaries of shape measures and coefficients of shape measures for each of the 101 brain images in the Mindboggle-101 dataset and for each of thousands of brain images in the ADNI and AddNeuroMed datasets. Shape measures are not independent of one another, and some related shape measures exaggerate values for certain morphological structures (such as geodesic vs. travel depth for the insula). Mindboggle’s thickinthehead cortical thickness measure is consistent across scans and across brains and generated values that are closer to published ranges of values than FreeSurfer or ANTs values. Mindboggle’s travel depth measure provides a more faithful representation of depth for shallow portions than FreeSurfer’s convexity measure. Mindboggle’s fundus are comparable to Gang Li’s fundi in terms of average proximity to manual label boundaries, but there was no clear winner in our evaluation of fundus extraction algorithms. Mindboggle’s shape measures are reasonably consistent across scans of the same brain, with some exceptions (such as entorhinal volume). We found that for the shape measures and populations we studied that shape differences between hemispheres were greater than shape differences between MRI scans of the same hemisphere, and that the variability within each brain hemisphere was higher than the variability between brain hemispheres in a participant or between participants. Finally, we reported which brain regions were significantly correlated with changes in ADNI-MEM cognitive score after a three-year interval as part of an international Alzheimer’s challenge.

Further evaluations and enhancements of Mindboggle

The Mindboggle software will continue to be subjected to evaluations of its algorithms as well as of its applicability to new datasets of healthy and diseased brains. Data exist to conduct evaluations of test/retest reliability and reproducibility [120,121], with different imaging parameters [122], with genetic information [123], with heritability information [124], at higher field
strengths [125], etc. Some data also exist for evaluating features such as sulcal pits [126]. Including different types of brain images can enable multivariate analyses, independent corroboration of morphology, and can even help to better interpret the factors that influence morphology [80]. We leave for future work real-world evaluations of Laplace-Beltrami spectra and Zernike moments, as well as comparisons of shape measures generated by Mindboggle and other brain morphometry software, such as BrainVISA. We also intend to evaluate Mindboggle output by analyzing interactions among shape measures to find higher order morphological relationships with brain shape differences.

There are many ways to enhance Mindboggle’s functionality and applicability to pathological brains. Taking advantage of different and multiple types of images, atlases, labels, features, and shape measures are clear ways to expand and improve Mindboggle, and the software was built using the Nipype framework specifically to enable modular and flexible inclusion of different algorithms, and to easily generate different outputs using different input data or parameter settings. We took advantage of this flexibility to generate multiple outputs for comparison in our evaluation studies. In the future, Mindboggle could accept different preprocessed inputs to take advantage of promising new algorithms that combine surface reconstruction with whole-brain segmentation in a way that is more robust to white-matter abnormalities [127]. The current version of Mindboggle does not take advantage of probabilistic labels, features, and shape measures, and such probabilistic assignments could lead to more careful interpretations of morphometry studies.

Alternative approaches to Mindboggle: Deep learning and beyond

The Mindboggle software extracts and identifies features for shape analysis. This approach is based on human-designed features (brain structure and label definitions and algorithmic implementations) and assumes the validity of the designed feature model. The tremendous success that machine learning (especially deep learning) approaches have had across domains [128] is strong evidence that such approaches may improve automated feature extraction, identification and labeling for features that a human would never consider designing. Machine learning has recently been demonstrated to recognize the multi-modal ‘fingerprint’ of cortical areas [24]. In 2011, we advocated a novel application of convolutional networks to build discriminative features and were able to demonstrate automated volumetric labeling of the cerebral cortex, without human intervention to build handcrafted features or to provide other prior knowledge [129]. At the time we had very limited training data (40 manually labeled brains), but with the Mindboggle-101 dataset, tables of shape statistics generated by Mindboggle, and with improved deep learning architectures, we may now be in a better position to apply deep learning to this problem. It may be helpful to explore ways in which priors and invariances can be modeled and integrated into deep learning approaches to reduce the amount of required training data and to integrate human expert information. This may be particularly beneficial for pathological conditions with tumors, lesions, and edemas, etc. that do not conform to a canonical reference brain or are difficult to obtain in sufficient quantities to train a deep learning algorithm. Indeed, thousands or millions of curated and labeled examples are usually required for deep learning algorithms, which points to the promise of unsupervised approaches that do not require expert feedback during training and can learn from messier data or from less data. Combining algorithmic approaches to feature extraction and morphometry with machine learning and unsupervised approaches has great potential applications in characterizing not just healthy human brain variation but in diagnosing, tracking, and predicting unhealthy conditions.
Supporting information

S1 Supplement. Mindboggle output directory tree.
(PDF)

S2 Supplement. Mindboggle flowchart. Nipype automatically generates a flow diagram of the processing steps when running Mindboggle.
(PDF)

S3 Supplement. Travel depth.
(PDF)

S4 Supplement. Tables of shape differences between scans and between hemispheres.
(PDF)

(PDF)

Acknowledgments

We sincerely thank everyone who has contributed to the Mindboggle open science project over the years, including Nolan Nichols, Oliver Hinds, Arthur Mikhno, Hal Canary, and Ben Cipollini. We also thank Gang Li, Denis Riviére, and Olivier Coulon for assistance with the fundus evaluation. Arno Klein would like to thank Deepanjana and Ellora for their continued patience and support, and dedicates this project to his mother and father, Karen and Arnold Klein.

Software and online resources that have benefitted the project include: GitHub.com, Read-theDocs.org, and PyCharm for software development; Open Science Framework, Harvard Dataverse, and Synapse.org for public data storage; and D3, Bokeh, Paraview, and the Viridis colormap (https://github.com/bids/colormap) for visualization.

Author Contributions

Conceptualization: AKl.
Data curation: AKl.
Formal analysis: AKl ECN.
Funding acquisition: AKl.
Investigation: AKl.
Methodology: AKl FSB JG YH.
Project administration: AKl.
Resources: AKl.
Software: AKl SSG FSB JG YH ES NL BR MR AKe.
Supervision: AKl.
Validation: AKl.
Visualization: AKl JG MR ECN AKe.
Writing – original draft: AKl.
Writing – review & editing: AKl SSG FSB JG YH ES NL MR.

References


