July 18, 2018

Bayesian Analysis of Retinotopic Maps

Noah C. Benson[1*], Jonathan Winawer[1,2]

[1]Department of Psychology and [2]Center for Neural Sciences, New York University, New York, NY

[*] To whom correspondences should be addressed: nben@nyu.edu

Abstract

Human visual cortex is organized into multiple retinotopic maps. Characterizing the arrangement of these maps on the cortical surface is essential to many visual neuroscience studies. Typically, maps are obtained by voxel-wise analysis of fMRI data. This method, while useful, maps only a portion of the visual field and is limited by measurement noise and subjective assessment of boundaries. We developed a novel Bayesian mapping approach which combines observation–a subject's retinotopic measurements from small amounts of fMRI time–with a prior–a learned retinotopic atlas. This process automatically draws areal boundaries, corrects discontinuities in the measured maps, and predicts validation data more accurately than an atlas alone or independent datasets alone. This new method can be used to improve the accuracy of retinotopic mapping, to analyze large fMRI datasets automatically, and to quantify differences in map properties as a function of health, development and natural variation between individuals.

Introduction

Visual responses in a substantial part of the human brain are organized in retinotopic maps, whereby nearby positions in the brain represent adjacent locations in the image. Accurate measurement of these maps using functional magnetic resonance imaging (fMRI) is essential to a wide range of neuroscience and clinical applications [1]. Maps provide a basis to compare or to aggregate measurements across individuals, groups, tasks, stimuli, and laboratories. Maps are used to study homology between species[2], cortical plasticity [3], and development [4]. Many studies of cortical visual function in human - whether in motion [5], color [6], object recognition [7], or attention [8] - include retinotopic mapping as a first step. Finally, basic properties of the maps themselves, such as the cortical magnification function (mm of cortex per degree of visual field), can be used to understand visual performance [9]. Hence, the ability to accurate and efficiently measure retinotopic maps is of broad importance in visual neuroscience.

Currently, most retinotopic analyses of fMRI data use a voxel-wise approach. The general method is to (1) measure responses to mapping stimuli, (2) derive retinotopic coordinates for each voxel or surface vertex by analyzing traveling waves [10, 11]or by solving a population receptive field (pRF) model for each voxel [12], and (3) identify areal boundaries by visual inspection. Aside from requiring significant time and effort, the maps that result from this process retain many common sources of error including distortion of the BOLD signal due to partial voluming [13], vessel artifacts [14], other sources of physiological noise, and model fitting biases [15]. Due to the various sources of noise, the measured maps have discontinuities and often systematically miss portions of the visual field, such as the vertical meridian [16-20]. Further, the measured maps are limited by the available stimulus field of view in the scanner, often as little as 6-12° of eccentricity, and have difficulty measuring the foveal representation [21], the portion of the maps most important for many visual tasks including reading and object

recognition [22]. These many shortcomings of the traditional retinotopic mapping process derive from the fact that it is organized around optimizing the explanatory power of the retinotopy solutions from individual voxels, rather than that of the entire visual field or cortical area. As a consequence, it yields maps that are neither smooth nor complete—nor grounded in any context of how the visual field is warped onto the cortical surface. Lacking these data, the comparison of maps between subjects is difficult, and precise quantitative examination of individual differences is impossible.

An alternative to voxel-wise modeling of fMRI data is to build a retinotopic atlas - a computational model of the mapping between visual field position and cortical structure. Atlases have been developed to describe the mean anatomical locations [23] and the mean functional organization [24, 25] of the V1-V3 retinotopic areas across subjects. The atlas, after being fit to training data, is applied to an individual anatomical MRI image without functional data. These atlases are useful, but they provide only a description of the mean and do not offer a method of characterizing maps in an individual subject. A tool capable of providing such a characterization would be of considerable utility to researchers interested in homology between species [26], individual differences among humans [27, 28], and basic principles of cortical organization [1] and development [29].

Individual variation in cortical maps between subjects can be quite large. For example, the surface area of V1 can vary by 2- to 3-fold across healthy adults [27, 30, 31]. Despite these very large differences, if the surface anatomies are co-registered between individuals, the map topographies sometimes show reasonable agreement. For example, atlases of V1-3 [24, 25] have been derived using topological models of the cortical field [24, 32] with anatomical registration between observers [33, 34]. The assumption underlying these atlases is that once a correspondence is found between the sulcal pattern in two subjects' visual cortices, the function will match. Hence if one were interested in individual variation in cortical topography *after* anatomical registration, this method is uninformative: it

assumes the answer is 0. In this paper, we develop a novel method and toolset for modeling retinotopic maps in individual subjects. This method provides a precise quantification of the space between average and individual maps, enabling us to examine differences in cortical organization, even after co-registration of the surface anatomies.

Here, we hypothesize that a Bayesian model of retinotopic maps, combining data (retinotopic measurements from voxels or vertices) with a prior (a full-field atlas derived from the anatomy), will eliminate many of the issues with retinotopic mapping (described above) by optimizing the description of cortical retinotopic maps in the context of the full visual field and cortical map. We propose that such methods can describe cortical retinotopic maps in individual subjects more accurately than an atlas alone or measurements alone. These hypotheses are motivated by two factors. First, previous work employing functional data to supplement global anatomical alignments between subjects has found an increase in the overlap of functional ROIs drawn from independent localizers [35]. Thus, even when subjects are aligned anatomically, appreciable and systematic differences in the structure-function relationship remain. Allowing the measurement from an individual subject to inform the alignment will, in part, capture these individual differences. Secondly, we believe that the basic form of the atlas (the prior) is sufficiently accurate that incorporating it will result in a more accurate estimate of the retinotopic map than the measurements alone.

In this paper, we employ a Bayesian maximum-likelihood optimization to describe the retinotopic maps in striate and extra striate cortex with previously infeasible precision. Unlike previous work on functional alignment [36], we perform alignment between each individual subject's retinotopic parameters and a model of retinotopy described on the (anatomically-aligned) group-average cortical surface. This optimization builds on previous work using iterative approaches to fit and interpolate smooth retinotopic maps in individual subjects [27] by incorporating an explicit prior in the place of

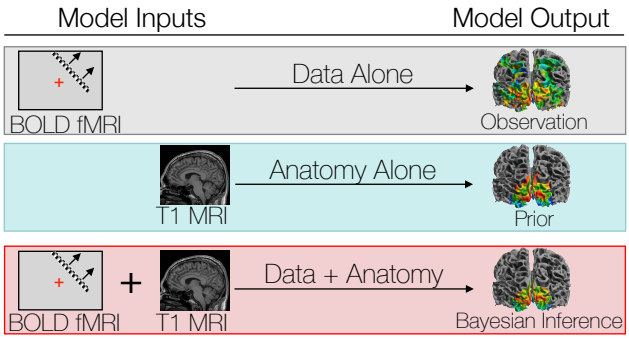

Figure 1. We compare three different ways to predict a subject's retinotopic maps. The first method is to perform a retinotopic mapping experiment. The fMRI measurements are converted to retinotopic coordinates by a voxel-wise model and projected to the cortical surface. Although a model is used to identify the coordinates for each vertex or voxel, we call this 'Data Alone' because no spatial template of retinotopy is used. The second method is to apply a retinotopic atlas to an anatomical scan (typically a T1-weighted MRI) based on the brain's pattern of sulcal curvature. This is called 'Data Alone' because no functional MRI is measured for the individual. The third method combines the former two methods via Bayesian inference, using the brain's anatomical structure as a prior constraint on the retinotopic maps while using the functional MRI data as an observation.

human intervention and adopting an explicitly Bayesian formulation. We additionally publish with this paper a tool capable of implementing the method we describe as well as all source code employed. We use these tools to characterize retinotopic maps from several subjects in terms of the precise warping from visual field to visual cortex. Using these characterizations, we are able to quantify the extent to which variations in retinotopic organization are due to anatomical differences versus differences in the structure-function relationship. We show that, in fact, these two sources of variability—differences in structure and differences in the structure-to-function mapping—are roughly equal and orthogonal across subjects. This means that after warping individual cortical surfaces to bring the anatomies into registration, an additional warping, equal in size, is needed to bring the functional maps into alignment, thereby demonstrating substantial variability in an early, sensory region of the human brain.

Results and Discussion

Retinotopic mapping experiments were performed on eight subjects using fMRI. Twelve individual retinotopy scans were performed on each subject then combined into 22 datasets for cross-validation as well as one full dataset of all scans for detailed analysis (Fig. S1; see also Materials and Methods).

Predicted maps were then generated using the datasets and compared to the validation dataset. We compared three methods for predicting retinotopic maps (Fig. 1): (1) using the training data alone as a prediction (the "observed" maps); (2) using the subject's anatomy alone to apply an anatomically-defined template of retinotopy (the "prior" maps); and (3) combining data with anatomy (observation with prior) using Bayesian inference (the "inferred" maps). We then leverage the differences between these methods to characterize the pattern of individual differences in retinotopic maps across subjects. The prior map, used for methods 2 and 3, was derived from fitting a template to a high-quality dataset derived from 181 subjects in the Human Connectome Project [37-39].

Bayesian inference has the advantages of an anatomical atlas while also respecting individual differences.

The inferred retinotopic maps, predicted by Bayesian inference, provide high-quality descriptions of the full retinotopic topology for each subject's V1-V3 regions. These maps can be produced even in the absence of observed retinotopic measurements (i.e., using the prior alone), or by combination with retinotopic data. In this latter case, the inferred maps have all the advantages of the retinotopic prior (topologically smooth maps, predictions beyond the stimulus aperture, etc.), while also accounting for idiosyncrasies in individual maps. Three examples of maps that demonstrate this advantage are shown in Figure 2. The first two columns show maps in which, relative to the validation maps, the predictions made from data alone have highly curved iso-eccentricity contours. These contours reflect noise rather than true curvature in the iso-eccentricity contours, as shown by the validation data. For these two columns, the predictions from the prior alone have iso-eccentric contours that are too smooth (as

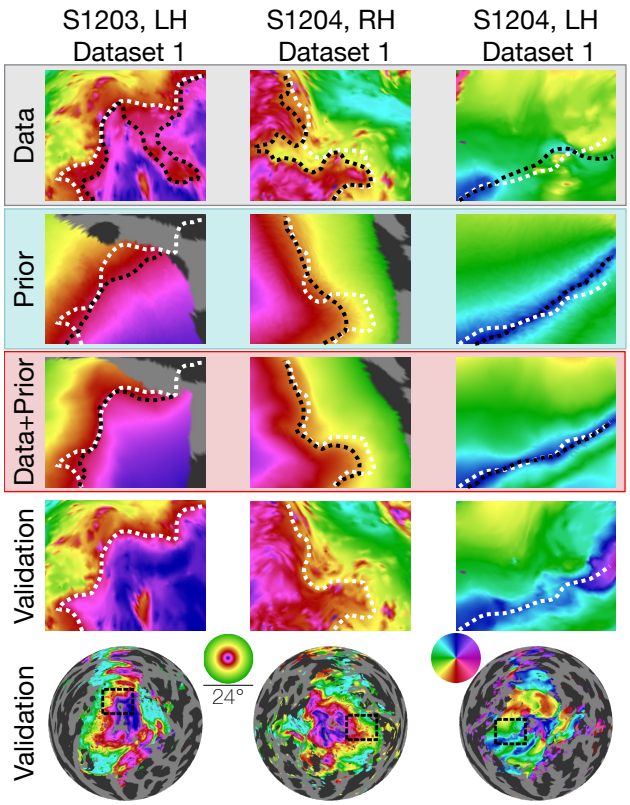

Figure 2. Inferred retinotopic maps accurately predict features of validation retinotopy. Twelve close-up plots of the retinotopic maps of three hemispheres are shown with predictions made from Data alone, Prior alone, or Data + Prior. The bottom two rows show the validation dataset with the bottom row indicating the context of the close-up patches. The first three rows show different methods of predicting the retinotopic maps, as in Figure 1. Approximate iso-eccentricity or iso-angular contour lines for the validation dataset have been draw in white on all close-up plots while black contour lines show approximate contour lines for the three prediction methods. Flattened projections of cortex were created using an orthographic projection (Fig. S2*A*).

compared to the validation data). The correct lines appear to lie between the training data and the prior. Hence when data is combined with the prior (Fig. 2*C*, 3rd row) the iso-eccentric contours resemble those of the validation dataset. The third column of Figure 2 shows an instance in which, even lacking a coherent polar angle reversal to define the ventral V1/V2 boundary near the fovea in the predictions made from data alone, combination of the data with prior more accurately predicts that boundary in the validation dataset than the prior alone.

In constructing the Bayesian-inferred retinotopic map for a single hemisphere, we perform two deformations, detailed in Figure 3 (see also Materials and Methods): (1) we first deform that hemisphere's inflated surface to register it to an average anatomical atlas using FreeSurfer (leftmost arrow in Fig. 3*C*); and (2) we then further deform the surface to register it to the retinotopic prior, based on the hemisphere's retinotopic measurements (second arrow in Fig. 3*C*). These steps together account for the individual differences in the organization of the subjects' retinotopic maps. In step 1, we account

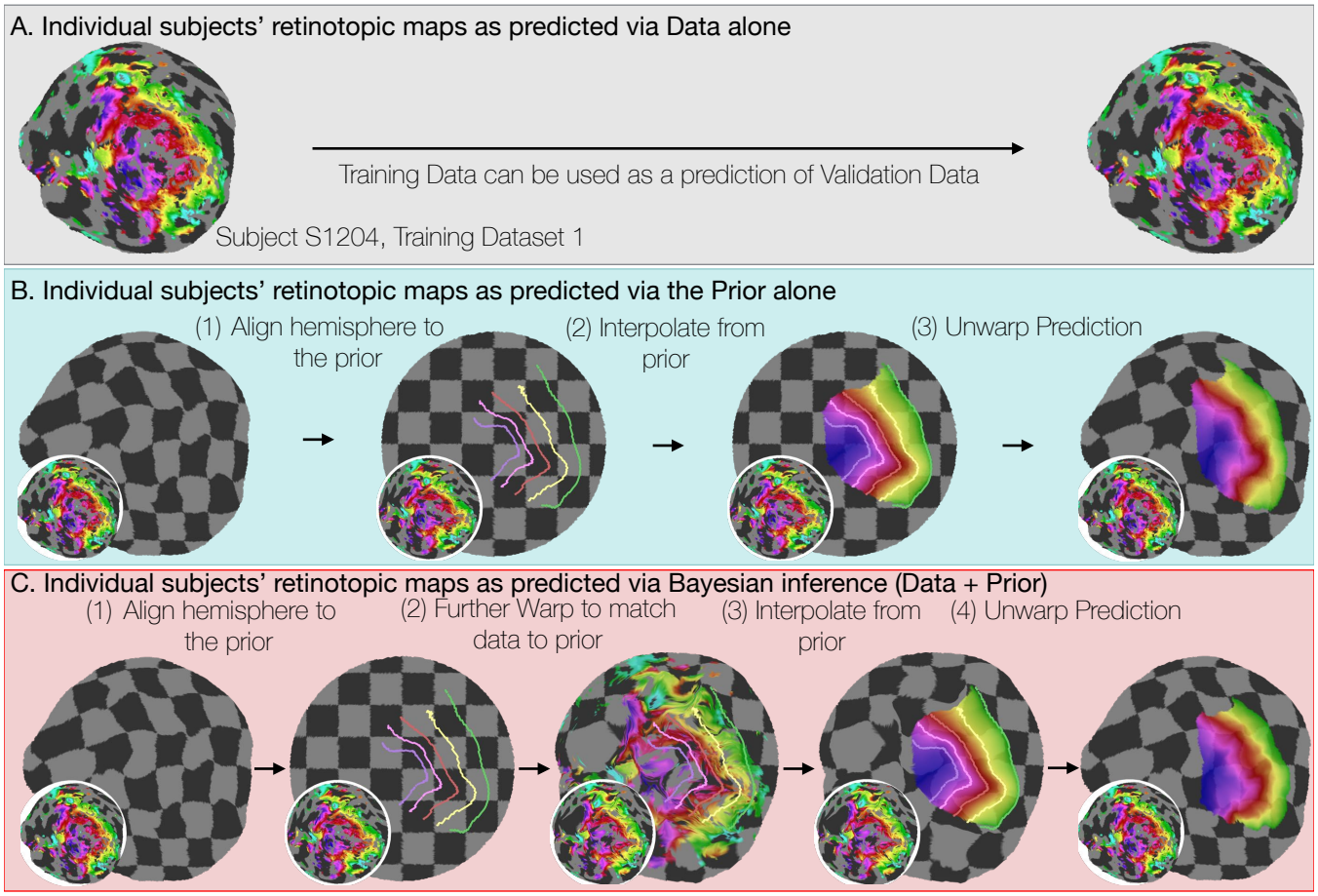

Figure 3. Deriving retinotopic predictions. Three methods of predicting retinotopic maps (as in Fig. 1) for an example subject. ***A*,** Predicted retinotopic maps based on training data alone are found by solving the pRF models for each voxel and projecting them to the cortical surface. The training data (left) and prediction (right) are identical. ***B*,** To predict a retinotopic map using the prior alone, the subject's cortical surface is aligned to FreeSurfer's *fsaverage* anatomical atlas (represented by rectilinear checkerboards), bringing the subject's anatomy into alignment with the anatomically-based prior, which is represented by iso-eccentricity contour lines in the figure (see also Supplemental Fig. 2*C*). The model of retinotopy is then used to predict the retinotopic parameters of the vertices based on their anatomically-aligned positions. After the predictions have been made, the cortical surface is relaxed. Maps are shown as checkerboards in order to demonstrate the warping (insets show original data and curvature). ***C*,** Bayesian inference of the retinotopic maps of the subject are made by combining retinotopic data with the retinotopic prior. This is achieved by first aligning the subject's vertices with the *fsaverage* anatomical atlas (as in *B*) then further warping the vertices to bring them into alignment with the data-driven model of retinotopy (shown as iso-eccentricity contour lines). The warping was performed by minimizing a potential function that penalized both the deviation from from the prior (second column) as well as deviations between the retinotopic observations and the retinotopic model.

for structural differences across subjects—the deformation that occurs for this registration is unique for each subject. This is where prior work ended [24, 25]. In step 2, we account for the differences in the relationship between structure and function *across* subjects. Although it is possible that deformations in step 2 partly compensate for imperfections in the first two registrations, we propose that, to a first

approximation, the deformations applied in the last step indicate meaningful individual differences in the structure-function relationship.

Individual differences in the V1-V3 structure-function relationship across subjects are substantial. The Bayesian model fitting allows us to parcellate two sources of variation between individuals: differences in surface anatomy (sulcal patterns) and differences in structure-to-function mapping, i.e. how retinotopic features map onto the surface anatomy. These two sources of variation map approximately to the two deformations in our atlas fitting: differences in surface anatomy are reflected in the deformation used for the surface alignment, and differences in structure-to-function mapping are reflected in the deformation to align the retinotopic data to the prior (the Bayesian fit). There are some

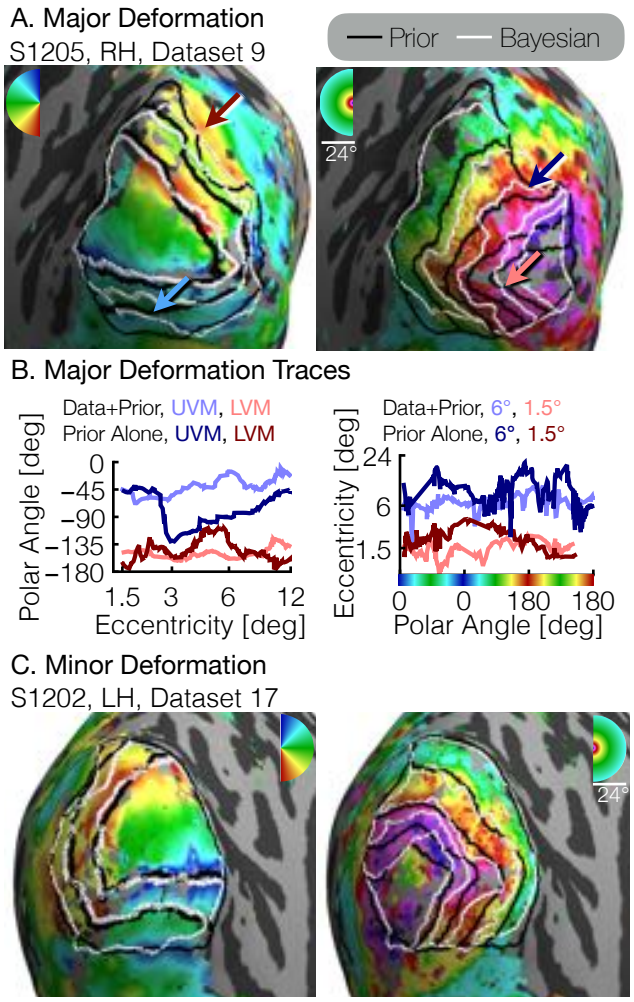

Figure 4. Comparison of inferred and prior maps. ***A,*** a subject whose maps were poorly predicted by the retinotopic prior and thus required major deformation (S1205, RH, dataset 9). ***B,*** To illustrate the differences between the Prior Alone (black lines in *A*) and the combination of Data + Prior (white lines in *A*), traces of the polar angle (right) and eccentricity (left) values beneath the lines indicated by arrows are shown. The eccentricities traced by the iso-angle lines and the polar angles traced by the iso-eccentricity lines of the Bayesian-inferred maps more closely match the angles/eccentricities of their associated trace lines than do the polar angles/eccentricities beneath the lines of the Prior alone ***C,*** A subject whose retinotopic maps were well-predicted by the prior and thus required relatively minor deformation during the Bayesian inference step (subject S1202, LH, dataset 17). In both *A* and *C*, black lines show the retinotopic prior and white lines show the maps inferred by Bayesian inference.

subjects for whom there are large differences between the retinotopic atlas defined by the prior and the atlas defined by the Bayesian fit (Figs. 4*A*, 4*B*). In this example subject, the iso-eccentricity lines in the Bayesian atlas are substantially more compressed along the posterior-anterior axis compared to the anatomical atlas, and the iso-angle lines in V2/V3 are more dorsal compared to the anatomical atlas. Where there are discrepancies, the Bayesian inferred map is more accurate. For example, the polar angles and the eccentricities in the validation data are approximately constant where the Bayesian map predicts iso-angle and iso-eccentricity lines, but not where the prior map predicts them (Fig. 4*B*). This indicates that even after anatomical alignment, the retinotopy in this subject differs systematically from the prior. For some other subjects the two atlases are in closer agreement such that the prior alone is a good fit to the retinotopic data (Fig. 4*C*).

To quantify the two types of deformations, we calculated the mean 3×3 distance matrix between (1) a vertex's native position, (2) its position in the anatomical alignment (*fsaverage* position), and (3) its "retinotopic position" after alignment to the retinotopic prior. All vertices in the V1-V3 region within 12° of eccentricity, as predicted by the Bayesian inference on the full dataset (all retinotopy scans combined), were used. We then performed 2D metric embedding to determine the mean deformation steps and the mean angle between them (Fig. 5*A*). Overall, the mean deformation distance across vertices is 3.1° ± 0.8° ($\mu \pm \sigma$ across subjects) of the cortical sphere for the anatomical alignment and 3.3° ± 0.4° for the retinotopic alignment. The mean angle between these deformations is 82.5° ± 9.5° (note that this last measurement is in terms of degrees of rotational angle rather than degrees of the cortical sphere). Of these two deformations, the latter, in which vertices are driven away from their anatomically-aligned positions according to their associated retinotopic measurements, is about 10%

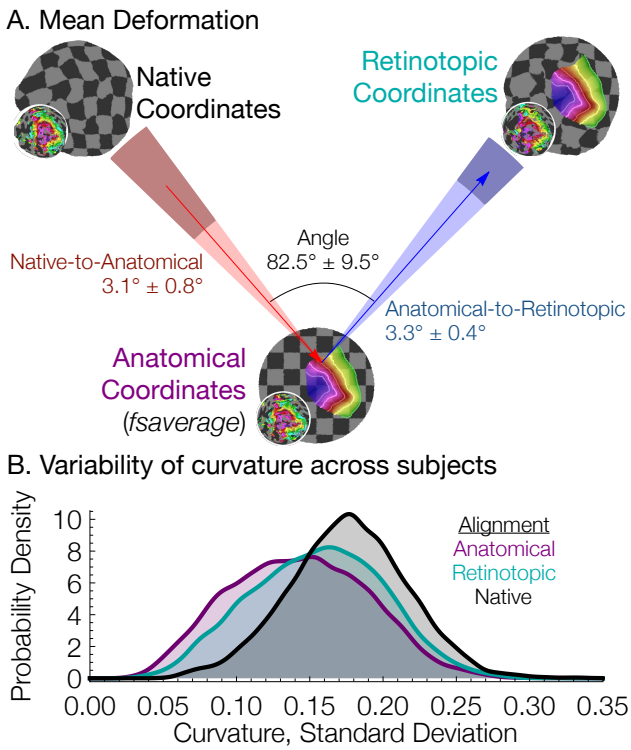

Figure 5. Individual differences between subjects in the structure-function relationship are significant. ***A,*** The mean deformation vectors, used to warp a surface vertex from its Native to its Anatomical (*fsaverage*-aligned) position and from its Anatomical to its Retinotopic position, are shown relative to each other. The wedges plotted beneath the mean arrows indicate ±1 standard deviation of the angle across subjects while the shaded regions at the end of the wedges indicate ±1 standard deviation of the lengths of the vectors. Note that because registration steps are always performed on a subject's inflated spherical hemispheres, these distances were calculated in terms of degrees of the cortical sphere and are not directly equivalent to mm of cortex. ***B,*** The alignment of the V1-V3 region to the retinotopic prior increases the standard deviation of the surface curvature across subjects, suggesting that retinotopic alignment is not simply an improvement on FreeSurfer's curvature-based alignment. Histograms show the probability density of the across-subject standard deviation of curvature values for all vertices in the V1-V3 region with a Bayesian-inferred eccentricity between 0 and 12°.

larger than the the anatomical alignment, in which vertices are driven away from their positions in the subject's native cortical sphere according to the surface curvature. The anatomical alignment corresponds to variation in the surface topology and is accounted for in anatomical atlases [24, 25]; the retinotopic alignment corresponds to variation in the structure-function mapping and is accounted for in the Bayesian model. An additional summary measurement of these deformations, the root-mean-square deviation (RMSD) distances across vertices near the occipital pole in a particular retinotopy dataset, provides a metric of the total warping applied during each step of the alignment process for each subject. A summary of this measurement, as well as various other summary statistics is provided in Table 1.

For both left (t_8 = -8.72; p < 0.0001) and right (t_8 = -5.19; p = 0.0013) hemispheres, the retinotopic deformation distances were significantly larger than the the anatomical deformation distances, which were still substantial; this is true whether one calculates the mean deformation distance over the entire patch of cortex immediately around the occipital pole (shown in the maps in Fig. *2*) or over only the vertices that are predicted to be in V1-V3; Figure 5 is calculated over the latter of these

two ROIs. Note that if the retinotopic deformation distances had been much larger, the prior anatomical atlases would have been less accurate. Had they been close to 0, then the anatomical atlas alone would have been as accurate as the Bayesian model.

We interpret the warping performed to align retinotopic data to the anatomical prior (the retinotopic alignment) as evidence of individual differences in the way in which retinotopy maps to

Table 1. Summary statistics for each subject.

Subject	Hemi	V1 Area (mm^2)[1]	V1 Volume (mm^3)[1]	Anatomical RMSD[2]	Retinotopic RMSD[2]
S1201	RH	1,308	3,733	2.88	3.15
S1201	LH	1,315	4,133	1.82	2.47
S1202	RH	2,024	3,706	1.21	2.73
S1202	LH	2,085	4,199	1.28	2.65
S1203	RH	1,574	3,152	2.13	3.27
S1203	LH	1,489	2,941	2.06	3.77
S1204	RH	1,906	3,325	2.29	3.00
S1204	LH	1,645	3,015	2.18	3.10
S1205	RH	1,995	3,926	1.99	2.91
S1205	LH	1,884	3,372	1.76	3.31
S1206	RH	1,647	3,116	2.12	2.73
S1206	LH	1,632	2,692	1.63	3.22
S1207	RH	1,648	3,402	1.84	2.41
S1207	LH	1,421	2,764	1.74	3.15
S1208	RH	2,069	4,290	2.01	2.35
S1208	LH	1,838	4,547	2.04	2.81

[1] The V1 boundary was determined from the Bayesian-inferred map constructed by combining the retinotopic prior with the full retinotopy dataset.

[2] RMSD values were calculated using the registration of the validation dataset of each hemisphere to the retinotopic prior, over the inner 12° of eccentricity of the V1-V3 region. Use of a larger patch of cortex (e.g., the flattened map projections in Figure 3*A*) does not qualitatively change the relationship between anatomical and retinotopic RMSD values. Units of the RMSD values are degrees of the cortical sphere; these are approximately equivalent to mm, but exact measurements in mm are distorted during inflation of the surface. "Anatomical" RMSD refers to the deviation between the subject's native anatomical sphere and the *fsaverage*-aligned sphere while "Retinotopic" RMSD refers to the deviation between the *fsaverage*-aligned sphere and the retinotopically-aligned sphere.

sulcal topology. An alternative possibility is that the retinotopic alignment corrects for incomplete or incorrect warping performed by FreeSurfer during alignment of each subject's sulcal topology to that of the *fsaverage* hemispheres (the anatomical alignment). Two observations cast doubt on this interpretation, however. First, the angle between the alignments is roughly orthogonal, meaning that there is very little movement along the axis of the first (structural) alignment during the second (retinotopic) alignment. Had the first alignment been in the correct direction but too conservative, then we would have expected the retinotopic alignment to be in the same (or similar) direction, rather than orthogonal. Second, if the retinotopic alignment served to correct an incomplete or incorrect anatomical warping, then anatomical metrics such as curvature would become more uniform across subjects after retinotopic alignment compared to after only anatomical alignment. Figure 5*B* demonstrates that this explanation is unlikely by showing the distribution across surface vertices of the standard deviation of curvature across subjects. When curvature is compared across subjects without retinotopic or anatomical alignment (Native alignment in Fig. 5*B*) the standard deviation is quite high. When subjects are compared after anatomical alignment, the standard deviations are much lower. After further alignment to the anatomical prior of retinotopy (Retinotopic alignment), the standard deviation of curvature across subjects is between these two extrema. This suggests that the retinotopic alignment is sacrificing some amount of structural uniformity across subjects in order to accommodate the individual differences in subjects' structure-to-function mapping, and is consistent with our interpretation that there are substantial individual differences in the mapping between retinotopy and surface topology.

The large individual differences that remain, even after structural co-registration, point to the importance of using at least some individual subject functional data when inferring the maps, rather than assuming the atlas (prior) is correct. The specific nature of these deformations, and whether, for example, they fall into a few basic patterns, is an important question about the natural variation of

individual brains. Our new method, combined with large datasets such as the HCP retinotopy data set [39], could be used to address this question.

The inferred maps make highly accurate predictions using very little data.

To quantify the accuracy of our Bayesian-inferred retinotopic maps, and to compare the accuracy against other predictions, we used a cross-validation scheme such that predictions from data alone, the prior alone, or via Bayesian inference were compared against independent validation datasets (Fig. S1). The validation datasets were derived from 6 of the 12 scans; the predictions from data alone and from Bayesian inference were derived from training datasets, which were comprised of various combinations of 1-6 independent scans (between 3.2 and 19.2 minutes of data). The predictions from the prior alone did not use training data.

Within the validation dataset, all vertices from the inner 12° of eccentricity of the V1-V3 maps, as identified via Bayesian inference using the validation dataset, were compared to their counterparts in the predicted maps. Note that the inferred maps from the validation dataset were used only to identify the vertices included in the comparison; the retinotopic coordinates from the validation datasets themselves were taken as the "gold-standard" measurements. In computing prediction accuracy, we weighted the vertices by the fraction of variance explained for each vertex's pRF solution in the validation dataset. For the three types of training datasets (prior alone, data alone, Bayesian inference),

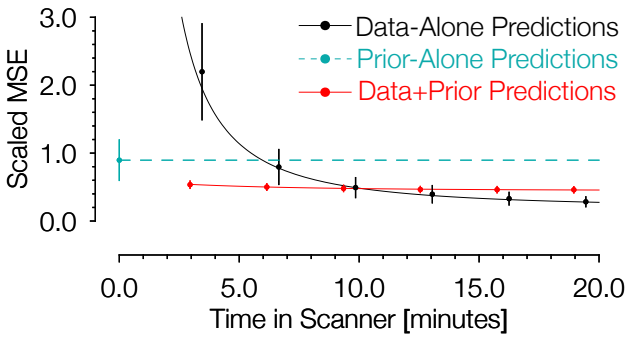

Figure 6. Comparison of prediction errors for three methods of predicting retinotopic maps. Errors are shown in terms of the number of minutes spent collecting retinotopic mapping scans (*x*-axis). The *y*-axis gives the mean squared eccentricity-scaled error. Each plotted point represents a different number of minutes in the scanner, with error bars plotting ± 1 standard error across subjects. For short scan times, errors are significantly higher for predictions made with the data alone than for those made using Bayesian inference. An offset has been added to the *x*-values of each of the black (-0.25) and red (+0.25) points to facilitate viewing.

we assume that each vertex makes a prediction regardless of the variance explained. To prevent errors at high eccentricity from dominating the error metric, we calculated the scaled error for a vertex to be the angular distance in the visual field between its retinotopic coordinates in the predicted map and the validation dataset, divided by the eccentricity in the validation dataset. Figure 6*A* shows the scaled mean squared error (MSE) for the various predicted maps in terms of the amount of time spent collecting retinotopic data for the map.

The maps predicted via Bayesian inference were highly accurate irrespective of the amount of data used to inform the fits (Fig. 6). Between those predicted maps informed by 3.2 minutes of scan time (1 scan) and 19.2 minutes (6 scans), the scaled MSE of the prediction remains in the range of ~0.4-0.5. These scaled errors are larger near the fovea because the denominator used for scaling the error metric (i.e., the eccentricity) could be very small; when the range is limited to 3 to 12 deg, the MSE is much lower, ~0.20-0.23. Expressed separately in units of polar angle and eccentricity, the mean absolute polar angle error from a Bayesian map derived from a single 3.2 minute scan is 22° ± 4.6° and the mean absolute eccentricity error is 0.69° ± 0.11° ($\mu \pm \sigma$ across 96 datasets). For the prior alone, these errors are substantially higher: 31° ± 11° for polar angle and 2.0° ± 1.3° for eccentricity. Note that these errors are approximately 3× higher than those reported for previous versions of the anatomical prior [25]; however these discrepancies are due to differences in the metric used, the amount of data collected, the thresholding applied, and the use of smoothing. Some of these factors we cannot reproduce exactly (amount of data) or have deliberately abandoned (smoothing), but by using the same metric (median absolute error across all vertices) and similar thresholding (1.25° < predicted eccentricity < 8.75°), we obtain errors very close to those previously reported: 12° of polar angle and 0.90° of eccentricity. In contrast to the inferred maps for which the accuracy is largely independent of scan time, the accuracy of the predictions from data alone was highly influenced by scan time. The scaled MSE of the maps

predicted from the training datasets alone for the same range of scan times ranged from ~2.2 (3.2 minutes of training data) to ~0.3 (19.2 minutes of training data). With more than ~11 minutes of scan time, the predictions made from the training datasets alone have a slightly lower scaled MSE than those made from Bayesian inference, though, notably, the improvements in scaled MSE for more than 11 minutes of scan time are small.

The prediction accuracy from the prior alone (0 min. scan time) was generally intermediate in accuracy between the predictions from the Bayesian model and the data alone. Predictions derived from the anatomically-defined atlas alone are more accurate than predictions from 3.2 minutes of scanning and only slightly less accurate than predictions from training datasets derived from 6.4 minutes of scanning; this is in agreement with previous analyses of prediction error versus measurement error in retinotopic mapping experiments [25]. The predictions using the prior alone were universally less accurate than the Bayesian predictions (Fig. 6, cyan)—for all subjects and all training datasets, the combination of prior with data improved prediction accuracy compared to the anatomical prior alone. These data demonstrate that the application of our new method to a small amount of retinotopic mapping data yields a higher quality retinotopic map than can be derived from other sources alone, with the possible exception of data derived from a very long retinotopic mapping session.

The fact that the increased prediction accuracy from the Bayesian maps is almost independent of the amount of scan time used for the observations suggests that much of the individual variability is captured by a low dimensional warping from the template, which can be inferred from a modest amount of data.

Another significant advantage of the method is that it eliminates the need for human intervention in the process of delineating retinotopic maps and visual areas. In most studies that requite retinotopic mapping data, one or more experimenters hand-label the visual area boundaries. While human raters are

better able to understand atypical retinotopic boundaries than our method, they are nonetheless subject to inter-rater disagreement and human error. Furthermore, although expert human raters have a much more nuanced prior about retinotopic map organization than our method, and thus may sometimes draw boundaries better than our method, our method at least makes its prior explicit and quantifiable, and, thus, comparable and replicable across studies.

What is ground truth?

The motivation for a Bayesian approach to retinotopic mapping can be found most clearly in the measured retinotopic maps themselves. In all of our measured retinotopic maps, there are numerous systematic imperfections (Fig. 7*A*), and the literature contains many reports of similar errors [14, 40-42]. These imperfections can arise from a variety of sources, including partial voluming [43], negative BOLD [44], and large, draining veins. Imperfections due to blood vessel artifacts in particular, such as the transverse sinus in ventral visual cortex, can have effects over large distances [14], and most perniciously, they may lead to large and reliable responses that nonetheless differ from the local neuronal activity in the voxel [42]. Such artifacts can be difficult to track down and are often not eliminated by typical methods of cleaning up maps such as smoothing, thresholding, or simply collecting larger datasets.

Although with large datasets (> 19 min of scan time), the prediction accuracy for the validation dataset is highest using the data alone rather than the data and prior, we believe that even in these cases the combination of data + prior is probably closest to ground truth. We defined accuracy operationally as the difference from the validation set, as this provides a single set of independent measures that can be used to assess the accuracy of all 3 types of models. The validation dataset is defined by at least as many scans as any of the training datasets, and hence is our best measurement. However, the validation dataset is not ground truth, as it is subject to errors from systematic and random measurement noise. The

inferred maps, unlike the maps from the validation and training datasets alone, produce a topologically smooth transformation into the visual field: they represent the complete visual field with no holes, redundancies, or discontinuities. Hence an enclosed region in the visual field will project to an enclosed region in the inferred map on the cortical surface. This is not the case for the maps predicted by the data alone without the use of an atlas (Fig. 7*B*). We consider this difference to be an advantage of the inferred maps, since it is generally accepted that the cortical surface of V1 is a topological map of the visual field.

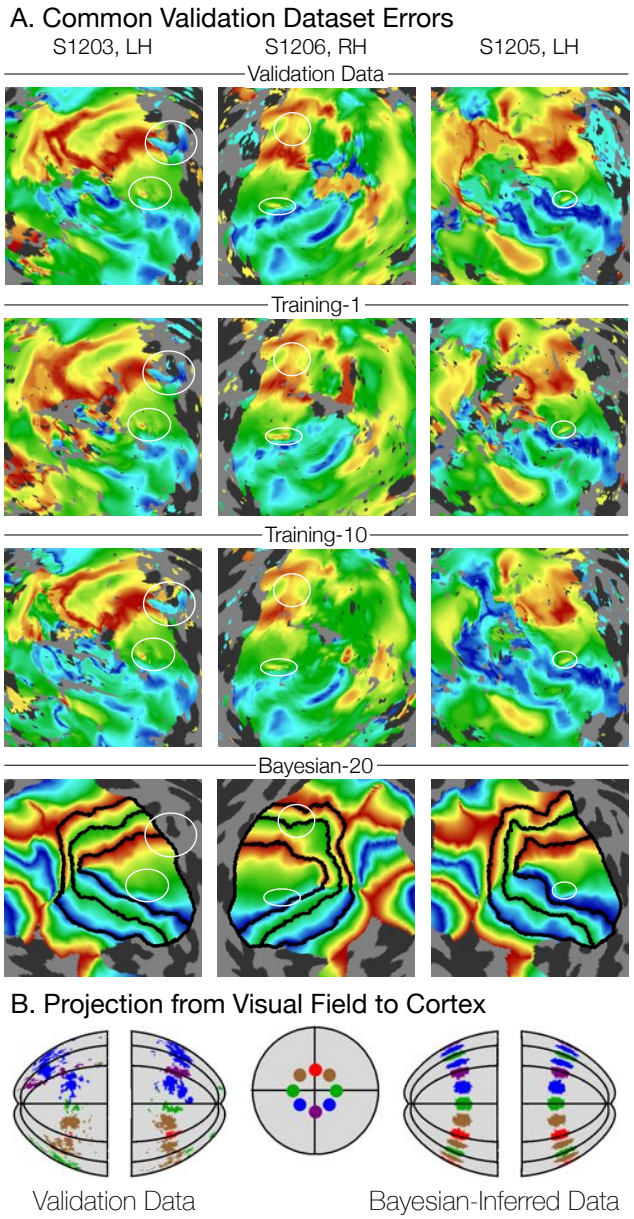

Figure 7. Systematic errors in training and validation datasets. ***A,*** Many small inconsistencies in the retinotopic maps are duplicated in both the validation dataset and the training datasets but not in the maps predicted by Bayesian inference. Maps for three example hemispheres are shown with validation datasets as well as training dataset 1, training dataset 10, and the Bayesian-inferred maps from dataset 20. Ellipses highlight blips of noise in the validation maps that are unlikely to represent the true underlying map, but that are correlated with the training maps. Such blips are significantly different in the inferred and validation maps, likely inflating the error of the inferred maps. Black lines show the V1-V3 boundaries in the Bayesian-inferred maps. ***B,*** Discontinuity errors. If the validation data is used to project the disks shown in the visual field in the middle panel to the cortical surface, the resulting map is messy and contains a number of inconsistencies due to measurement error. While the Bayesian inferred map may contain errors of its own, it will always predict a topologically smooth retinotopic map with respect to the topology of the visual field.

In short, since we do not correct for all of these potential sources of systematic error, we consider our estimates of error from the Bayesian-inferred maps to be conservative, and the estimate from the data-to-data predictions to be liberal.

The Bayesian model accurately predicts visual field positions not included in the training data. One important advantage of using the method of Bayesian inference outlined in this paper is that it provides predictions beyond the extent of the stimulus aperture in the retinotopy experiment. These peripheral predictions extend to 90° of eccentricity, even though the data used to derive the prior was based on stimuli that only extended to ~8º of eccentricity. Hence it is important to ask whether the model makes accurate predictions in the periphery. We demonstrate this in two ways. First, when our registration algorithm is run using only a subset of the eccentricity range (e.g., only data within the first 3° or the first 6° of eccentricity), the predicted maps remain accurate to 12° of eccentricity (Fig. 8*A*). Second, we compared wide-field retinotopy data, collected out to 48° of eccentricity from subject S1201

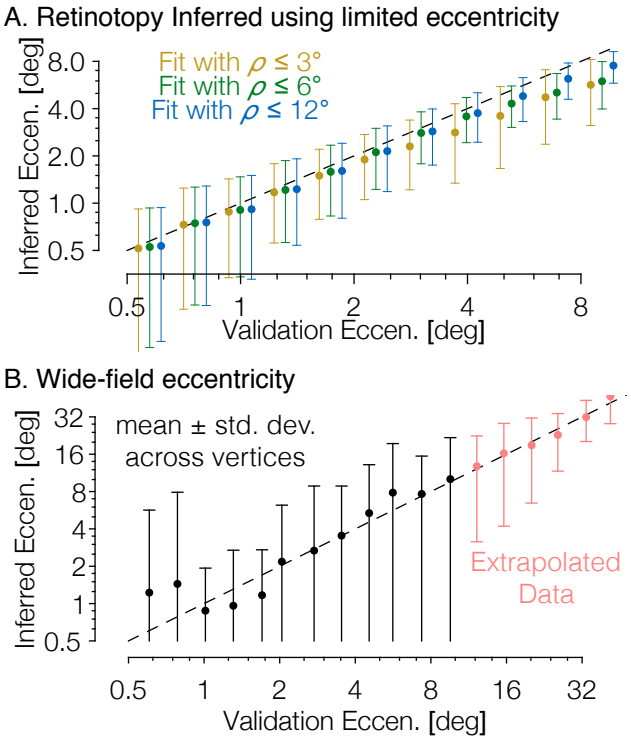

Figure 8. The Bayesian-inferred maps accurately predict eccentricity beyond the range of the stimulus. ***A*,** In order to examine how accurately the retinotopic maps predicted using Bayesian inference describe the retinotopic arrangement outside of the range of the stimulus used to construct them we constructed maps from all datasets using only the inner 3° or 6° of eccentricity then compared the predictions to the full validation dataset. Eccentricity is well predicted out to 12° regardless of the eccentricity range used to construct the predicted map, indicating that our inferred maps are likely accurate beyond the range of the stimulus. ***B*,** In addition, we compared the wide-field retinotopic mapping data from subject S1201 to the inferred retinotopic maps using only the 12° stimulus; the inferred eccentricity is shown in terms of the validation eccentricity. The highest errors appear in the fovea (< 3°), while predictions are most accurate in the periphery, indicating that eccentricity may be well-predicted far beyond the range of the stimulus (out to 48° of eccentricity in this case).

to the Bayesian-inferred map predictions made using our data with a 12°-aperture (Fig. 8*B*). We find that in both cases, our method is highly accurate despite lacking training data for peripheral measurements. Thus, even if prediction accuracy within the measured regions were similar for the Bayesian model and the data-to-data predictions, the Bayesian model is advantageous because it includes predictions for regions of the visual field beyond training data.

The Bayesian inferred maps accurately reproduce systematic properties of the visual field maps. Another aspect in which our work here extends previous methods is the addition of the pRF size to the retinotopic quantities predicted by the model in the inferred maps—previous work predicted only the pRF centers [24, 25]. Here, we predict the pRF sizes for the vertices based on the eccentricity inferred from the Bayesian map, and the relationship between eccentricity and pRS size in previous literature [45]. The inferred pRF size of a vertex is the pRF size predicted by the Kay et al. [45] model for the vertex's inferred visual area at the vertex's inferred eccentricity. Although pRF size can depend on variables such as the voxel size and the stimulus that may differ between scans, we find that this method approximately reproduces the pRF sizes found in the observed retinotopic maps (Figs. 9*A* and 9*B*).

Another metric inversely related to pRF size is the cortical magnification, usually measured in terms of mm^2 of cortex representing one $degree^2$ of the visual field. We summarize these measurements in Figures 9*C* and 9*D*. Our measurements of cortical magnification are broadly in agreement with previous work by Horton and Hoyt [46], shown by the dotted black line in the lower panels of Figure 9. The cortical magnification of the inferred maps is quite similar to those of the observed retinotopic maps. In both cases, V1 has slightly lower cortical magnification than V2 and V3 near the fovea, but higher magnification in the periphery. This difference is slightly exaggerated in the inferred maps relative to the observed maps; though this difference is slight and is in agreement with previous examinations of cortical magnification [21].

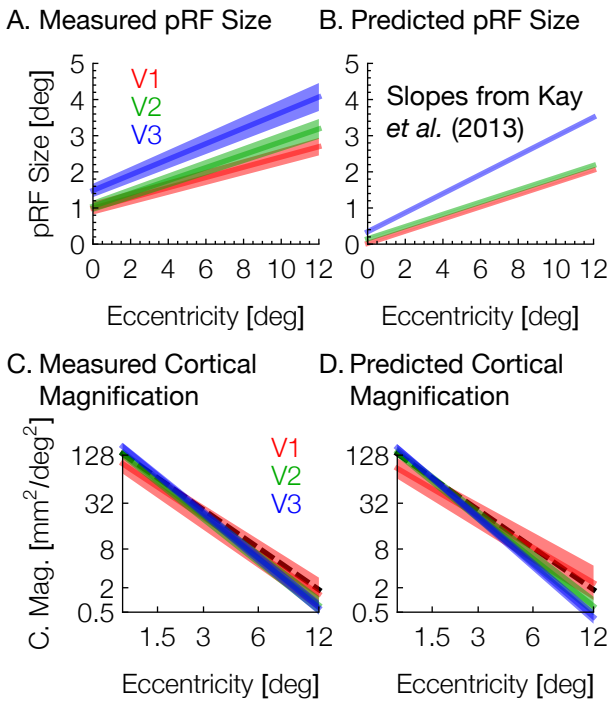

Figure 9. Aggregate pRF size and cortical magnification measurements are in agreement with previous literature. ***A,*** PRF sizes by eccentricity are shown for V1, V2, and V3, as calculated from the full datasets; shaded regions show standard errors across subjects. ***B***, PRF sizes predicted by the Bayesian inference; note that these values are predicted based on the inferred eccentricity of each vertex based on the slopes described by Kay et al. (2013). ***C,*** Cortical magnification is shown in terms of eccentricity for V1-V3, as calculated from the full datasets. Again, the shaded regions show standard errors across subjects. The dashed black line shows the equation for cortical magnification provided by Horton & Hoyt (1991). ***D,*** Cortical magnification as calculated using the pRF coordinates inferred by the Bayesian inference. Note that in both *A* and *C*, eccentricity refers to measured eccentricity while in *B* and *D*, eccentricity refers to Bayesian-inferred eccentricity.

The retinotopic prior and Bayesian-inferred maps include 12 visual areas.

Previous research on the retinotopic organization of visual cortex used a model of V1, V2, and V3 retinotopy described by Schira et al. [32] to produce a template of retinotopy that included only those visual areas. This "banded double-sech" model accurately describes the anisotropic magnification of the visual field on the cortical surface, particularly near the fovea. However, we have observed, particularly in individual data, that retinotopic data from outside the V1-V3 region described by the Schira model has a large impact on the quality of the inferred map. Accordingly, in creating our retinotopic prior, we constructed a new model that includes 9 additional visually active regions: hV4, VO1, VO2, V3a, V3b, LO1, LO2, TO1, and TO2. This model employed a new method of constructing 2D models of retinotopy that was specifically designed to accommodate the distortions caused by anatomical alignment, inflation, and flattening of the cortical surface. The new method is much simpler to extend to many more visual field maps, as it does not rely on an analytic description of the flattened (2D) retinotopic maps, which are only available for V1-V3. Rather, it requires as input a cortical map image on which estimates of the visual area boundaries have been drawn manually and labeled as either a foveal boundary, a peripheral

boundary, an upper vertical meridian, a lower vertical meridian, or a horizontal meridian. A minimization technique is then used to fill in the retinotopic coordinates between the drawn boundaries (see Models of Retinotopy in Materials and Methods). The new retinotopic prior, including all new areas can be seen in Figure S3. Although we consider the addition of 9 retinotopic areas to be an important development, we consider these areas preliminary and do not analyze them in detail here. One reason for this is that the organizations of many of these areas remain under dispute. Additionally, the responses to our stimuli in these areas is of a considerably lower quality than in V1-V3; thus even were we to analyze the accuracy of the predictions in these ares, our validation dataset would be a particularly poor standard. We do, however, include these areas in the predictions made by the Bayesian inference method so that interested researchers may analyze or use them. These areas are included in the data provided with this paper, and predictions of these areas are included when using the tools provided in the supplemental materials.

Making the Bayesian inference explicit.

Our method has the advantage of allowing the retinotopic atlas to act as a prior constraint on new observed data. This is a Bayesian model in the general sense of combining a prior belief with a measurement in order to make an inference. The computation can also be reformulated in an explicit Bayesian framework. We define a hypothesis H to be a particular warping of the cortical surface, and we define the evidence E to be a particular set of retinotopic measurements. We then convert the cost functions from Table 2 into probabilities by assuming an exponential relationship. Hence, the prior probability of H is defined in terms of the deviation from the retinotopic prior: $P(H) = \exp(-(F_e(\boldsymbol{x}) + F_\vartheta(\boldsymbol{x}) + F_p(\boldsymbol{x})))$, and the likelihood of the evidence under a given hypothesis, $P(E \mid H)$, is defined in terms of the fit between the retinotopic model and the retinotopic measurements: $P(E \mid H) = \exp(-F_\varphi(\boldsymbol{x}))$.

Table 2. Components of the registration potential function

Term	Description	Form
$F_e(\boldsymbol{x};\boldsymbol{x_0},\mathbf{E})$	Penalizes changes in the distances between neighboring vertices in the mesh.	$\frac{H_e(\boldsymbol{x};\boldsymbol{x_0},\mathbf{E})+G_e(\boldsymbol{x};\boldsymbol{x_0},\mathbf{E})}{\lvert\mathbf{E}\rvert}$
$F_\vartheta(\boldsymbol{x};\boldsymbol{x_0},\boldsymbol{\Theta})$	Penalizes the changes in the angles of the triangles in the mesh.	$\frac{H_\vartheta(\boldsymbol{x};\boldsymbol{x_0},\boldsymbol{\Theta})+G_\vartheta(\boldsymbol{x};\boldsymbol{x_0},\boldsymbol{\Theta})}{\lvert\boldsymbol{\Theta}\rvert}$
$F_p(\boldsymbol{x};\boldsymbol{x_0},\mathbf{P})$	Penalizes any change in the positions of the vertices on the perimeter of the map.	$\frac{1}{2}\sum_{u\in\mathbf{P}}\lVert(\boldsymbol{x})_u-(\boldsymbol{x_0})_u\rVert^2$
$F_\varphi(\boldsymbol{x};\boldsymbol{\Phi})$	Decreases as a retinotopic vertex *u* approaches its anchor-point ***y*** in the retinotopy model.	$\frac{\sum_{(u,\boldsymbol{y},\sigma,w)\in\boldsymbol{\Phi}} w\exp(-\frac{\lVert(\boldsymbol{x})_u-\boldsymbol{y}\rVert^2}{2\sigma^2})}{\sum_{(u,\boldsymbol{y},\sigma,w)\in\boldsymbol{\Phi}} w}$
$H_e(\boldsymbol{x};\boldsymbol{x_0},\mathbf{E})$	Harmonic component of the edge-length deviation penalty $F_e(\boldsymbol{x};\boldsymbol{x_0},\mathbf{E})$.	$\frac{1}{2}\sum_{(u,v)\in\mathbf{E}}\left(r_{\boldsymbol{x}}(u,v)-r_{\boldsymbol{x_0}}(u,v)\right)^2$
$G_e(\boldsymbol{x};\boldsymbol{x_0},\mathbf{E})$	Infinite-well component of the edge-length deviation penalty $F_e(\boldsymbol{x};\boldsymbol{x_0},\mathbf{E})$.	$\frac{1}{2}\sum_{(u,v)\in\mathbf{E}}\left(\left(\sqrt{\frac{r_{\boldsymbol{x_0}}(u,v)-q_0}{r_{\boldsymbol{x}}(u,v)-q_0}}-1\right)^2+\left(\sqrt{\frac{q_1-r_{\boldsymbol{x_0}}(u,v)}{q_1-r_{\boldsymbol{x}}(u,v)}}-1\right)^2\right)$
$H_\vartheta(\boldsymbol{x};\boldsymbol{x_0},\boldsymbol{\Theta})$	Harmonic component of the angle deviation penalty $F_\vartheta(\boldsymbol{x};\boldsymbol{x_0},\boldsymbol{\Theta})$.	$\frac{1}{2}\sum_{(a,b,c)\in\boldsymbol{\Theta}}\left(\alpha_{\boldsymbol{x}}(a,b,c)-\alpha_{\boldsymbol{x_0}}(a,b,c)\right)^2$
$G_\vartheta(\boldsymbol{x};\boldsymbol{x_0},\boldsymbol{\Theta})$	Infinite-well component of the angle deviation penalty $F_\vartheta(\boldsymbol{x};\boldsymbol{x_0},\boldsymbol{\Theta})$.	$\frac{1}{2}\sum_{(a,b,c)\in\boldsymbol{\Theta}}\left(\left(\sqrt{\frac{\alpha_{\boldsymbol{x_0}}(a,b,c)}{\alpha_{\boldsymbol{x}}(a,b,c)}}-1\right)^2+\left(\sqrt{\frac{\pi-\alpha_{\boldsymbol{x_0}}(a,b,c)}{\pi-\alpha_{\boldsymbol{x}}(a,b,c)}}-1\right)^2\right)$

During registration, we seek the hypothesis *H* that maximizes the posterior probability $P(H \mid E) = P(E \mid H)\, P(H)/P(E)$. Because $P(E)$ is a constant, we can ignore it and instead maximize the function given in Equation 1, which is equivalent to minimizing $F(\boldsymbol{x})$. This operation is performed during registration. Thus, to derive our cost function from Bayes' rule, we write:

$$
\begin{aligned}
&P(H \mid E) = P(E \mid H)\, P(H)/P(E)\\
&P(H \mid E) \propto P(E \mid H)\, P(H)\\
&P(H \mid E) \propto \exp(-(F_e(\boldsymbol{x}) + F_\vartheta(\boldsymbol{x}) + F_p(\boldsymbol{x}))) \exp(-F_\varphi(\boldsymbol{x}))\\
&P(H \mid E) \propto \exp(-(F_e(\boldsymbol{x}) + F_\vartheta(\boldsymbol{x}) + F_p(\boldsymbol{x}) + F_\varphi(\boldsymbol{x})))\\
&P(H \mid E) \propto \exp(-F(\boldsymbol{x})) \qquad (1)
\end{aligned}
$$

Individual differences in structure-function relationship.

An important question in human neuroscience is the degree to which different brains, when brought into anatomical registration, share the same functional mapping. There is no single, agreed-upon method to register the brains of different individuals, but a general finding is that cortical function shows better inter-observer agreement when the brains are aligned based on sulcal topology (surface registration) rather than volume registration [23, 47]. Here, using surface registration, we find that substantial individual differences in functional mapping remains, e.g. as evidenced by the amount of additional warping needed to align individual brains to retinotopic measurements (Fig. 5). These results are consistent with studies showing differences in structural and functional alignment of primate and human area MT [48]. They are also consistent with studies combining structural and functional alignment in the absence of a model or template of the underlying function [35, 36]. Such studies show that the extra warping driven by functional alignment leads to better predictions of functional responses in cross-validated data.

The anatomical atlas, described previously [25], was adapted into the first step of our method (Fig. 3*C*) and is equivalent to the prediction using the retinotopic prior alone that we present here (Fig. S2*B*). Consistent with previous results, we find that the prior alone produces reasonably good predictions for the retinotopic maps of most subjects (Fig. 6). Additionally, maps predicted using the prior alone contain many of the advantages described here such as topological smoothness, complete coverage of the visual field, and prediction of peripheral retinotopy. However the idiosyncrasies of individual subjects' retinotopic maps are often not well predicted by the retinotopic prior alone (Figs. 2, 4*A*, and 4*B*). By combining both the retinotopic prior and a small amount of measured data we are able to produce higher-quality predictions that not only share these advantages but also improve the prediction accuracy of the maps beyond that of the measurement error (Fig. 6).

Method Availability and Usage.

The method described in this paper has been made publicly available in a number of ways in order to enable easy use by other researchers. The Bayesian inference method itself is implemented as part of a free software library called Neuropythy; we have publicly archived the version of this library used in the preparation of this manuscript at DOI:10.5281/zenodo.1312983. Additionally, we have created a universally executable Docker image and have publicly archived it at DOI:10.5281/zenodo.1313859. Detailed instructions for the use of both the library and the universally executable Docker image are available at the Open Science Foundation repository associated with this paper (https://osf.io/knb5g/). In brief, the method may be run with only a few inputs: a FreeSurfer directory for the subject (which can be generated from an anatomical image), and a set of files containing the measured retinotopic parameters for the subject's cortical surface. The outputs produced are a similar set of files describing the inferred retinotopic parameters of the subject's cortical surface. Runtime on a contemporary desktop computer is less than an hour per subject.

Materials and Methods

Scientific Transparency.

All source code and notebooks as well as all anonymized data employed in this Methods section and the preparation of this manuscript have been made publicly available at the Open Science Foundation: https://osf.io/knb5g/

Subjects.

This study was approved by the New York University Institutional Review Board, and all subjects provided written consent. A total of eight subjects (4 female, mean age 31, range 26-46) participated in the experiment. All scan protocols are described below.

Magnetic Resonance Imaging.

All MRI data were collected at the New York University Center for Brain Imaging using a 3T Siemens Prisma scanner. Data were acquired with a 64-channel phased array receive coil. High resolution whole-brain anatomical T1-weighted images (1 mm^3 isotropic voxels) were acquired from each subject for registration and segmentation using a 3D rapid gradient echo sequence (MPRAGE). BOLD fMRI data were collected using a T2*-sensitive echo planar imaging pulse sequence (1s TR; 30ms echo time; 75° flip angle; 2.0 × 2.0 × 2.0 mm^3 isotropic voxels, multiband acceleration 6). Two additional scans were collected with reversed phase-encoded blips, resulting in spatial distortions in opposite directions. These scans were used to estimate and correct for spatial distortions in the EPI runs using a method similar to [49] as implemented in FSL [50].

Anatomical images were processed using the FreeSurfer image analysis suite, which is freely available online (http://surfer.nmr.mgh.harvard.edu/) [33, 34, 51, 52]. Subject brains were inflated and aligned to FreeSurfer's anatomical *fsaverage* atlas.

Stimulus Protocols.

Each subject participated in 12 retinotopic mapping scans using the same stimulus emptied in the Human Connectome Proect [39]. Briefly, bar apertures on a uniform gray background swept gradually across the visual field at four evenly-spaced orientations while the subject maintained fixation. Bar apertures contained a grayscale pink noise background with randomly placed objects, faces, words, and scenes. All stimuli were presented within a circular aperture extending to 12.4° of eccentricity. The bars were a constant width (1.5°) at all eccentricities. Subjects performed a task in which they were required to attend to the fixation dot and indicate when its color changed.

The 12 scans were split into several subsets and analyzed as independent datasets. Six of the scans (two of each bar width) were allocated to the subject's *validation dataset*, while the remaining

three scans were used for 21 training datasets: 6 datasets with 1 scan each, 5 datasets with 2 scans each, 4 datasets with 3 scans each, 3 datasets with 4 scans each, 2 datasets with 5 scans each, and 1 dataset with all 6 training scans. Additionally, all 12 scans were included in a *full dataset* which was used for all analyses not related to the accuracy of the Bayesian inference method.

Additionally, one previously published retinotopy dataset for with a wide field of view (48º of eccentricity) was re-analyzed [1] (their Fig. 3). The subject for this dataset was also included in the newly acquired data, S1201. The wide-field-of-view dataset was used as a further validation set for models derived from the newly acquired data for S1201, as it enabled us to test the accuracy of model predictions in the far periphery from models derived from data with limited eccentricity.

FMRI processing.

Spatial distortions due to inhomogeneities in the magnetic field were corrected using in-house software from NYU's Center for Brain Imaging (http://cbi.nyu.edu/software). The data were then motion-corrected by co-registering all volumes of all scans to the first volume of the first scan in the session. The fMRI slices were co-registered to the whole brain T1-weighted anatomy, and the time series resampled via trilinear interpolation to the 1 mm^3 voxels within the cortical ribbon (gray matter). Finally, the time series were averaged for each voxel across all scans with the same stimulus within a given dataset.

PRF solutions.

Retinotopic maps were produced by solving population receptive field (pRF) models for each voxel using Vistasoft, as described previously [12]. pRF models were solved using a 2-stage coarse-to-fine approach on the time series in the 1 mm^3 gray matter voxels. The first stage of the model fit was a grid fit, solved on time series that were temporally decimated (2×), spatially blurred on the cortical surface

using a discrete heat kernel (approximately equal to a Gaussian kernel of 5 mm width at half height), and subsampled by a factor of 2. The decimation and blurring helps to find an approximate solution that is robust to local minima. The parameters obtained from the grid fit were interpolated to all gray matter voxels and used as seeds for the subsequent nonlinear optimization. Finally, the pRF parameters were projected from the volume to the cortical surface vertices for white, mid-gray, and pial surfaces using nearest-neighbor interpolation; values were then averaged across the three layers using a weighted-mean in which the fraction of BOLD signal variance explained by the pRF model was used as a weight. All vertices with a pRF variance explained fraction less than 0.1 were ignored.

Models of Retinotopy.

In previous work [25], group-average retinotopic maps were registered to a model of V1, V2, and V3 retinotopy described by Schira et al. [32]. This "banded double-sech" model accurately describes the anisotropic magnification of the visual field on the cortical surface, particularly near the fovea. However, we have observed, particularly in individual data, that retinotopic data from outside the V1-V3 region described by the Schira model has a large impact on the quality of the inferred map. Accordingly, we constructed a new model that includes 9 additional visually active regions: hV4, VO1, VO2, V3a, V3b, LO1, LO2, TO1, and TO2. This model employed a new method of constructing 2D models of retinotopy that was specifically designed to accommodate the distortions caused by anatomical alignment, inflation, and flattening of the cortical surface. The new method is much simpler to extend to many more visual field maps, as it does not rely on an analytic description of the flattened (2D) retinotopic maps, which are only available for V1-V3. Rather, it requires as input a cortical map image on which estimates of the visual area boundaries have been drawn manually and labeled as either a foveal boundary, a peripheral boundary, or an upper-, horizontal- or lower-visual-field meridian (i.e., the visual-field extrema are labeled). These boundaries need only be drawn once for a single group-average

retinotopic map. The boundaries are projected onto the cortical surface, and the retinotopic coordinates for each vertex on the surface are deduced via a minimization procedure. This minimization is motivated by two principles: (1) the retinotopic fields (polar angle and eccentricity) should be as orthogonal to each other as possible and (2) the retinotopic fields should be as smooth as possible. To this end, the minimization routine simultaneously maximizes both the smoothness of the retinotopic fields between vertices connected by edges in the mesh as well as the overall orthogonality between the polar angle field and the eccentricity field. The hand-drawn boundary values are held constant during the minimization. Equation 2 gives the function *f* that is minimized, where $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ represent vectors of the polar angle and eccentricity values, respectively; **E** represents the set of edges between vertices in the mesh; and $\boldsymbol{x}$ represents the matrix of vertex coordinates (i.e., $\boldsymbol{x}_u$ represents the coordinate vector of the vertex u). During minimization the values of the polar angle and eccentricity vectors are scaled such that both fields ranged from -1 to 1 (e.g., polar angle boundary values of 0° and 180° were assigned values of -1 and 1 respectively) so that the fields could be evaluated easily for orthogonality; after the minimization, the polar angle was linearly rescaled back to the range 0°-180° while eccentricity was rescaled so as to have an exponential distribution that best fit the group-average. We employed this model generation routine using boundaries drawn over the group-average data (see Anatomically-defined Atlas of Retinotopy: Group Average, below) as well as the Wang et al. [23] atlas and the Hinds et al. [53] V1 boundary as rough guidelines.

$$f(\boldsymbol{\theta}, \boldsymbol{\rho}) = (\boldsymbol{\theta} \cdot \boldsymbol{\rho})^2 + \sum_{(u,v)\in\mathbf{E}} \frac{(\theta_u - \theta_v)^2 + (\rho_u - \rho_v)^2}{2\|(\boldsymbol{x})_u - (\boldsymbol{x})_v\|} \tag{2}$$

A full description of the model, including how it can be projected onto an *fsaverage* spherical surface or an individual subject's *fsaverage*-aligned spherical surface, is provided in the open source Neuropythy library (https://github.com/noahbenson/neuropythy).

Anatomically-defined Atlas of Retinotopy.

Construction of the anatomically-defined atlas of retinotopy is summarized in Figure S2. Previous work employed a mass-spring-damper system combined with a nonlinear gradient-descent minimization in order to register group-average retinotopic data, averaged on FreeSurfer's *fsaverage_sym* hemisphere [54], with a model of V1, V2, and V3 retinotopy [32]. In this paper, we modify this technique slightly to bring it more in line with previous established methods such as those used by FreeSurfer for surface-based anatomical registration [33, 34]. In brief, retinotopy is measured in a group of subjects via fMRI; the subjects' cortical meshes are aligned to the *fsaverage* surface via FreeSurfer's surface registration; the retinotpoic coordinates are then averaged across subjects at each vertex on a single atlas of the cortical surface; a 2D atlas of retinotopy is then placed on this cortical surface; and finally, the cortical surface is warped to match the retinotopic atlas as best as possible given constraints on the warping. Each of these steps is described in more detail below.

Group-average Data. Group-average retinotopic maps (Fig. S2*B*) were obtained from 181 subjects whose data were published and made freely available as part of the Human Connectome Project [39]. The resulting group-average retinotopic maps are shown in Figure S2B.

Cortical Map Projection. The cortical surfaces of the *fsaverage* left and right hemispheres, on which the group-average data were constructed, were inflated both to a smooth hemisphere (FreeSurfer's "inflated" surface) as well as to a sphere (FreeSurfer's "sphere" surface); the vertices on the spherical surfaces were then flattened to 2D maps using an orthographic map projection. Precise parameters of this projection and the source code used to generate it are included in the supplemental materials. We refer to the 2D vertex coordinates in this resulting map as the "initial vertex coordinates" because they precede the warping of the vertex coordinates that occurs during registration.

Registration. The initial vertex coordinates of the map projections described above were warped in order to bring the polar angle and eccentricity measurements of the vertices into alignment with the 2D model's predictions of retinotopy while maintaining topological constraints: i.e., preventing triangles in the triangle mesh representing the 2D cortical map from inverting and penalizing excessive stretching or compression of the map. This process was achieved by minimizing a potential function defined in terms of the edges of the triangle mesh, the angles of the triangle mesh, and the positions of the vertices with polar angle and eccentricity measurements above the weight threshold (see *Group-Average Data*, above). Equation 3 gives this potential function, $F(\boldsymbol{x})$, which is further broken down into four components detailed in Table 2. Fundamentally, the potential function F is a sum of two kinds of penalties: penalties for deviations from the reference mesh and penalties for mismatches between the vertices with retinotopic coordinates and their positions in the retinotopic model. In the case of the former, the reference mesh is given by the parameters $\boldsymbol{x}_0$, $\mathbf{E}$, $\boldsymbol{\Theta}$, $\mathbf{P}$, and $\boldsymbol{\Phi}$, and the potential of the deviations are defined by F_e, F_ϑ, and F_p. The latter is described by F_φ. In these functions, $\boldsymbol{x}$ represents the $n \times 2$ matrix of the 2D-coordinates of each vertex while $\boldsymbol{x}_0$ represents the same coordinates in the reference mesh; $\mathbf{E}$ represents the set of undirected edges (represented as (u, v) pairs such that (u, v) and (v, u) are not both in $\mathbf{E}$) in the reference mesh; $\boldsymbol{\Theta}$ represents the set of angle triples (a, b, c) such that the angle is between edge (a, b) and edge (a, c); $\mathbf{P}$ is the set of vertices that lie on the perimeter of the 2D map projection; $r_{\boldsymbol{x}}(u,v)$ is the Euclidean distance between vectors $(\boldsymbol{x})_u$ and $(\boldsymbol{x})_v$; and $\alpha_{\mathbf{x}}(a,b,c)$ is the counter-clockwise angle between vectors $((\boldsymbol{x})_b - (\boldsymbol{x})_a)$ and $((\boldsymbol{x})_c - (\boldsymbol{x})_a)$; $\boldsymbol{\Phi}$ represents the set of anchors defined by the retinotopic model in which each anchor is represented by a tuple $(u,\mathbf{y},\sigma,w)$ where w is the weight of the anchor, u is the vertex drawn to the anchor, σ is the standard deviation of the anchor's Gaussian potential well, and $\boldsymbol{y}$ is the 2D point to which the anchor is attached; the constants q_0 and q_1 are the minimum and maximum allowable edge lengths, respectively.

$$F(\boldsymbol{x}; \boldsymbol{x_0}, \mathbf{E}, \boldsymbol{\Theta}, \mathbf{P}, \boldsymbol{\Phi}) = F_e(\boldsymbol{x}; \boldsymbol{x_0}, \mathbf{E}) + F_\vartheta(\boldsymbol{x}; \boldsymbol{x_0}, \boldsymbol{\Theta}) + F_p(\boldsymbol{x}; \boldsymbol{x_0}, \mathbf{P}) + F_\varphi(\boldsymbol{x}; \boldsymbol{\Phi}) \tag{3}$$

The term of the potential function devoted to the retinotopic model is given in F_φ (Eq. 3; Tab. 2). This potential term is a set of inverted-Gaussian potential wells called anchors. Each anchor represents the attraction of a single vertex u, with measured polar angle θ, eccentricity ρ, and weight w, to a 2D point $\mathbf{y}$, at which the retinotopic model predicts a polar angle value of θ and an eccentricity value of ρ. Note that each visual area represents every point (θ, ρ) in the visual field, there are multiple anchors per vertex with retinotopic data. In fact, the retinotopic model used in this paper defines 9 maps in addition to the V1-V3 maps (see Model of Retinotopy, above), bringing the total number of anchors per retinotopic vertex to 12. The additional areas are intended partly to prevent vertices immediately outside of V1-V3 from being drawn incorrectly into the V1-V3 section of the model and are not analyzed in detail in this paper. Each anchor additionally defines a parameter σ; this value is the width (standard deviation) of the anchor's Gaussian potential well; σ is defined as the minimum distance from the given anchor to any other anchor to which u is also attracted; this value was given a maximum value of 20ε where ε is the mean edge-length in the projected map.

The potential function was minimized using a gradient descent algorithm sensitive to the singularities in the terms G_e, and G_ϑ (Tab. 2); whenever the singularity is accidentally crossed, the minimizer backtracks and chooses a smaller step-size. This approach prevents the inversion (from counter-clockwise ordering to clockwise ordering) of any triangle in the mesh, as such an inversion would require the minimization trajectory to pass through a singularity at the point where $\alpha = 0$ or $\alpha = \pi$. The source code used to minimize the potential function as well as specifications of the gradients of each term is provided in the open-source library included with the Neuropythy and Neurotica libraries (https://github.com/noahbenson/nben).

Minimization was run for at least 2500 steps in which the step-size was constrained such that the displacement of each vertex in each step was at most 1/50th of the average edge-length in the map projection. A small amount of exponentially distributed random noise was added to the gradient at each step with the constraint that the gradient direction at each vertex be conserved; this noise did not affect the minimum obtained by the search but did speed up convergence significantly (see associated libraries for further details). Convergence was generally observed within 1000-2000 steps. The set of vertex coordinates that resulted from this minimization brings the retinotopic measurements associated with the vertices in V1, V2, and V3 referred to as the *registered vertex coordinates*.

Prediction. The registered vertex coordinates, once obtained, give the alignment of the subject's cortical surface to the model of retinotopy; accordingly, a prediction of any vertex's associated pRF and visual area label can be derived by comparing the the vertex's registered coordinates with the model. Every vertex whose registered coordinates fall within the model's V1 boundary, for example, is labeled as part of V1. Because only the vertex coordinates, and not the vertex identities, are changed during the registration process, there is no need to invert the registration: visual area label, polar angle, and eccentricity values assigned to each vertex apply as readily to the vertices whether they are visualized in the registered vertex coordinates or in the coordinates that define the subject's white-matter surface, for example. The retinotopic map predictions for the group-average data is shown in Figure S2D (left column).

Because the group-average retinotopic data were used in the registration, the predicted map that results provides a reasonable estimate of any subject's expected retinotopic map, as shown previously [24, 25], although the predicted map does not account for further individual differences in the structure to function relationship, as we show in this paper. Additionally, because the predicted map from the group-averaged data is defined on the *fsaverage* subject's cortical surfaces, a retinotopic map prediction

for any new subject, for whom retinotopic mapping measurements may not be available, can be easily obtained: one can use FreeSurfer to align the new subject's cortical surface with the *fsaverage* subject's surface (anatomical structure alignment) then to project the retinotopic maps from the *fsaverage* subject to the new subject based on the anatomical similarity between them. Because of this, we refer to this group-average retinotopic prediction as the *anatomically-defined atlas of retinotopy.* This atlas is used as *the prior* for the Bayesian model fit, described below. The atlas is similar but not identical to one presented previously [25].

Bayesian Retinotopic Maps.

The anatomically-defined atlas of retinotopy, while providing a good prediction for most subjects' individual retinotopic maps, nonetheless does not account for individual differences in the mapping between anatomical location and retinotopic coordinates. Accordingly, predicted retinotopic maps for individual subjects were refined starting from the anatomically-defined atlas of retinotopy using a similar method as was used to generate the atlas originally; this process is detailed in Figure 3C. For each subject, their cortical surface was aligned based on anatomical structure to the *fsaverage* subject's cortical surface using FreeSurfer, then their retinotopic data were projected to a map using the identical map projection described above in the section on the anatomically-defined atlas of retinotopy. Note that, in this case, the anatomical alignment to the *fsaverage* subject serves to make the map projections as similar as possible between individual subjects and the anatomically-defined atlas of retinotopy. If we were not interested in incorporating information obtained from the anatomically-defined atlas of retinotopy (which represents a prior belief of retinotopic organization based on group-average data), this step would not be necessary.

The individual subject's projected map is then arranged according to the registered vertex coordinates from the anatomically-defined atlas of retinotopy; this step reflects the prior belief that the

group-average registration to the retinotopy model is generally accurate for an individual subject when that subject's anatomical structure has been aligned to the *fsaverage* subject's. Critically, none of the steps taken so far in processing the individual subject's data relies on any measurements of retinotopy that might be associated with that subject. Rather, these steps have relied only on anatomical structure. If, for a subject, no retinotopic measurements are made, then there is no data with which to modify this prior belief; accordingly, the prediction of retinotopy for that subject would be identical to the prediction of retinotopy contained in the anatomically-defined atlas. In other words, without observation, the prior remains the prediction.

The next step registers the individually measured retinotopy data to the anatomically-defined atlas. Before registration, the individual subject's data is resampled onto a uniform triangular mesh, and each vertex whose retinotopic measurements are above threshold are given a weight, w, based on the variance explained, ω, of its pRF model solution. The mesh is resampled to the same uniform triangle mesh used as the initial vertex coordinates in the registration of the anatomically-defined atlas of retinotopy in order to speed up registration. Triangles that are tightly pinched (i.e., triangles with internal angles near 0 or π) can drastically slow the registration progress by forcing the minimizer to frequently backtrack steps; resampling makes such behavior much less likely during the initial minimization. Aside from the weight, other parameters tracked by the potential field, including anchors parameters used by the function $F_{\varphi}(\boldsymbol{x})$, are obtained identically as with the anatomically-defined atlas of retinotopy. These anchors inherit the weight of the vertex to which they apply, but are reduced when the field sign of the triangles adjacent to the vertex does not match the field sign of the visual area to which it is tied by the anchor or when the pRF size predicted by the model does not match that of the vertex's measured pRF. Details regarding the weights on anchors are provided in the neuropythy library.

For each training dataset of each subject, minimization was run for 2500 steps using the same protocol that was used with the anatomically-defined atlas of retinotopy. Retinotopic map prediction, based on the positions of the registered vertex coordinates in the retinotopy model, were also computed identically to those in the anatomically-defined atlas. Identical minimization and prediction methods were run for each test dataset as well, but these results were not used to measure the accuracy or effectiveness of the prediction methods.

Cortical Magnification.

Cortical magnification was calculated using both the observed retinotopic maps and the inferred maps that were produced by combining each subject's full retinotopy dataset with the retinotopic prior. This combination of data should, in theory, produce the highest-quality retinotopic map predictions of which we are capable (see Results & Discussion). Cortical magnification was calculated by first projecting all vertices in a single visual area (such as V1) into the visual field based on their pRF centers. The cortical magnification of a particular polar angle and eccentricity is then the total white vertex surface-area (as calculated by FreeSurfer) of all pRF centers within a disk of some radius α, divided by the area of the disk ($\pi\alpha^2$). For an eccentricity ρ, we used a radius $\alpha = \rho / 3.$ on these maps in two ways.

Acknowledgements

The research was supported by NIH grants R00-EY022116 and R01-MH111417. The authors declare no competing financial interests.

References

1. Wandell, B.A. and J. Winawer, Imaging retinotopic maps in the human brain. Vision research, 2011. **51**(7): p. 718-37.

2. Sereno, M.I. and R.B. Tootell, From monkeys to humans: what do we now know about brain homologies? Curr Opin Neurobiol, 2005. **15**(2): p. 135-44.

3. Wandell, B.A. and S.M. Smirnakis, Plasticity and stability of visual field maps in adult primary visual cortex. Nat Rev Neurosci, 2009. **10**(12): p. 873-84.

4. Conner, I.P., et al., Retinotopic organization in children measured with fMRI. J Vis, 2004. **4**(6): p. 509-23.

5. Huk, A.C., D. Ress, and D.J. Heeger, Neuronal basis of the motion aftereffect reconsidered. Neuron, 2001. **32**(1): p. 161-72.

6. Engel, S., X. Zhang, and B. Wandell, Colour tuning in human visual cortex measured with functional magnetic resonance imaging. Nature, 1997. **388**(6637): p. 68-71.

7. Grill-Spector, K., et al., A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum Brain Mapp, 1998. **6**(4): p. 316-28.

8. Martinez, A., et al., Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat Neurosci, 1999. **2**(4): p. 364-9.

9. Duncan, R.O. and G.M. Boynton, Cortical magnification within human primary visual cortex correlates with acuity thresholds. Neuron, 2003. **38**(4): p. 659-71.

10. Sereno, M.I., et al., Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science, 1995. **268**(5212): p. 889-93.

11. Engel, S.A., G.H. Glover, and B.A. Wandell, Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cerebral Cortex, 1997. **7**(2): p. 181-92.

12. Dumoulin, S.O. and B.A. Wandell, Population receptive field estimates in human visual cortex. Neuroimage, 2008. **39**(2): p. 647-60.

13. Dukart, J. and A. Bertolino, When structure affects function--the need for partial volume effect correction in functional and resting state magnetic resonance imaging studies. PLoS One, 2014. **9**(12): p. e114227.

14. Winawer, J., et al., Mapping hV4 and ventral occipital cortex: the venous eclipse. Journal of vision, 2010. **10**(5): p. 1.

15. Binda, P., et al., Minimizing biases in estimating the reorganization of human visual areas with BOLD retinotopic mapping. J Vis, 2013. **13**(7): p. 13.

16. Silver, M.A., D. Ress, and D.J. Heeger, Topographic maps of visual spatial attention in human parietal cortex. J Neurophysiol, 2005. **94**(2): p. 1358-71.

17. Larsson, J. and D.J. Heeger, Two retinotopic visual areas in human lateral occipital cortex. J Neurosci, 2006. **26**(51): p. 13128-42.

18. Swisher, J.D., et al., Visual topography of human intraparietal sulcus. J Neurosci, 2007. **27**(20): p. 5326-37.

19. Arcaro, M.J., et al., Retinotopic organization of human ventral visual cortex. J Neurosci, 2009. **29**(34): p. 10638-52.

20. Mackey, W.E., J. Winawer, and C.E. Curtis, Visual field map clusters in human frontoparietal cortex. Elife, 2017. **6**.

21. Schira, M.M., et al., The foveal confluence in human visual cortex. J Neurosci, 2009. **29**(28): p. 9050-8.

22. Malach, R., I. Levy, and U. Hasson, The topography of high-order human object areas. Trends Cogn Sci, 2002. **6**(4): p. 176-184.

23. Wang, L., et al., Probabilistic Maps of Visual Topography in Human Cortex. Cereb Cortex, 2015. **25**(10): p. 3911-31.

24. Benson, N.C., et al., The retinotopic organization of striate cortex is well predicted by surface topology. Curr Biol, 2012. **22**(21): p. 2081-5.

25. Benson, N.C., et al., Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy. PLoS Comput Biol, 2014. **10**(3): p. e1003538.

26. Van Essen, D.C., et al., Mapping visual cortex in monkeys and humans using surface-based atlases. Vision Res, 2001. **41**(10-11): p. 1359-78.

27. Dougherty, R.F., et al., Visual field representations and locations of visual areas V1/2/3 in human visual cortex. J Vis, 2003. **3**(10): p. 586-98.

28. Harvey, B.M. and S.O. Dumoulin, The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. The Journal of neuroscience : the official journal of the Society for Neuroscience, 2011. **31**(38): p. 13604-12.

29. Van Essen, D.C., A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature, 1997. **385**(6614): p. 313-8.

30. Stensaas, S.S., D.K. Eddington, and W.H. Dobelle, The topography and variability of the primary visual cortex in man. J Neurosurg, 1974. **40**(6): p. 747-55.

31. Andrews, T.J., S.D. Halpern, and D. Purves, Correlated size variations in human visual cortex, lateral geniculate nucleus, and optic tract. J Neurosci, 1997. **17**(8): p. 2859-68.

32. Schira, M.M., et al., Modeling magnification and anisotropy in the primate foveal confluence. PLoS Comput Biol, 2010. **6**(1): p. e1000651.

33. Dale, A.M., B. Fischl, and M.I. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage, 1999. **9**(2): p. 179-94.

34. Fischl, B., M.I. Sereno, and A.M. Dale, Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. NeuroImage, 1999. **9**(2): p. 195-207.

35. Frost, M.A. and R. Goebel, Functionally informed cortex based alignment: an integrated approach for whole-cortex macro-anatomical and ROI-based functional alignment. Neuroimage, 2013. **83**: p. 1002-10.

36. Haxby, J.V., et al., A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron, 2011. **72**(2): p. 404-16.

37. Ugurbil, K., et al., Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. NeuroImage, 2013. **80**: p. 80-104.

38. Van Essen, D.C., et al., The WU-Minn Human Connectome Project: an overview. Neuroimage, 2013. **80**: p. 62-79.

39. Benson, N., et al., The HCP 7T Retinotopy Dataset. bioRxiv, 2018.

40. Press, W.A., et al., Visual areas and spatial summation in human visual cortex. Vision Res, 2001. **41**(10-11): p. 1321-32.

41. Gardner, J.L., Is cortical vasculature functionally organized? Neuroimage, 2010. **49**(3): p. 1953-6.

42. Boubela, R.N., et al., fMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. Sci Rep, 2015. **5**: p. 10499.

43. Huettel, S.A., A.W. Song, and G. McCarthy, *Functional magnetic resonance imaging*. Third edition. ed. 1 volume (various pagings).

44. Shmuel, A., et al., Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. Neuron, 2002. **36**(6): p. 1195-210.

45. Kay, K.N., et al., Compressive spatial summation in human visual cortex. J Neurophysiol, 2013. **110**(2): p. 481-94.

46. Horton, J.C. and W.F. Hoyt, The representation of the visual field in human striate cortex. A revision of the classic Holmes map. Arch Ophthalmol, 1991. **109**(6): p. 816-24.

47. Van Essen, D.C., et al., Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. Proc Natl Acad Sci U S A, 1998. **95**(3): p. 788-95.

48. Large, I., et al., Individual Differences in the Alignment of Structural and Functional Markers of the V5/MT Complex in Primates. Cereb Cortex, 2016. **26**(10): p. 3928-3944.

49. Andersson, J.L., S. Skare, and J. Ashburner, How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage, 2003. **20**(2): p. 870-88.

50. Smith, S.M., et al., Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage, 2004. **23 Suppl 1**: p. S208-19.

51. Fischl, B., et al., High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp, 1999. **8**(4): p. 272-84.

52. Fischl, B. and A.M. Dale, Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A, 2000. **97**(20): p. 11050-5.

53. Hinds, O.P., et al., Accurate prediction of V1 location from cortical folds in a surface coordinate system. Neuroimage, 2008. **39**(4): p. 1585-99.

54. Greve, D.N., et al., A surface-based analysis of language lateralization and cortical asymmetry. J Cogn Neurosci, 2013. **25**(9): p. 1477-92.

Supplemental Material

Raw MRI data as well as analyses and source code for reproducing figures and performing additional analyses can be found on the Open Science Foundation website https://osf.io/knb5g/

Performing Bayesian inference using your own retinotopic maps.

To perform Bayesian inference on a FreeSurfer subject, one can use the neuropythy Python library (https://github.com/noahbenson/neuropythy). For convenience, this library has also been packaged into a Docker container that is freely available on Docker hub (https://hub.docker.com/r/nben/neuropythy). The following command will provide an explanation of how to use the Docker:

```
> docker run -it --rm nben/neuropythy:v0.5.0 register_retinotopy --help
```

Detailed instructions on how to use the tools documented in this paper are included in the Open Science Foundation website mentioned above.

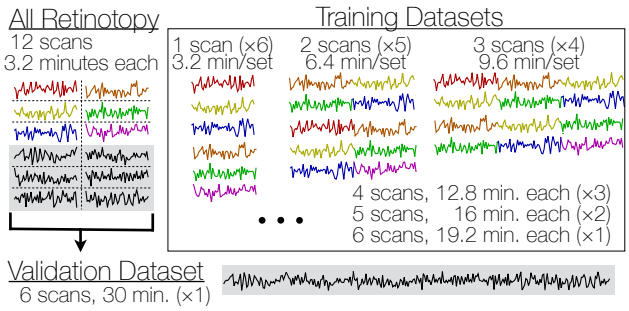

Figure S1. Cross-validation schema. To evaluate the accuracy of the predictions of retinotopic maps, we employ a cross-validation schema. Each subject's 12 retinotopic mapping scans were divided into one large set of validation data as well as 21 smaller sets of training data. An additional dataset of all 12 scans was used for analysis of retinotopic properties not linked to evaluation of the quality of the predicted maps.

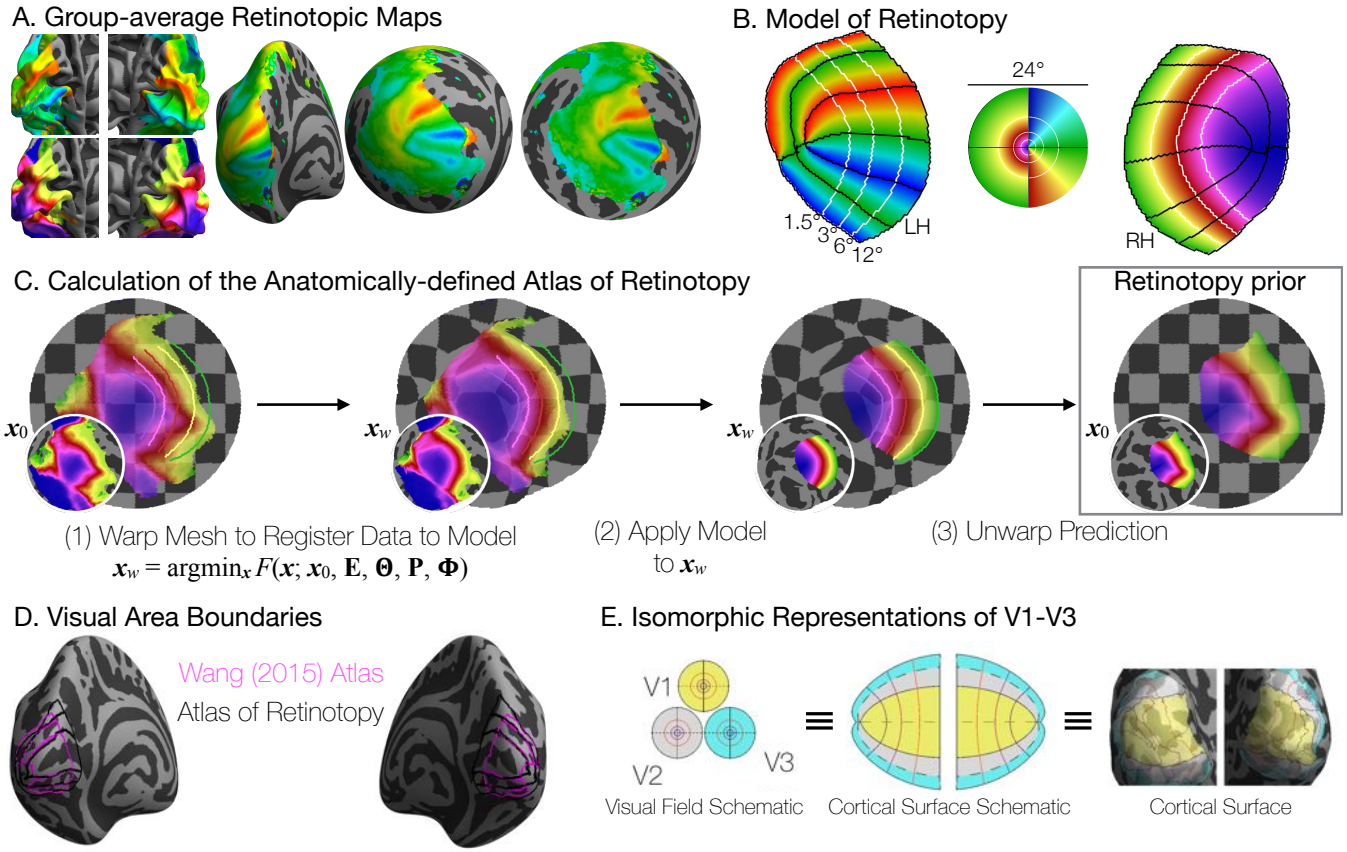

Figure S2. Deriving the anatomically-defined atlas of retinotopy (the prior). ***A,*** The group-average polar angle (top) and eccentricity (bottom) maps. The cortical surface is inflated to a sphere then flattened to a map. ***B,*** The model of retinotopy shown with polar angle plotted on the left and eccentricity plotted on the right hemispheres. ***C,*** The retinotopic prior is constructed from the group-average data using an updated version of the method described by Benson *et al.* (2014). Note that while only eccentricity is shown, polar angle and eccentricity are registered simultaneously. The checkerboard underlay illustrates the anatomical warping. ***D,*** There is approximate agreement between the boundaries of visual areas V1, V2, and V3 as defined by two atlases. The Wang et al. maximum probability atlas (2015) and the retinotopic template defined here have similar boundaries. The template extends from 0° to 90° eccentricity, whereas the Wang et al atlas is limited to the field of view of their experiments (14°), hence the template maps are larger. ***E,*** Because there is a topological isomorphism between the cortical surface, the left and right hemifields, and the model of retinotopy, the three representations have exact one-to-one correspondences.

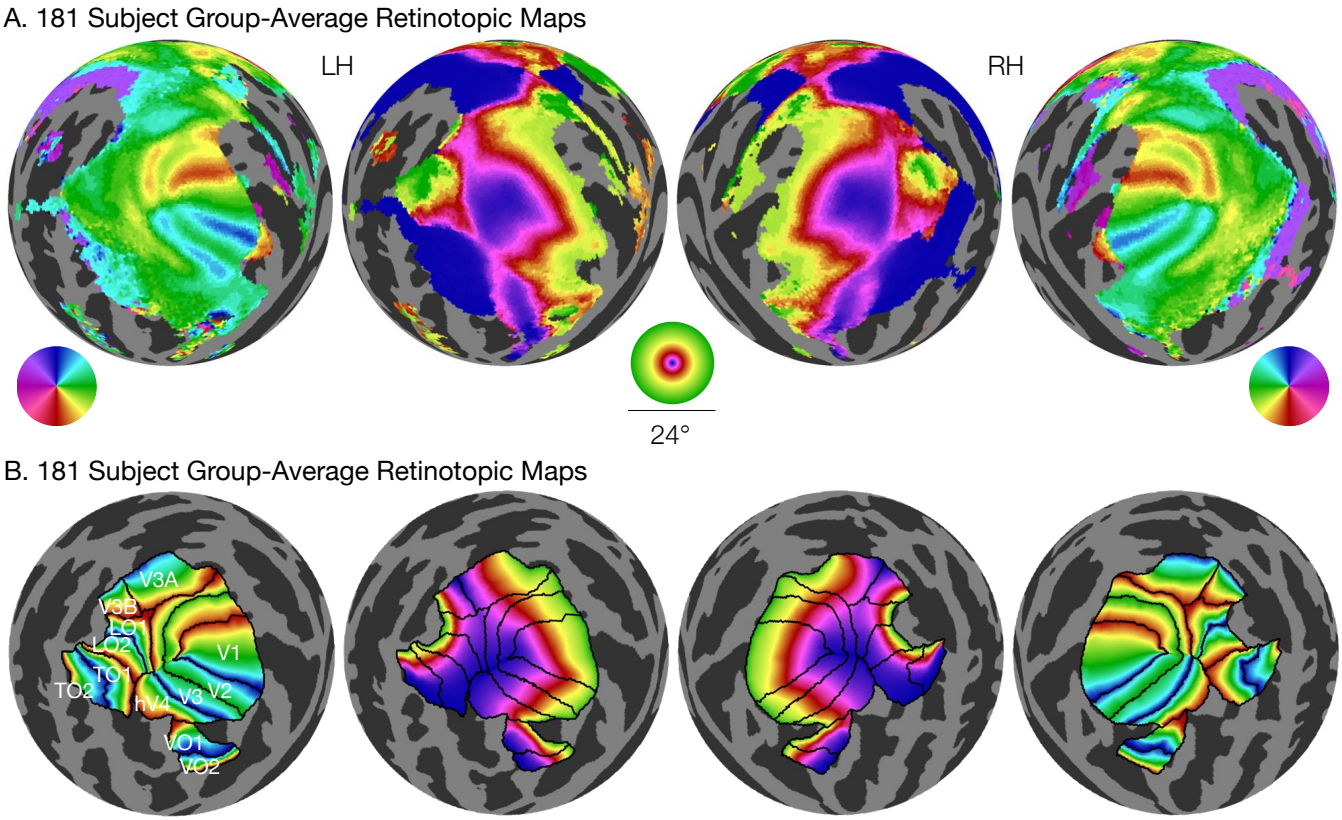

Figure S3. The retinotopic prior. ***A***, The 181-subject group-average retinotopic maps from the Human Connectome Project 7T Retinotopy Dataset are shown. These maps were used to construct the prior. B, The retinotopic prior is shown from 0-12° of eccentricity with boundary lines between areas. All 12 retinotopic areas included in the prior are shown.