

Thickness of deep layers in the fusiform face area predicts face recognition

Abbreviated title: Cortical layer thickness predicts behavior

Authors: Rankin W. McGugin^{1*}, Allen T. Newton^{2,3,4}, Benjamin Tamber-Rosenau⁵, Andrew Tomarken¹, Isabel Gauthier¹

Affiliations:

¹Department of Psychology, Vanderbilt University, Wilson Hall, PMB 407817, 2301 Vanderbilt Place, Nashville, TN 37240-7817, USA

²Vanderbilt Medical Center, Department of Radiology and Radiological Sciences

³Vanderbilt University Institute of Imaging Science

⁴Monroe Carell Jr. Children's Hospital at Vanderbilt

⁵Department of Psychology, University of Houston, 126 Heyne Building, 3695 Cullen Blvd, - Houston, TX 77204, USA

*Correspondence to: Rankin.McGugin@gmail.com

Number of pages: 19

Number of figures: 4

Number of tables: 4

Number of words for Abstract: 250

Number of words for Introduction: 648

Number of words for Discussion: 1449

Conflict of Interest: none

Acknowledgments: We thank Carla Shatz for suggesting that different ages of acquisition should translate into differential effects on cortical layers. We thank Magen Speegle and Susan Benear for manual tracing on the grey-matter/white-matter boundaries, Anita Disney, Jesse Gomez, Thilo Womelsdorf and Wolfe Zinke for comments, Maarten Versluis and Ed Mojahed for sharing pulse sequence designs which were used as a basis for those developed here. R.W.M. and I.G. conceived the study. R.W.M. and A.T.N. collected the primary data. A.T.N. developed the image acquisition techniques and average susceptibility-weighted image reconstruction techniques. R.W.M., A.T.N. and B.T.R. developed the methods for laminar analysis of susceptibility-weighted images. A.T. conducted reliability, bootstrapping, and Bayesian analyses. All authors coordinated second level analyses. R.W.M., A.T.N. and I.G. wrote the paper with input from the other authors. This work was supported by the NSF (SBE-0542013 and SMA-1640681) and the Vanderbilt Vision Research Center (P30-EY008126) and the Vanderbilt University Medical Center (1 S10RR023047 01).

Abstract:

People with superior face recognition have relatively thin cortex in face-selective brain areas, while those with superior vehicle recognition have relatively thick cortex in the same areas. We suggest that these opposite correlations reflect distinct mechanisms influencing cortical thickness (CT) as abilities are acquired at different points in development. We explore a new prediction regarding the specificity of these effects through the depth of the cortex: that face recognition selectively and negatively correlates with thickness of the deepest laminar subdivision in face-selective areas. With ultra-high resolution MRI at 7T, we estimated the thickness of three laminar subdivisions, which we term MR layers, in the right fusiform face area (rFFA) in 14 adult male humans. Face recognition was negatively associated with the thickness of deep MR layers, while vehicle recognition was positively related to the thickness of all layers. Regression model comparisons provided overwhelming support for a model specifying that the magnitude of the association between face recognition and CT differs across MR layers (deep vs. superficial/middle) while the magnitude of the association between vehicle recognition and CT is invariant across layers. The total CT of rFFA accounted for 69% of the variance in face recognition, and thickness of the deep layer alone accounted for 84% of this variance. Our findings demonstrate the functional validity of MR laminar estimates in FFA. Studying the structural basis of individual differences for multiple abilities in the same cortical area can reveal effects of distinct mechanisms that are not apparent when studying average variation or development.

Significance Statement:

Face and object recognition vary in the normal population and are only modestly related to each other. The recognition of faces and vehicles are both positively related to neural responses in the fusiform face area (FFA), but show different relations to the cortical thickness of FFA. Here, we use very high-resolution MRI, and find that face recognition ability (a skill acquired early in life) is negatively correlated with thickness of FFA's deepest MR-defined layers, whereas recognition of vehicles (a skill acquired later in life) is positively related to thickness at of all cortical layers. Our methods can be used in the future to characterize sources of variability in human abilities and relate them to distinct mechanisms of neural plasticity.

Introduction

Brain volumetrics are behaviorally relevant, task-independent biomarkers that can be studied in terms of their relation to visual abilities. Those abilities depend on both genetic and environmental factors, as well as on the developmental state of the brain. Importantly, many structural biomarkers can be measured independent of behavior, and are therefore more likely to reflect stable variability across individuals rather than performance during a specific task. The relation between brain structure and behavioral abilities can provide a useful window into the underlying forces that culminate into individual differences.

The current work builds on our previous finding that differences in total cortical thickness (CT) of *the same brain area in the same people* predict different abilities in opposite directions: Vehicle recognition (VR) was associated with a relatively thicker fusiform face area (FFA), whereas face recognition (FR) was associated with a relatively thinner FFA ((McGugin, Van Gulick, & Gauthier, 2016). The current study extends that previous work, using ultra-high resolution imaging to probe laminar thickness differences underlying the differences in total cortical thickness, and their relationship to behavioral abilities.

When we reported the surprising pattern of FFA CT variations across individuals, we suggested that those opposing results (thinner CT associated with better FR, thicker CT associated with better VR, all within the same region of the brain) could be explained by differences in the developmental state (plasticity) of the brain at the time these visual skills were learned (McGugin et al., 2016). This was predicted since individuation for exemplars in the categories used, faces and vehicles, were likely learned at different times in development. It seemed plausible that learning occurring at different times in development would have different effects on underlying brain structure, with some of these effects being selective to specific cortical layers. In other words, our predictions are based on intuitions relevant to two different fields: the development of behavioral recognition abilities across domains and the development of the associated cortical underpinnings. These are two areas of study that are difficult to study, but we based our conjectures on the following findings.

First, we predicted that observed differences in brain correlates of FR and VR in adults may reflect differential ages of acquisition for these abilities (McGugin et al., 2016). There is growing evidence that faces dominate the visual diet of infants. Three month-old babies see faces for 25% of their waking time, an amount reduced to 10% by 12 months (Fausey, Jayaraman, & Smith, 2016; Sugden & Moulson, 2018), and early face exposure has been found to be critical to the development of face-selectivity (Arcaro, Schade, Vincent, Ponce, & Livingstone, 2017). The age at which infants learn to individuate objects such as vehicles is less clear (Clerkin, Hart, Rehg, Yu, & Smith, 2016). Although face selectivity continues to develop during adolescence (Gomez et al., 2017; Scherf, Thomas, Doyle, & Behrmann, 2014) and FR continues to improve in adults (Germine, Duchaine, & Nakayama, 2011), it is reasonable to assume that FR also starts earlier than VR.

Second, we conjectured that if the earlier acquisition of FR (relative to VR) was responsible for a pattern where thinner cortex related to better performance, then the underlying structural biomarkers for the thinner cortex may be related to the early developmental pruning of connections critical to the acquisition of FR. The amygdala is critical to the development of FR (Schultz, 2005) and is connected to the occipital lobe via the inferior longitudinal fasciculus (ILF). Research on the connectivity between the inferior temporal (IT) areas TE and TEO (the monkey analog of the FFA) and the limbic system has revealed projections in the infant monkey

that are pruned in the adult (Webster, Ungerleider, & Bachevalier, 1991). Such connections from IT to limbic areas are labeled predominantly in infragranular (deep) layers of IT (Stefanacci & Amaral, 2000). Accordingly, if this variability in the pruning of the ILF connections from IT to the amygdala is relevant to FR in humans, it should be most reflected in the thickness of deep cortical layers. In contrast, we expect the positive correlation between VR and variability in FFA's CT to reflect mechanisms that support learning-dependent structural plasticity in adults (Wenger et al., 2012). These mechanisms are not well understood but are proposed to include processes such as gliogenesis and angiogenesis that accompany synaptogenesis (Zatorre, Fields, & Johansen-berg, 2012). We have no reason to expect these mechanisms to be confined to deep layers (but acknowledge no evidence against laminar specificity).

It is important to note that multiple mechanisms may underlie any observed differences in structure that are associated with development. For example, variation in the pruning of connections could occur as described in the predictions that motivated the current work, affecting the thickness of different layers. However, the predicted pattern of laminar-specific correlations could be consistent with pruning but could also reflect other mechanisms. In particular, myelination increases during development and the effects of pruning and myelination are difficult to distinguish, as myelination could shift the apparent gray-white matter boundary and result in thinner cortical measurements, an effect that would be attributed to deep cortical layers ((Sowell et al., 2004). It is fortuitous that both of these plausible mechanisms predict that the earlier development of FR would result in variance across individuals for FR to be specifically related to the thickness of the deep layers in our measurements.

In summary, here we test whether changes in CT that are associated with behavior (McGugin et al., 2016) are predominantly driven by changes in deep layers of the cortex, suggestive of developmental underpinnings. We additionally demonstrate that ultra-high resolution MRI can be used to image laminar thickness within the cortex, providing a potentially useful biomarker for probing the structure-function relationship of the brain in vivo. In testing our predictions, we contrasted two models (Fig. 1). The first represents a null hypothesis with regards to laminar specificity: it simply assumes that the effects observed in overall CT in McGugin et al. (2016) will replicate here but will be the same across all layers (Model 1). In contrast, Model 2, instantiates our predictions based on differential early development and posits a selective linkage between the deep MR layer and variability in FR.

Materials and Methods

Subjects. 14 right-handed men aged 19 to 29 (mean=22.46, s.d.=3.60) were recruited from the Vanderbilt University community based on previous participation. Only men were recruited to limit variability in brain anatomy. Our sample size was limited by the short period of time between when we developed scan protocols with sufficient resolution and before a major scanner upgrade. We recruited individuals with experience in the scanner, and who, as a group, offered a wide range of performance with faces. One subject was excluded due to excessive head motion. Written informed consent was obtained from all subjects in accordance with guidelines of the Vanderbilt University Institutional Review Board and Vanderbilt University Medical Center. All subjects reported normal or corrected-to-normal visual acuity, and received monetary compensation for participation.

Behavioral Testing. Subjects performed a battery of matching and recognition memory tests outside of the scanner. First, they completed the extended Cambridge Face Memory Test (CFMT+; (Russell, Duchaine, & Nakayama, 2009) in which they studied several views of

unfamiliar male faces and were tested in forced-choice trials of three faces on which they selected the face that matched a studied face. Second, subjects completed various sections of the Vanderbilt Expertise Test (VET; (McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012). This included 2 face tests (VET-Males, VET-Females), 2 vehicle tests (VET-Car, VET-Plane), and 2 animal tests (VET-Bird, VET-Butterfly). In each case, subjects studied six identities, models or species for as long as needed, followed by 3-alternative forced choice test trials in which they indicated which item was studied. Third, subjects completed the Vanderbilt Face Matching Test (VFMT; (Sunday, Lee, & Gauthier, 2018). On each trial, subjects studied 2 faces for 4000 ms, then were shown a 3-alternative forced choice and indicated which face was a new image of a face studied on that trial. Matching faces were different images from the original presentation. Fourth, subjects completed sequential matching tests using cars, planes, birds and butterflies (McGugin, Gatenby, Gore, & Gauthier, 2012). On each trial, an image appeared for 1000 ms, followed by a 500-ms mask, then a second image. Subjects judged if the two images showed cars/planes of the same make and model regardless of year, or birds/butterflies of the same species. See Table 1 for mean test performance, reliability of each test, and consistency across tests for each category. Aggregate Face, Vehicle and Animal indices were calculated based in each case on standardized performance for the four relevant tests.

MRI acquisition. All subjects were imaged using a Philips Achieva whole-body 7T MRI scanner with a quadrature, head only transmit coil and a 32 channel receive coil array. Imaging proceeded in three stages: whole brain structural imaging, functional localization, and ultra-high resolution susceptibility-weighted imaging (SWI).

Whole brain structural imaging acquisition. Whole brain structural imaging consisted of a sagittal whole brain 3D T1 weighted MPRAGE image gathered at 1 mm isotropic resolution and resampled into axial and coronal orientations for accurate localization of SWI acquisitions. This was used to identify anatomic landmarks important to the planning of the subsequent ultra-high resolution SWI acquisitions.

Functional MRI acquisition, stimuli, design and analyses. Functional localization of the fusiform face area (FFA) was accomplished via isotropic whole brain fMRI data. Images were acquired with minimal distortion using a multishot 3D gradient echo PRESTO EPI pulse sequence (Versluis et al., 2010). TR/TE=21.93/28.13ms, flip=12°, FOV=211.4mm, number of slices=30, voxel size=2.5mm isotropic, EPI factor=11, water/fat shift=6.705pix, number of dynamics=160, time/dynamic=2.02s).

Experimental Design and Statistical Analysis. We presented all stimuli with an Apple Macintosh computer running Matlab 2014a (MathWorks, Natick, MA) using Psychophysics Toolbox (Brainard, 1997). Stimuli were displayed on a rear-projection screen using an Avotec projector. Immediately following the HR structural scan, each subject completed 1-2 runs (depending on time) of a functional localizer scan (160 dynamics/run). We used grayscale images (36 faces, 36 objects) in a 1-back detection task across 20 alternating blocks of face and object images. Each image was presented for 1 s, with 16 images per category block. The HR T1-weighted structural scans were transformed into the standard sagittal view with 1mm isometric voxels to facilitate coregistration of functional and anatomical datasets. Functional data were analyzed using Brain Voyager (www.brainvoyager.com) and in-house Matlab scripts (MathWorks, Natick, MA). Preprocessing in Brain Voyager included registration to the structural scan, slice scan time correction (cubic spline), 3D motion correction (sinc interpolation) and temporal filtering (high-pass criterion of 2 cycles per run) with linear trend removal. No spatial smoothing was applied. Regions of interest (ROIs) were defined using the

Face>Object contrast from the face-localizer scan. We localized an ROI that responded more to faces than objects in the right middle fusiform gyrus (FFA2) of all subjects.

Right FFA2s were localized on the 3 mm³ (interpolated) statistical maps from the localizer scans as the peak face-selective voxel, then grown to incorporate contiguous 3-mm³ voxels passing statistical threshold within a 4-voxel or 108 mm³ fixed volume around the peak of face-selectivity. Solely for the purpose of comparing the location of our rFFA2s with those reported in the literature and demonstrating homology of the rFFA2 across subjects, the HR T1-weighted structural scans were normalized to Talairach space and functional data re-aligned to the Talairach-transformed images. Talairach coordinates for our ROIs (mm $\pm$ SD) are (X = 38.92 $\pm$ 3.86; Y = -47.54 $\pm$ 6.73; Z = -20.62 $\pm$ 3.01).

SWI image acquisition, processing and averaging. A minimum of six ultra-high resolution T₂^{*} weighted images were acquired as separate and independent acquisitions in each subject using slice-selective Cartesian gradient echo acquisitions. Both real and imaginary images were recorded. Measurement of individual differences in CT presents additional challenges over characterizing average laminar profiles because of the various sources of measurement error that can confound inter-individual variability. Thus, care was taken to plan image acquisitions such that the frequency encoding direction was perpendicular to the ventral surface of the temporal lobe underneath the right FFA, to minimize differences in partial volume effects across layers from subject to subject. Prior to analysis, ultra-high resolution data were reconstructed at 0.1875 mm in-plane resolution.

SWI imaging parameters were as follows: FOV=240x180.194x21.9mm, yielding voxel dimensions of 0.194x0.194x1.00 mm, 20 slices, 0.1 mm gap, 'shortest' (878.8 $\pm$ 8.29ms) repetition time, 'shortest' (27.5 $\pm$ 0.31ms) echo time, 27.26 pix water/fat shift, 55 degree flip angle, flow compensation, 9 min 11 $\pm$ 5.2 sec total duration. Images were acquired with a 1D phase navigator to correct for phase errors arising from respiration during acquisition, as has been previously shown to be effective in very high resolution SWI imaging (Versluis et al., 2010).

Real and imaginary images were used to calculate magnitude and phase images, which were then processed to create susceptibility weighted scans (Haacke, Xu, Cheng, & Reichenbach, 2004). Within each subject, each of the six resulting SWI scans were co-registered using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>) and averaged in MATLAB (<https://www.mathworks.com>). For those subjects that had more than six T₂^{*} weighted scans acquired, the optimal six were identified for analysis by calculating the normalized mutual information (NMI; (Studholme, Hill, & Hawkes, 1999; Viola & Wells III, 1997; Wells, Viola, Atsumi, Nakajima, & Kikinis, 1996) between all pairs of acquired and co-registered SWI scans, and ranking each according to the sum of its pairwise NMI values. The six scans with the highest summed pairwise NMI were kept for further analysis. This ensured that the six scans most similar to each other, and likely containing the least substantial run-dependent artifacts or misalignments, were selected for the purpose of averaging in later stages of the analysis.

Image registration. Standard-resolution functional localizer scans were registered to high-resolution (1mm isotropic) structural scans in BrainVoyager (Fig. 2A). Because we sought to avoid unnecessary interpolations of the ultra-high-resolution (UHR) images, all intensity analyses were performed directly on the UHR images. We registered the high-resolution structural image and the functionally-defined rFFA2 to the UHR slices. Specifically, high-resolution structural images and rFFA2 masks were converted from BrainVoyager's proprietary image formats to NifTI-1 images using the NeuroElf toolbox for Matlab. Rigid-body

transformation matrices of the high-resolution structural images to the UHR images were calculated via SPM's *SPM_coreg()* coregistration function, and these matrices were used to reslice the rFFA2 masks into the UHR space. The 108 mm³ rFFA2 masks traversed 4 to 7 UHR slices depending on the subject (mean= 5.57, SD= .85). For each UHR slice, 16-bit TIFF-format two-dimensional slice images were saved using Matlab.

Image Segmentation and Trace Identification. Using Adobe Photoshop, the gray matter/white matter (deep) and gray matter/CSF (superficial) cortical boundaries were manually drawn by two image analysts trained to recognize differences in brain structures but blind to the experimental goals (Fig. 2B). The inter-rater reliability of total rFFA2 thickness was very high (intraclass correlation, ICC=0.98). In addition to segmenting cortex from other tissue, blood vessels and image artifacts were manually marked as well. The resulting segmented data images were masked with the ROI images to create a series of ROI-specific cortical segmentation images. These segmented ROI images were next used to trace within-slice line segments through the thickness of cortex, along which we sampled the cortical signal intensity. Prior to tracing, each boundary was smoothed to avoid single-pixel concavities or convexities. Specifically, for each voxel that was part of the boundary, the three contiguous nearest neighbor boundary voxels in either direction (for a total of seven) were identified and a line segment fit to those voxels using linear regression. The set of such line segments was aggregated as the boundary was traversed, and skeletonized (to removed thickenings) and cleaned of spurs using the Matlab function *bwmorph()*.

Because of the convoluted shape of the cortical sheath, points on the deep and superficial boundaries of cortex do not have a one-to-one correspondence. Using custom Matlab code, we selected lines originating in each superficial boundary voxel and terminating in the nearest deep boundary voxel within that ROI in that slice (Fig. 2C). We conducted the same procedure in the reverse direction as well, originating in each deep boundary voxel and terminating in each superficial boundary voxel. The resulting traces were subject to five constraints: First, traces could not intersect cortical boundaries (aside from at trace origination or termination), to ensure that all samples were restricted to cortical tissue even in areas of high cortical convolution. Second, no trace could cross a previously-identified trace originating from the same boundary to avoid over-representing certain pieces of the cortical mantle. Third, the angle between the trace and boundary from which it originated could not be more than 20 degrees off perpendicular, to ensure that traces were approximately perpendicular to the cortical boundary. Fourth, any trace that was three or more standard deviations longer or shorter than the mean of all traces in a given slice was removed, to control for outlier paths. Fifth, no trace could cross an area marked for exclusion, such as a vein, since MRI intensity values are not valid indicators of gray matter signal in veins. For any trace that violated one of these constraints, a new trace was drawn from the originating voxel to the next-closest voxel in the other boundary.

Intensity sampling and MR layer identification. First, average CT was defined in each ROI in each slice per subject as the average distance between the GM-CSF border and GM-WM border. Next, three cortical depths were identified by the points at which gray matter intensity trends changed (Fig. 2D). Specifically, to estimate these depths, we fit a 4th-order polynomial to the set of intensity samples from each ROI in each slice. Gray matter intensity was sampled at 500 equally-spaced points along the length of each trace. This ensured that voxel intensities were sampled even when a trace only crossed a corner of a voxel. The sampled intensities were linearly interpolated to 1/100th mm intervals. The resulting intensities were thresholded since pixels above or below a certain intensity are assumed to reflect partial-volume or whole-pixel

capture of non-grey-matter. Specifically, any intensity values three or more standard deviations from the mean (over the entire set of trace intensities) were not included. Subsequently, intensity data were de-trended.

In all slices, inflection points were identified as the depths at which the second derivative of the polynomial was equal to zero. Inflection points outside the range of depths that fell within cortex were discarded and the remaining points were taken as estimates of the boundaries between cortical layers. Thus, we operationalized three MR cortical layers as the distances between (1) GM-CSF border and the first inflection point (superficial cortical depth), (2) first and second inflection points (middle cortical depth), and (3) second inflection point and the GM-WM border (deep cortical depth). We hypothesize that these MR layers are related to, although they are not a direct measure of, traditional cortical layers defined based on histological properties (Lifshits et al., 2018).

In 2 of 69 slices analyzed, the variation of intensity through the cortex was insufficient for accurately fitting a polynomial to the underlying data. These slices were omitted from further analyses.

Transcortical microvasculature. Transcortical microvasculature was segmented and mapped using a contour based local minimization routine. Having manually identified the deep and superficial boundaries of the functionally defined rFFA2, intermediary contours parallel to the cortical surface were drawn throughout the cortical thickness. Transcortical vessels were segmented by mapping local minima along each contour, and the vascular area was quantified as the percent of voxels within the rFFA2 labeled as belonging to transcortical vessels. The measure for each subject was an average from at least 4 and up to 7 slices, with a weighted average ICC of .80.

Results

Reliability of CT measurements.

Slices. We estimated the reliability of cortical measurements across slices (after centering them on the peak of face selectivity) using the correlation of thickness estimates for each slice relative to all the other slices in each subject. For total regional CT, which was based on manual tracings of high-contrast boundaries, correlations were higher than .8 for all slices. As a result, total CT measurements were averaged over all available slices. Inter-rater reliability of total CT was calculated using intraclass correlations for distance from GM-WM boundary to GM-CSF boundary, averaged across all slices for the two analysts.

Thickness measurements of the middle MR layer were based on relatively lower contrast boundaries identified using an automated algorithm (see below). Correlations reached .5 for anterior slices but slices that were posterior relative to the peak of face selectivity showed little to no evidence of reliability. Inspecting individual images, this proved to be because FFA2 ROIs were located closer to the posterior slices of the ultra-high resolution stack of 20 images for most subjects. Our measurements were made after co-registration of six ultra-high resolution scans to increase contrast, thus the final slices of the stack were more likely to show co-registration artifacts due to motion or respiration. We therefore used only the 2-4 reliable anterior slices for each subject and dropped the 2-3 most posterior and unreliable slices, for all measurements involving an internal laminar boundary (thickness estimates for deep, superficial and middle MR layers).

Average Thickness Measures. The white matter (WM)/grey matter (GM) boundary and the GM/cerebrospinal fluid (CSF) boundary were traced manually by two image analysts. Across

subjects, the mean total CT of rFFA2 (distance from the WM/GM boundary to the GM/CSF boundary) was consistent with previous work. Between the two analysts blind to behavioral measurements, inter-rater reliability of total rFFA2 thickness was very high: intraclass correlation (ICC) = 0.98). MR-defined laminar subdivisions are defined by two internal boundaries identified using an automated procedure (see Methods). The representation of each MR laminar subdivision proportional to total CT was also consistent with previous work (Fig. 2D). Across subjects, the mean total CT of rFFA2 (distance from the WM/GM boundary to the GM/CSF boundary) was in the range reported in extrastriate cortex (Fischl & Dale, 2000): average rFFA2 GM thickness = 3.37 mm, with a range between 2.50 and 4.67 mm. Furthermore, the representation of each laminar subdivision proportional to total CT was consistent with MR microscopy in area V4 in nonhuman primates (Chen, Wang, Gore, & Roe, 2012): superficial MR layer (1.07 mm, range: 0.74 – 1.61 mm; 31%), middle MR layer (1.04 mm, range: 0.64 – 1.49 mm; 31%), deep MR layer (1.26 mm, range: 0.56 – 2.72 mm; 38%). As an index of consistency across subjects, ICCs for the three laminar subdivisions were: superficial, 0.77, middle, 0.73, deep, 0.80.

Association between Behavioral Measures and Total rFFA2 CT.

Each behavioral measure was computed as the average of the standardized values of four tests. ICCs were greater than .84 in all cases. We replicated previous results showing that total FFA CT is negatively correlated with face recognition, ($r=-0.83$ $p<.001$) and positively correlated with the recognition of vehicles ($r=0.70$, $p<0.001$, Table 2), with the two correlations significantly different ($t=8.86$, $p<.001$). Face and Vehicle performance were not significantly related ($r=-.33$, $p=.27$) and partial correlations with rFFA2 controlling for performance on the other categories were comparable in value and statistical significance to the zero-order r s (Table 2, Fig 3A). We replicated these correlational results using the univariate observed-imposed bootstrap (e.g., (Beasley et al., 2007). Multiple regression indicated that both Face and Vehicle scores independently predicted rFFA2 thickness (both p s $\leq .001$) and together accounted for 60% of the variance on adjusted R^2 .

As in prior work (McGugin et al., 2016), the correlation between overall CT and recognition of animals, which may not be acquired especially early or late, was in the same direction as face recognition although not significantly so ($r=-0.12$, $p=0.34$). Animal recognition also did not predict thickness of MR layers (superficial: $r=0.15$; middle: $r=0.26$; deep: $r=-0.26$). Our results cannot be explained by variability in vascular volume in FFA2 because such differences were not related to total CT ($r=-.31$), or behavioral performance (face, $r=.11$; vehicle, $r=-.09$, animal, $r=.21$), despite being reliable (ICC=.75).

Association between Behavior Measures and Thickness of rFFA2 MR layers.

Correlations. In accord with predictions of the differential early development model, face recognition was strongly negatively correlated with the thickness of deep MR layers of rFFA2 ($r=-0.76$, $p=.002$, Fig. 3B, Table 2). This relation is evident when examining subject-specific brain data listed as a function of face recognition performance (Fig. 4). There were no significant correlations between face recognition and thickness of superficial or middle MR layers (Fig. 3B, Table 2). In contrast, correlations between vehicle recognition and thickness of all three MR layers were consistently positive in sign (r s ranged from .44 to .48, two-tail p s ranging from .13 to .10), although not statistically significant. For the deep MR layer, the

correlations of thickness with face recognition significantly differed from that with vehicle recognition ($t=3.69$, $p<.005$). These effects were replicated when bootstrapping was used.

Regression Models Predicting Behavioral Measures From MR Layer-Specific CT.

While correlations revealed bivariate relations among pairs of variables, they do not address the degree to which the combination of MR layer-specific CT measures predict individual differences in behavior. To address this question, we conducted multiple regression analyses in which CT in the superficial, middle, and deep MR layers predicted face recognition and vehicle recognition scores. According to adjusted R^2 , the combination of superficial, middle, and deep MR layer thickness accounted for 65% of the variance in face recognition (FR; overall regression $F(3,9)=8.29$, $p=.006$; unadjusted $R^2=.73$). Both standard and bootstrapped confidence intervals indicated that the thickness of the deep layer was the only statistically significant predictor of face recognition. We also computed semi-partial correlations indicating the correlation between FR and each CT measure with the latter adjusted for the other CT measures. Only the semi-partial correlation for the deep MR layer ($r=.72$, $p=.0125$) was significantly greater than 0 (superficial $r=-.29$, $p=.40$; middle $r=-.25$, $p=.45$). Squared semi-partial correlations quantifying the proportion of variance uniquely contributed by each predictor indicated that the thickness of the deep MR layer accounted for slightly over half the total variance of FR (deep $sr^2=.52$, superficial and middle $sr^2=.08$ and $.06$ respectively).

The multiple regression analyses with vehicle recognition (VR) as the dependent variable had a different pattern of results. In this case, the overall regression model was significant ($F(3,9)=3.98$, $p<.05$) although none of the regression coefficients for the three predictors were significant (ps ranged from $.07$ to $.12$). Although the latter result could be due to the relatively small sample size, this result also suggests that it was the shared variance among the three predictors rather than the unique variance of each predictor that accounted for effects on VR.

Regression Models Predicting CT of MR Layers from Behavior. Our most extensive analyses used a multiple regression framework to assess alternative models of the effects of face recognition and vehicle recognition on CT measures. We emphasized these analyses because they allowed us to test alternative hypotheses about the *complete pattern of associations* between the two behavioral measures and thickness of the three MR layers.

Our regression models used a repeated measures framework, with the 3 MR layer-specific CT measures per person serving as the repeated dependent measure. Predictors were layer, FR, and VR. We created two new variables that decomposed the overall effect of the layer factor into two orthogonal (i.e., uncorrelated) contrasts. As noted above, we hypothesized that the association between FR and thickness would be more strongly negative in the deep MR layer relative to the superficial and middle MR layers. To capture this difference, we gave one contrast variable (denoted L_{DvSM}) a code of -1 for observations from the deep MR layer and codes of .5 for observations from the superficial and middle MR layers. The second contrast variable for layer (denoted L_{SvM}) was orthogonal to the first and assigned a code of 0 to the deep layer, 1 to the superficial layer, and -1 to the middle layer. When both contrast variables are included in a model as a set, they estimate the overall (2 degree of freedom) main effect of MR layer.

We compared six different regression models that imposed different restrictions on the association between measures of the CT of MR layers (hereafter “layers” for short in the model descriptions) and FR and VR.

Model 0 (M_0): Baseline Model.

In this model, the only predictor of CT levels was the overall main effect of layer:

$$\text{Model 0: } CT_{ij} = B_0 + B_1L_{DvsSM} + B_2L_{SvM} + \varepsilon_{ij},$$

where i denotes subject (1 through 13), j denotes layer (1 through 3), and

ε denotes the prediction error for the CT measure in the j th layer for the i th subject.

We used M_0 simply as a baseline comparison to verify that FR and VR measures, introduced in subsequent models, were significantly associated with CT.

Model 1(M_1): Equal Coefficients Across Layers.

The next two models reflected our two primary alternative hypotheses. Model 1 predicted CT in a given layer from the two contrast variables for layer, FR, and VR:

$$\text{Model 1: } CT_{ij} = B_0 + B_1L_{DvsSM} + B_2L_{SvM} + B_3FR_i + B_4VR_i + \varepsilon_{ij}$$

M_1 is a “main effects only” model with no interaction terms. As a result it specifies that the regression coefficient denoting the effect of FR on CT (B_3) is equal (i.e., invariant) across all three layers and the coefficient for the effect of VR on CT (B_4) is also equal across layers (see Figure 1). This model is the most parsimonious extension of our previous finding using overall CT (McGugin et al., 2016).

Model 2 (M_2): FR Differentially Predicts Deep vs Superficial/Middle MR Layer CT.

Model 2 tested the hypothesis that: (1) The association between FR and CT is more pronounced (i.e., more negative) in the deep layer than the middle and superficial layers; but (2) Layer does not moderate the relation between VR and CT measures. To embody this specification, an interaction (i.e., product) term between FR and the Deep vs. Superficial/Middle contrast variable (L_{DvsSM}) was added to Model 1:

$$\text{Model 2: } CT_{ij} = B_0 + B_1L_{DvsSM} + B_2L_{SvM} + B_3FR_i + B_4VR_i + B_5(FR_i \times L_{DvsSM}) + \varepsilon_{ij}$$

In model 2, three unique regression coefficients are sufficient to account for the linkages between FR and VR and CT across layers: the coefficient for the regression of CT on FR within the deep layer (equal to $B_3 - B_5$); the coefficient for the regression of CT on FR within the middle and superficial layers (equal to $B_3 + \frac{B_5}{2}$); and, the single coefficient for the regression of CT on VR within each of the three layers (B_4). Thus, this model specifies a constant effect of VR across all three layers and the hypothesized variable effect for FR.

The remaining models are designed to assess patterns of association other than those directly hypothesized in Models 1 and 2:

Model 3 (M_3) extended Model 2 by adding an interaction term between the second Layer contrast variable (L_{SvM}) and FR. This addition relaxes the assumption that the regression coefficients for FR in the middle and superficial layers were equal to one another. Thus, M_3 goes beyond M_2 by allowing all three regression coefficients for FR to differ from one another. Like M_2 , however it restricts all three regression coefficients for VR to be equal.

Model 4 (M₄) added the full VR X Layer interaction to Model 1. This model specifies that the regression coefficients involving FR are all the same but allows for unique regression coefficients for VR in each layer.

Model 5 (M₅) extended M₃ and M₄ by adding in both the overall FR X Layer interaction and the overall VR X Layer interaction. This model allows for 6 unique regression coefficients (FR/VR x Superficial/Middle/Deep) to capture the association between FR and VR and CT measures. Support for this model would imply that layer moderates the relation between *both* FR and VR and CT.

We conducted analyses using the gls function (Gałecki & Burzykowski, 2013) in the nlme package (Pinheiro et al., 2018) in R (R Core Team, 2018). For all models, we accounted for the dependence among the three observations for each subject by imposing an unstructured covariance matrix among the residuals. Maximum likelihood (ML) estimation was used (e.g., (Gałecki & Burzykowski, 2013)). We evaluated the relative fit of the models using several indices. When models were nested, we used likelihood ratio tests to compare models. We also used three information indices to compare all models: The Akaike Information Criterion (AIC; (Akaike, 1973), the small-sample corrected AIC (AICc; (Sugiura, 1978), and the Bayesian Information Criterion (BIC; (Schwarz, 1978)). These indices reward model fit but penalize for model complexity, and thus are designed to prevent overfitting. Lower values indicate better fit. We also report -2 X the log likelihood (-2LL) value of each model computed at the maximum likelihood estimates. -2LL values are used both in the computation of likelihood ratio tests (for which the difference in the -2LL values of two models are distributed as chi-square variables with degrees of freedom equal to the difference in the number of parameters estimated by each model) and of information indices.

Table 3 shows the values of -2LL and the information indices for the six models. The values for M₀ very clearly fit worse than all other models (e.g., all LR tests comparing other models to M₀ ps < .005), thus indicating that FR and VR contributed significantly to the prediction of CT. Most importantly, Table 4 indicates that M₂ (“FR Differentially Predicts Deep vs Superficial/Middle CT”) is clearly the best-fitting model. It has the lowest values for all three information indices. Likelihood ratio tests comparing M₂ to other models yielded a pattern consistent with the information indices. M₂ fit significantly better than M₁ ($\chi^2_1 = 5.94, p = .02$). Although M₂ is a restricted version of M₃ and M₅, neither fit any better than M₂ (M₂ vs. M₃: $\chi^2_1 = .01, p > .90$; M₂ vs M₅: $\chi^2_3 = .20, p > .97$). These results indicate that M₂ is the model that best combines fit to the data with parsimony. Table 5 shows the unstandardized and standardized regression coefficients for the regression of CT on FR and VR that are implied by M₂. FR has a stronger (negative) effect on CT in the deep MR layer relative to the superficial and middle MR layers. This difference is statistically significant ($t(33) = 2.53, p < .02$).

Although the fit indices discussed so far clearly indicate that M₂ is the best-fitting model, they do not precisely quantify the strength of the evidence for M₂ relative to the other hypotheses. For this reason, we additionally computed Bayes factors (BFs; (Jeffreys, 1961; Kass & Raftery, 1995)) using the BayesFactor package in R (Morey & Rouder, 2018). We specified the same six regression models. Instead of correlated residuals, we specified a random effect for subject to model the correlation among the CT values for each person. For all models, we used the Jeffreys-Zellner-Snow mixture of g-priors specification introduced by (Liang, Paulo, Molina, Clyde, & Berger, 2008) (see also (Rouder & Morey, 2012)). Because BFs can be sensitive to prior

distributions, we used three alternative specifications of the scale parameter of the inverse gamma distribution that, in turn, is used to specify the variances of the prior distribution for the regression coefficients. All three specifications produced identical conclusions. Below, we report the results for the default “medium” values of the scaling parameters. For each pair of models, BFs were computed as the ratio of the marginal likelihood of the first model to the marginal likelihood of the second. Although we caution against rules of thumb, Jeffreys’ criteria specify that BFs in the range 1-3, 3-10, 10-30, 30 to 100, and > 100 indicate, respectively, weak, substantial, strong, very strong, and decisive support for the hypothesis linked to the numerator of the BF. Reciprocal values indicate degree of support for the model in the denominator.

An examination of the BFs yielded by the 15 pairings of models indicated that M_2 was clearly the best-fitting model (i.e., for each of its 5 comparisons, M_2 had the best fit). The final column in Table 3 displays the Bayes Factors comparing M_2 to each of the other models. Perhaps the most important index here is the BF of 38.85 for the M_2/M_1 comparison. This value indicates that the evidence in the data for M_2 is over 30 times as strong as the evidence for M_1 . All other comparisons yielded what Jeffrey’s termed “strong” support for M_2 except that involving M_3 . This is not surprising given that both models are highly similar. Both allow for a difference in the association between FR and CT between the deep and the superficial layers, and both specify that the regression coefficients for VR are the same across layers. Unlike M_2 , however, M_3 allows for differences between the FR coefficients in the superficial and middle layers. That the evidence favoring M_2 is over three times stronger indicates that the superficial and middle layer coefficients can be constrained to be equal. Thus, the best-fitting model in both the regression and Bayesian analyses (M_2) required only three unique regression coefficients to capture the relation between CT and FR and VR across the three MR layers: the coefficient denoting the effect of FR on CT in the deep MR layer, the coefficient denoting the effect of FR on CT in the superficial and middle MR layers, and the coefficient denoting the effect of VR on CT in all three MR layers.

Discussion

We used susceptibility-weighted MRI in a 7T magnet to measure the thickness of deep, middle and superficial MR layers, distinctions that cannot be resolved with conventional MRI approaches at 3T or below. While new developments in MRI have allowed some visualization of intracortical layers, validation of their functional relevance has been reserved to primary visual cortex (Trampel, Ott, & Turner, 2011). We were constrained to a single region of interest due to a limited field of view and the fact that accurate measurement of laminar CT with non-isotropic voxels depends upon careful alignment of slices perpendicular to the cortex of a targeted brain area. We selected the rFFA2 based on earlier robust functional expertise effects for non-face objects (McGugin, Van Gulick, Tamber-Rosenau, Ross, & Gauthier, 2015).

Here, we replicate and extend the finding that vehicle recognition is positively correlated with CT of FFA (McGugin et al., 2016), whereas face recognition is negatively correlated with FFA thickness (Bi, Chen, Zhou, He, & Fang, 2014; McGugin et al., 2016). This surprising pattern of correlations led us to postulate different mechanisms of plasticity behind these effects, with the mechanism relevant to face recognition operating earlier and selectively influencing the deep layers of rFFA2.

To measure the thickness of MR layers, we applied an automated algorithm to define borders between layers using signal intensity shifts along perpendicular traces through rFFA2 (Methods and Fig. 2). We postulate that the three most prominent intensity differences are

related to observed laminar divisions in the fusiform (Lorenz et al., 2017), corresponding to the densely-packed deep/infragranular layers (layers V-VI), middle/internal granular (layer IV), and superficial/supragranular (layers I-III). We successfully visualized laminar structure in rFFA2 and our measurements of the thickness of MR layers were reliable.

Although we compared several different models of the data, our primary focus was on two specific alternatives (see Fig. 1). The first (M_1) specified that the pattern of correlations observed in total CT (negative for faces and positive for vehicles) is constant across all laminar subdivisions. The second (M_2) specified that the association between individual differences in FR and the thickness of the deep MR layer was stronger (i.e., more negative) than the association between FR and thickness of the superficial and middle MR layers. Like M_1 , M_2 specified that the association between VR and CT was invariant across layer. Regression analyses consistently favored M_2 relative to M_1 and to the other models that we specified, across multiple indices (likelihood ratio tests, three information indices, and Bayes Factors). Comparisons of regression coefficients yielded by M_2 indicated that the association between FR and the deep MR layer was significantly stronger than the association between FR and the superficial and middle MR layers. The thickness of the deep MR layer accounted for a much higher proportion of variance in FR than the thickness of the other two MR layers. Indeed, when considered in isolation, CT of the deep MR layer accounted for 58% of the variance in FR, which is 84% of the variance accounted for by all the layers together (using total CT). Thus, the strong association between total CT and FR is largely, if not exclusively, due to the deep MR layer. In contrast, for VR there was no evidence that the relation was moderated by layer, and the clearly best-fitting regression model (M_2) constrained the regression coefficients for VR to be equal across layers.

The sample size of the present study ($N=13$) was small. Accordingly, we focused on regression models that specified the overall pattern of association across layers and behavioral measures, to increase power and precision (e.g., see (Hojtink, 2011)). In the present case, for example, while the individual correlations between VR and CT in the superficial, middle, and deep layers were not significantly greater than 0, our regression models yielded significant effects for VR on CT when these were constrained to be equal across layers. More generally, that our analyses indicated the clear superiority of M_2 even given the relatively small sample size is an indication of its strength as a representation of the underlying pattern of associations.

These results are amongst the most structurally precise correlates of behavioral ability to date, and an important validation of the functional relevance of our in-vivo laminar measurements. Future development of these methods to a larger field of view should allow exploration of how these fine structural effects relate to correlations of face recognition ability on a much more distributed scale (Elbich & Scherf, 2017). Our results also represent perhaps the most striking difference to date between face and non-face recognition. Face recognition has often been deemed special relative to the recognition of other categories (Farah, Wilson, Drain, & Tanaka, 1998), but many behavioral hallmarks of face expertise, such as the inversion effect or holistic processing, are observed with non-face objects like cars in expert subjects (Bukach, Phillips, & Gauthier, 2010; Curby, Glazek, & Gauthier, 2009). Face and car recognition also show similar degrees of heritability (Shakeshaft & Plomin, 2015) and similar correlations with domain-general object recognition ability (Richler, Wilmer, & Gauthier, 2017). In fMRI studies, car recognition ability correlates positively with selective neural responses to images of cars in FFA (e.g., (McGugin, Gatenby, et al., 2012)), while face recognition ability also correlates positively with selective responses to faces in the same area (McGugin, Ryan, Tamber-Rosenau, & Gauthier, 2018). Despite these similarities, our results suggest that behaviors supported by

common circuitry in a brain region may have developed at different times, with variability in structure as a footprint of the different acquisition history.

Our predictions were inspired by the idea that face recognition ability begins earlier than vehicle recognition and that a skill acquired during this period would be influenced by the pruning of connections from IT to limbic areas, though this is not the only possibility. We do not know whether the relevant period of influence on FFA's structure occurs before about 5 years of age (by that age, children show many signs of the expertise adults display with faces (Jeffery et al., 2011)), when pruning may be more important, or in later childhood and adolescence, when myelination may dominate developmental changes (Natu et al., 2018). An alternative explanation is that greater myelination near FFA during childhood and adolescence is associated with better face recognition. Recent work has questioned the prevalent theory that pruning is the main force behind cortical thinning during development, with more evidence for myelination than for pruning after 5 years of age in the fusiform gyrus (Natu et al., 2018). Also unknown are the specific mechanisms of plasticity that would support structural effects of visual learning in adults, leading us to assume the null hypothesis of no laminar specificity of the positive correlation with CT observed for vehicles. Importantly, the mechanism driving the largest amount of average change in CT in an area may not be the same mechanism driving individual differences in CT.

It is unknown whether negative correlations between CT and performance will prove to be specific to deep MR layers for other kinds of abilities and other brain regions. It is also unknown which of the two patterns (positive correlation with total regional CT or negative correlation with thickness of deep MR layers) is more representative of the majority of object categories. Here, as in prior work at 3T (McGugin et al., 2016), there was no significant correlation with animal recognition but the trend was more similar to effects found with faces than vehicles (a negative correlation for thickness of deep MR layers). Future studies should explore how laminar CT in FFA relates to a more extensive set of visual abilities, and how this relation is affected by sex. Indeed, as of now, the cross-over interaction between CT and visual recognition abilities has only been described in men, yet there are sex differences in the recognition of living and non-living objects (McGugin, Richler, et al., 2012), and sex also moderates CT effects (Plessen, Hugdahl, Bansal, Hao, & Peterson, 2014).

By the time CT is measured in adults, effects in any given brain area will reflect a complex history of changes due to both experience- and non-experience-dependent developmental factors. This complicates interpreting relationships between behavior and local brain structure, including how structure underlies various neurological disorders. For example, both abnormal thinning of fusiform cortex over development (Raznahan et al., 2017) and lack of developmental improvements in face recognition (O'hearn, Schroer, Minshew, & Luna, 2010) have been reported in Autism Spectrum Disorders. The study of the laminar specificity of structural correlates of behavioral abilities may provide a useful window into these complex dynamics.

References

- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. Petrov & F. Csaki (Eds.), *Proceedings of the 2nd International Symposium on Information Theory* (pp. 267–281). https://doi.org/10.1007/978-1-4612-1694-0_15
- Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, 20(10), 1404–1412. <https://doi.org/10.1038/nn.4635>
- Beasley, W. H., DeShea, L., Toothaker, L. E., Mendoza, J. L., Bard, D. E., & Rodgers, J. L. (2007). Bootstrapping to test for nonzero population correlation coefficients using univariate sampling. *Psychological Methods*, 12(4), 414–433. <https://doi.org/10.1037/1082-989X.12.4.414>
- Bi, T., Chen, J., Zhou, T., He, Y., & Fang, F. (2014). Function and structure of human left fusiform cortex are closely associated with perceptual learning of faces. *Current Biology*, 24(2), 222–227. <https://doi.org/10.1016/j.cub.2013.12.028>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Bukach, C. M., Phillips, W. S., & Gauthier, I. (2010). Limits of generalization between categories and implications for theories of category specificity. *Attention, Perception, and Psychophysics*, 72(7), 1865–1874. <https://doi.org/10.3758/APP.72.7.1865>
- Chen, G., Wang, F., Gore, J. C., & Roe, A. W. (2012). Identification of cortical lamination in awake monkeys by high resolution magnetic resonance imaging. *NeuroImage*, 59(4), 3441–3449. <https://doi.org/10.1016/j.neuroimage.2011.10.079>
- Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2016). Real-world visual statistics and infants’ first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160055. <https://doi.org/10.1098/rstb.2016.0055>
- Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 94–107. <https://doi.org/10.1037/0096-1523.35.1.94>
- Elbich, D. B., & Scherf, K. S. (2017). Beyond the FFA: Brain-behavior correspondences in face recognition abilities. *NeuroImage*, 147, 409–422. <https://doi.org/10.1016/j.neuroimage.2016.12.042>
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “Special” about face perception? *Psychological Review*, 105(3), 482–498. <https://doi.org/10.1037/0033-295X.105.3.482>
- Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, 152, 101–107. <https://doi.org/10.1016/j.cognition.2016.03.005>
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20), 11050–11055. <https://doi.org/10.1073/pnas.200033797>
- Gałecki, A., & Burzykowski, T. (2013). Linear Mixed-Effects Models Using R: A Step-by-Step Approach. *Springer*, New York. <https://doi.org/10.1113/jphysiol.1932.sp002901>
- Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, 118(2), 201–110. <https://doi.org/10.1016/j.cognition.2010.11.002>

- Gomez, J., Barnett, M. A., Natu, V., Mezer, A., Palomero-Gallagher, N., Weiner, K. S., ... Grill-Spector, K. (2017). Microstructural proliferation in human cortex is coupled with the development of face processing. *Science*, 355(6320), 68–71. <https://doi.org/10.1126/science.aag0311>
- Haacke, E. M., Xu, Y., Cheng, Y.-C. N., & Reichenbach, J. R. (2004). Susceptibility weighted imaging (SWI). *Magnetic Resonance in Medicine*, 52(3), 612–618. <https://doi.org/10.1002/mrm.20198>
- Hojtink, H. (2011). Informative hypotheses: Theory and practice for behavioral and social scientists. *CRC Press*, Boca Raton, FL. https://doi.org/10.1207/S15327906MBR3604_04
- Jeffery, L., Rhodes, G., McKone, E., Pellicano, E., Crookes, K., & Taylor, E. (2011). Distinguishing norm-based from exemplar-based coding of identity in children: Evidence from face identity aftereffects. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1824–1840. <https://doi.org/10.1037/a0025643>
- Jeffreys, H. (1961). *Theory of Probability*. Oxford: Clarendon Press.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481), 410–423. <https://doi.org/10.1198/016214507000001337>
- Lifshits, S., Tomer, O., Shamir, I., Barazany, D., Tsarfaty, G., Rosset, S., & Assaf, Y. (2018). Resolution considerations in imaging of the cortical layers. *NeuroImage*, 164, 112–120. <https://doi.org/10.1016/j.neuroimage.2017.02.086>
- Lorenz, S., Weiner, K. S., Caspers, J., Mohlberg, H., Schleicher, A., Bludau, S., ... Amunts, K. (2017). Two New Cytoarchitectonic Areas on the Human Mid-Fusiform Gyrus. *Cerebral Cortex*, (27), 373–385. <https://doi.org/10.1093/cercor/bhv225>
- McGugin, R. W., Gatenby, J. C., Gore, J. C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proceedings of the National Academy of Sciences*, 109(42), 17063–17068. <https://doi.org/10.1073/pnas.1116333109>
- McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt Expertise Test reveals domain-general and domain-specific sex effects in object recognition. *Vision Research*, 69, 10–22. <https://doi.org/10.1016/j.visres.2012.07.014>
- McGugin, R. W., Ryan, K. F., Tamber-Rosenau, B. J., & Gauthier, I. (2018). The role of experience in the face-selective response in right FFA. *Cerebral Cortex*, 28(6), 2071–2084. <https://doi.org/10.1093/cercor/bhx113>
- McGugin, R. W., Van Gulick, A. E., & Gauthier, I. (2016). Cortical thickness in fusiform face area predicts face and object recognition performance. *Journal of Cognitive Neuroscience*, 28(2), 282–294. https://doi.org/10.1162/jocn_a_00891
- McGugin, R. W., Van Gulick, A. E., Tamber-Rosenau, B. J., Ross, D. A., & Gauthier, I. (2015). Expertise effects in face-selective areas are robust to clutter and diverted attention, but not to competition. *Cerebral Cortex*, 25(9), 2610–2622. <https://doi.org/10.1093/cercor/bhu060>
- Morey, R. D., & Rouder, J. N. (2018). *Bayesfactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.2*. <https://doi.org/10.1007/978-3-642-21037-2>
- Natu, V. S., Gomez, J., Barnett, M. A., Kirilina, E., Jaeger, C., Cox, S., ... Grill-Spector, K. (2018). Apparent thinning of visual cortex during childhood is associated with myelination, not pruning. *BioRxiv*, 368274. <https://doi.org/10.1101/368274>

- O’hearn, K., Schroer, E., Minshew, N., & Luna, B. (2010). Lack of developmental improvement on a face memory task during adolescence in autism. *Neuropsychologia*, 48(13), 3955–3960. <https://doi.org/10.1016/j.neuropsychologia.2010.08.024>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., & Willigen, B. Van. (2018). *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-137*. Retrieved from <https://cran.r-project.org/web/packages/nlme/index.html>
- Plessen, K. J., Hugdahl, K., Bansal, R., Hao, X., & Peterson, B. S. (2014). Sex, age, and cognitive correlates of asymmetries in thickness of the cortical mantle across the life span. *Journal of Neuroscience*, 34(18), 6294–6302. <https://doi.org/10.1523/JNEUROSCI.3692-13.2014>
- R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria*. <https://doi.org/10.1108/eb003648>
- Raznahan, A., Toro, R., Daly, E., Robertson, D., Murphy, C., Deeley, Q., ... Murphy, D. G. M. (2017). Cortical anatomy in autism spectrum disorder: an in vivo MRI study on the effect of age. *Cerebral Cortex*, 20(6), 1332–1340. <https://doi.org/10.1093/cercor/bhp198>
- Richler, J. J., Wilmer, J. B., & Gauthier, I. (2017). General object recognition is specific: Evidence from novel and familiar objects. *Cognition*, 166, 42–55. <https://doi.org/10.1016/j.cognition.2017.05.019>
- Rouder, J. N., & Morey, R. D. (2012). Default bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47, 877–903. <https://doi.org/10.1080/00273171.2012.734737>
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16(2), 252–257. <https://doi.org/10.3758/PBR.16.2.252>
- Scherf, K. S., Thomas, C., Doyle, J., & Behrmann, M. (2014). Emerging structure-function relations in the developing face processing system. *Cerebral Cortex*, 24(11), 2964–2980. <https://doi.org/10.1093/cercor/bht152>
- Schultz, R. T. (2005). Developmental deficits in social perception in autism: The role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience*, 23(2–3), 125–141. <https://doi.org/10.1016/j.ijdevneu.2004.12.012>
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences*, 112(41), 12887–12892. <https://doi.org/10.1073/pnas.1421881112>
- Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. *Journal of Neuroscience*, 24(38), 8223–8231. <https://doi.org/10.1523/JNEUROSCI.1798-04.2004>
- Stefanacci, L., & Amaral, D. G. (2000). Topographic organization of cortical inputs to the lateral nucleus of the macaque monkey amygdala: A retrograde tracing study. *Journal of Comparative Neurology*, 421(1), 52–79. [https://doi.org/10.1002/\(SICI\)1096-9861\(20000522\)421:1<52::AID-CNE4>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1096-9861(20000522)421:1<52::AID-CNE4>3.0.CO;2-O)
- Studholme, C., Hill, D. L. G., & Hawkes, D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1), 71–86. [https://doi.org/10.1016/S0031-3203\(98\)00091-0](https://doi.org/10.1016/S0031-3203(98)00091-0)

- Sugden, N. A., & Moulson, M. C. (2018). These are the people in your neighbourhood: Consistency and persistence in infants' exposure to caregivers' relatives' and strangers' faces across contexts. *Vision Research*. <https://doi.org/10.1016/j.visres.2018.09.005>
- Sugiura, N. (1978). Further analysis of the data by Anaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, 7(1), 13–26. <https://doi.org/10.1080/03610927808827599>
- Sunday, M. A., Lee, W.-Y., & Gauthier, I. (2018). Age-related differential item functioning in tests of face and car recognition ability. *Journal of Vision*, 18(1–2), 1–17. <https://doi.org/10.1167/18.1.2>
- Trampel, R., Ott, D. V. M., & Turner, R. (2011). Do the congenitally blind have a Stria of Gennari? First intracortical insights in vivo. *Cerebral Cortex*, 21(9), 2075–2081. <https://doi.org/10.1093/cercor/bhq282>
- Versluis, M. J., Peeters, J. M., van Rooden, S., van der Grond, J., van Buchem, M. A., Webb, A. G., & van Osch, M. J. P. (2010). Origin and reduction of motion and f0 artifacts in high resolution T2*-weighted magnetic resonance imaging: Application in Alzheimer's disease patients. *NeuroImage*, 51(3), 1082–1088. <https://doi.org/10.1016/j.neuroimage.2010.03.048>
- Viola, P., & Wells III, W. M. (1997). Alignment by maximization of mutual information. *The International Journal of Computer Vision*, 24(2), 137–154. <https://doi.org/10.1109/ICCV.1995.466930>
- Webster, M. J., Ungerleider, L. G., & Bachevalier, J. (1991). Lesions of inferior temporal area TE in infant monkeys alter cortico-amygdalar projections. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*, 2(12), 769–772. <https://doi.org/10.1097/00001756-199112000-00010>
- Wells, W. M., Viola, P., Atsumi, H., Nakajima, S., & Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1), 35–51. [https://doi.org/10.1016/S1361-8415\(01\)80004-9](https://doi.org/10.1016/S1361-8415(01)80004-9)
- Wenger, E., Schaefer, S., Noack, H., Kühn, S., Mårtensson, J., Heinze, H. J., ... Lövdén, M. (2012). Cortical thickness changes following spatial navigation training in adulthood and aging. *NeuroImage*, 59(4), 3389–3397. <https://doi.org/10.1016/j.neuroimage.2011.11.015>
- Zatorre, R. J., Fields, R. D., & Johansen-berg, H. (2012). Plasticity in gray and white : neuroimaging changes in brain structure during learning. *Nature Publishing Group*, 15(4), 528–536. <https://doi.org/10.1038/nn.3045>

Table 1. Behavioral results across twelve independent tests. For one subject, performance on VET-Male was near chance and an outlier relative to other subjects and his own other scores, so this data point was dropped. The ICC4 (with 95% HPD interval) for the aggregate of Face tests was .89 (.77,.96), for the aggregate Vehicle tests, .83 (.66,.95) and for the aggregate Animal tests, .85 (.70,.96).

		Mean (std dev)	Cronbach's Alpha
Faces	CFMT+	0.65 (0.13)	0.88
	VET-Male (Acc)	0.83 (0.17)	0.78
	VET-Female (Acc)	0.83 (0.12)	0.86
	VFMT (Acc)	0.6 (0.07)	0.86
Vehicles	VET-Car (Acc)	0.73 (0.18)	0.93
	VET-Plane (Acc)	0.76 (0.12)	0.93
	Matching-Car (d')	1.79 (1.00)	0.75
	Matching-Plane (d')	1.57 (0.77)	0.86
Animals	VET-Bird (Acc)	0.68 (0.17)	0.80
	VET-Butterfly (Acc)	0.56 (0.17)	0.83
	Matching-Bird (d')	1.46 (0.39)	0.92
	Matching-Butterfly (d')	1.65 (0.52)	0.82

Table 2. Pearson's correlations between behavioral performance and cortical thickness and laminar depth measurements. Zero-order Pearson correlation coefficients are shown for Face and Vehicle categories correlated with total regional thickness and cortical depths operationalized by the 4th order polynomial and its 2nd order derivative. Partial-order Pearson correlation coefficients are given for each behavioral category regressing out the other two (including Animals), correlated with total regional and laminar cortical depths. Significant correlations bolded.

		Total Regional Thickness	Superficial Cortical Subdivision	Middle Cortical Subdivision	Deep Cortical Subdivision
Zero-Order Correlations	Faces	-0.83	-0.38	-0.28	-0.76
	Vehicles	0.70	0.44	0.48	0.45
Partial Correlations	Faces	-0.90	-0.28	-0.15	-0.73
	Vehicles	0.83	0.37	0.43	0.33

Table 3: Fit Indices for Regression Models Predicting Cortical Thickness. $-2LL = -2 * \log$

likelihood of the model. This index is used in the computation of likelihood ratio tests comparing nested models and information indices. AIC = Akaike Information Criterion. AICc = AIC with small-sample correction. BIC = Bayesian Information Criterion. As lower values of information indices indicate better fit, all favor M2 (in bold). Bayes Factors compare M2 to each of the other models. All BF's indicate greater support for M2 than all other models.

Model	Description	-2LL	AIC	AICc	BIC	M2/M# Bayes Factor
M ₀	Layer Only	43.97	61.97	68.18	76.95	161.04
M ₁	M ₀ + Main effects for FR and VR	13.51	35.51	45.29	53.81	38.85
M ₂	M ₁ + L _{DvsSM}	7.56	31.56	43.56	51.53	-----
M ₃	M ₁ + Layer x FR Interaction	7.54	33.56	48.12	55.18	3.07
M ₄	M ₁ + Layer x VR Interaction	12.52	38.52	53.09	60.15	139.52
M ₅	M ₁ + Layer x FR + Layer x VR Interactions	7.34	37.34	58.21	62.29	19.90

Table 4: Unstandardized and Standardized Regression Coefficients from Model 2. Model 2 is the best-fitting regression model that predicts that the Face Recognition is selectively associated with CT in the deep layer but that Vehicle Recognition has the same association with CT in each layer. Unstandardized regression coefficients are on top and standardized regression coefficients are below in parentheses. * = $p < .001$

Layer	Face Recognition	Vehicle Recognition
Superficial	-.07 (-.10)	.18* (.27)
Middle	-.07 (-.10)	.18* (.27)
Deep	-.62* (-.57)	.18* (.27)

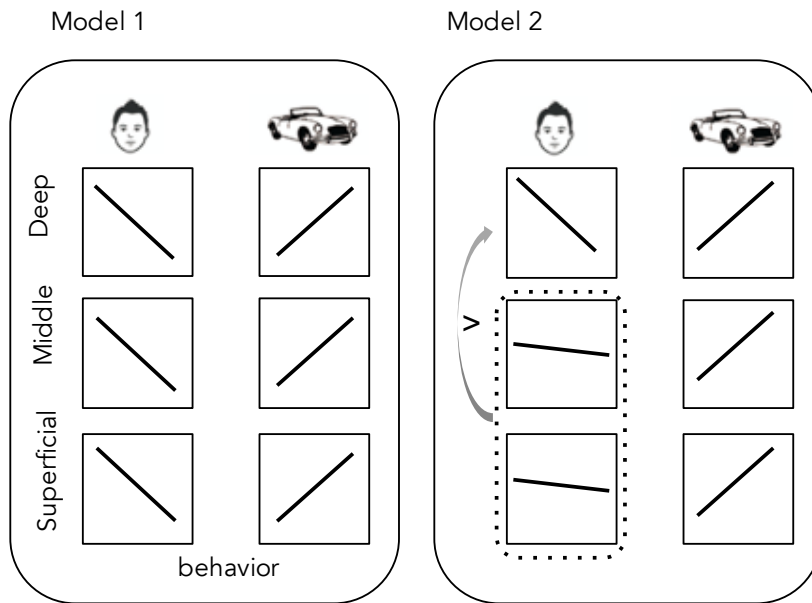


Figure 1. Composite Models tested. Model 1 specified that correlations and regression coefficients between face recognition and cortical thickness are negative in sign and equal across MR cortical layers and that correlations and regression coefficients between vehicle recognition and cortical thickness are positive in sign and equal across layers. Model 2 specified a layer-specific pattern for the associations between face recognition and cortical thickness across layers, with the deep MR layer having a stronger, more negative pattern of association than the superficial and middle MR layers. Like Model 1, Model 2 specified an invariant of association between vehicle recognition and CT across layers. As described in the text, we also tested three additional substantive models positing different patterns of association.

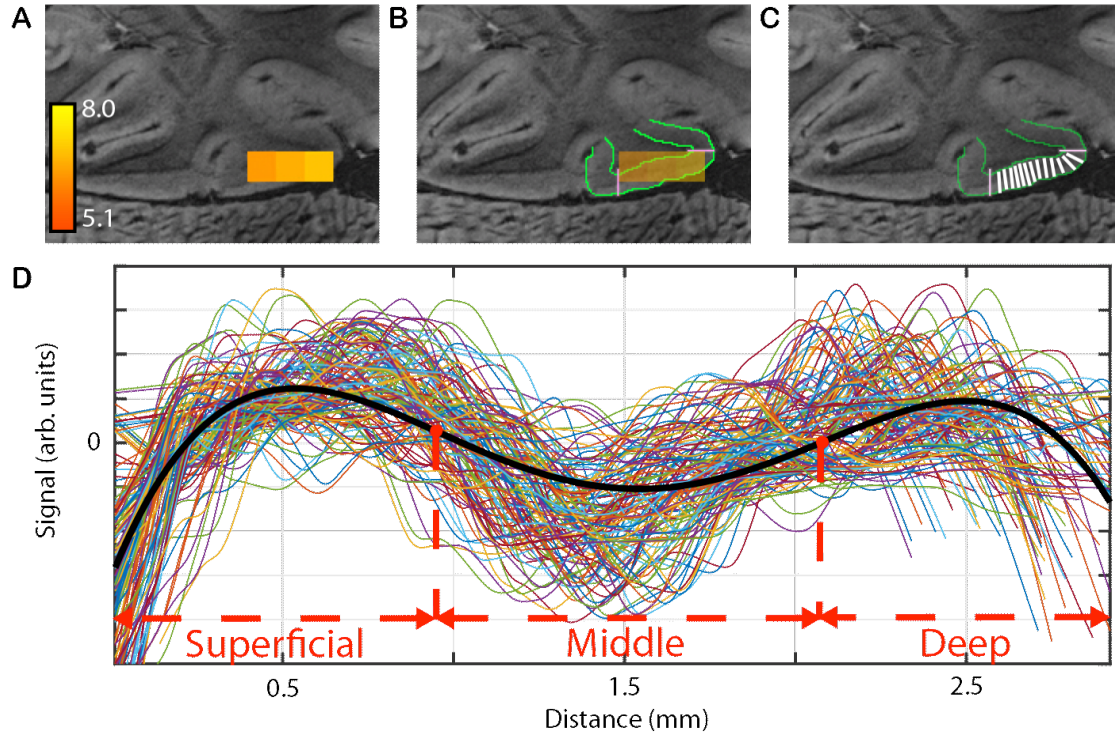


Figure 2. Trace acquisition, intensity sampling and layer identification. (A) rFFA2 was defined functionally with a Face > Object contrast and at a fixed size of 108 mm^3 in each subject, then co-registered onto the ultra-high resolution slices. Color bar and values represent the face selectivity measured from the 3mm scan. (B) The edges of the rFFA2 were traced (magenta) over the underlying cortical mantle. Boundaries between grey matter (GM) and cerebrospinal fluid (CSF) (superficial layer boundary), and between GM and white matter (WM) (deep layer boundary), were manually drawn (green). (C) Traces were acquired as the series of lines originating at each superficial boundary voxel and terminating in each deep boundary voxel, and vice versa. Gray matter intensity was sampled across the length of the trace. (D) The line plot visualizes all traces shown in varying colors contributing to one slice of one subject's rFFA2. A 4th-order polynomial (bold black curve) was fit to an average trace, with points of inflection defining the depths at which image intensity changed.

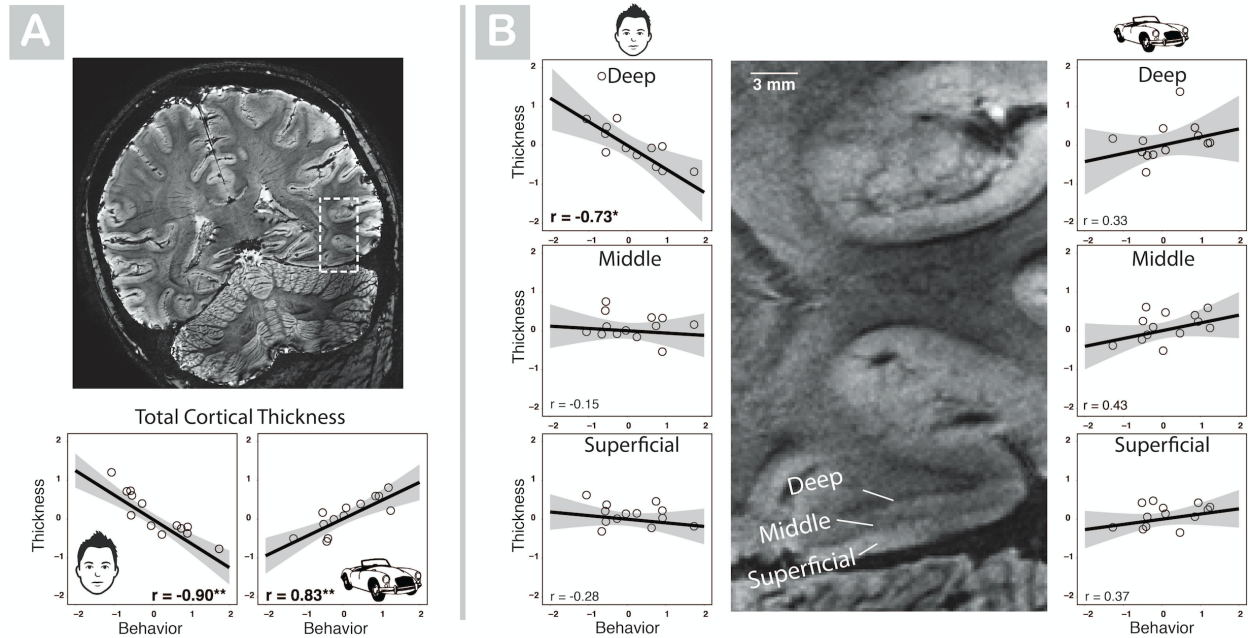


Figure 3. The relation between grey matter thickness of rFFA2, or its laminar subdivisions, and behavioral performance. Behavioral performance on the x-axes and cortical thickness and depth measurements on the y-axes show residualized values (see text for actual values in mm). The scatterplots on the left (with face icons) show partial correlations with Face recognition controlling for performance with Vehicles, while scatterplots on the right (with car icons) show partial correlations with Vehicles controlling for Face recognition. **(A)** The brain is tilted as a result of slices being aligned perpendicular to the rFFA2. Shown is one slice for one subject showing the average of six susceptibility-weighted scan acquisitions. The dashed box is centered on the functionally-defined rFFA2. Below the whole brain slice are scatterplots depicting the relation between total regional CT of rFFA2 and behavioral performance with Faces (left) and Vehicles (right). Error bars represent the 95% confidence intervals. **(B)** The central inset shows a magnified view from the box in (A), centered on rFFA2 for one slice in one subject, with the three laminar subdivisions perceivable to the naked eye. Scatterplots represent partial correlations between the thickness of Deep, Middle or Superficial MR layers as a function of face and vehicle recognition, controlling for performance with the other category. Error bars show the 95% confidence intervals.

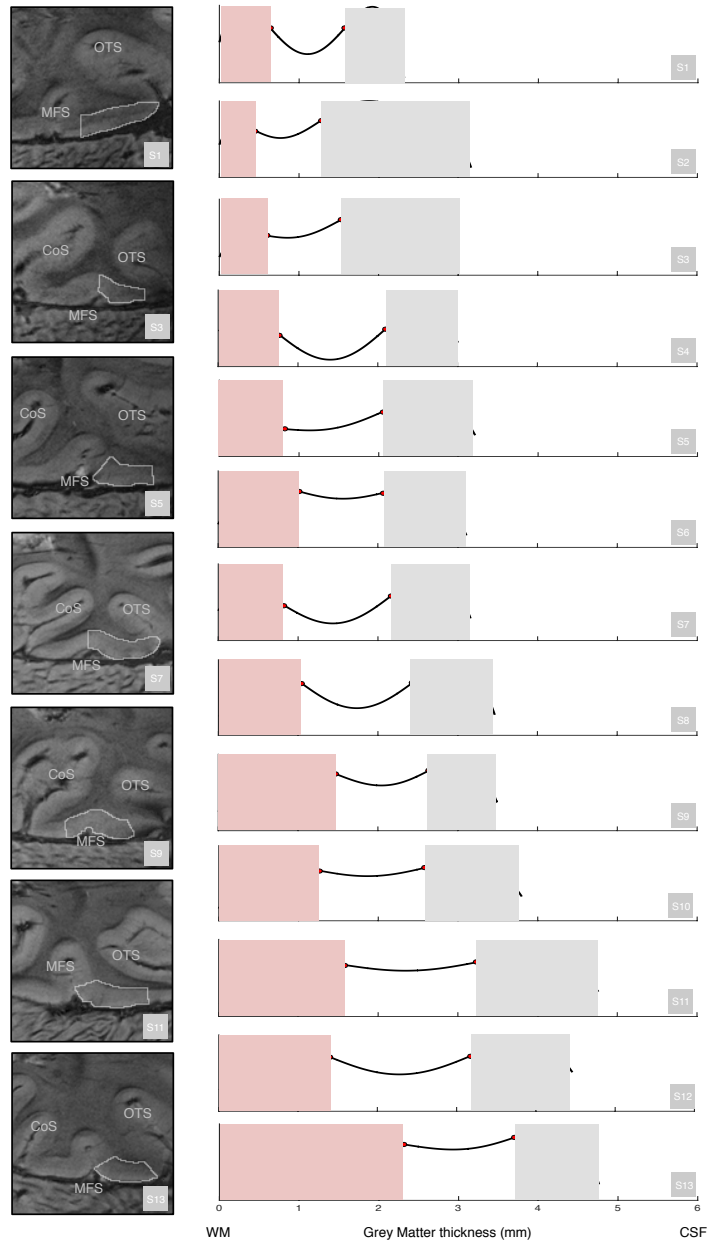


Figure 4. Plots of the 4th order polynomial fit to the average rFFA2 intensity curve of the center-most slice for each subject. Subjects are rank-ordered from highest (S1) to lowest based on face recognition performance, regressing out performance with vehicles. Inflection points marking the transitions from deep MR layers (red shading) to middle MR layers (no shading) to superficial MR layers (grey shading) are marked with red dots. Axis labels: white matter (WM), cerebrospinal fluid (CSF). For every other subject, the centermost slice of rFFA2 is shown to the left. The spread of face-selective activation along the fusiform gyrus is outlined in grey. Labels: Occipital temporal sulcus (OTS), Collateral Sulcus (CoS), Middle Fusiform Sulcus (MFS). The rFFA2 most often fell on the lateral aspect of the fusiform gyrus, between MFS and OTS, corresponding to cytoarchitectonic FG4.