Cognitive insights from evolutionarily new brain structures in prefrontal cortex

Willa I. Voorhies[2], Jacob A. Miller[1], Jewelia K. Yao[2], Silvia A. Bunge,[1,2*] & Kevin S. Weiner[1,2*]

[1] Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley CA, 94720 USA

[2] Department of Psychology, University of California, Berkeley, Berkeley CA, 94720 USA

* Shared senior authorship

Corresponding Author:

Willa I. Voorhies, Department of Psychology, 2121 Berkeley Way West, UC Berkeley, Berkeley CA, 94720, wvoorhies@berkeley.edu

Acknowledgments:

This research was supported by a T32 HWNI training grant and an NSF-GRFP fellowship (WIV), as well as start-up funds from UC Berkeley (KSW). Funding for the original data collection and curation was provided to SAB by NINDS R01 NS057156 and NSF BCS1558585. Data analysis was supported by NICHD R21HD100858, awarded to KSW and SAB. We thank Ishana Raghuram for assistance with sulcal labeling and former members of the Bunge laboratory for assistance with data collection, and the families who participated in the study.

ABSTRACT

While the disproportionate expansion of lateral prefrontal cortex (LPFC) throughout evolution is commonly accepted, the relationship between evolutionarily new LPFC brain structures and uniquely human cognitive skills is largely unknown. Here, we tested the relationship between variability in evolutionarily new LPFC tertiary sulci and reasoning skills in a pediatric cohort. A novel data-driven approach in independent *discovery* and *replication* samples revealed that the depth of specific LPFC tertiary sulci predicts individual differences in reasoning skills beyond age. These findings support a classic, yet untested, theory linking the protracted development of tertiary sulci to late-developing cognitive processes. We conclude by proposing a mechanistic hypothesis relating the depth of LPFC tertiary sulci to anatomical connections. We suggest that deeper LPFC tertiary sulci reflect reduced short-range connections in white matter, which in turn, improve the efficiency of local neural signals underlying cognitive skills such as reasoning that are central to human cognitive development.

Introduction

A fundamental question in cognitive neuroscience is how the structure of the brain supports complex cognition. While much progress has been made in answering this question, especially in animal models, human brains differ in both their micro- and macrostructural properties from widely used animals in neuroscience research such as mice, marmosets, and macaques[1]. These cross-species differences are especially pronounced in association cortices such as lateral prefrontal cortex (LPFC). LPFC is a late-developing cortical expanse that is enlarged in humans compared to non-human primates[2] and is critical for cognitive control, executive function, reasoning, and goal-directed behavior[3–6]. Yet there is still much progress to be made in understanding how the development of evolutionarily new brain structures in the expanded human LPFC support the development of complex, largely human, cognitive skills achieved by neural circuits within LPFC.

Of all the cognitive skills and anatomical features to focus on, we investigate the relationship between relational reasoning and overlooked macroanatomical structures known as tertiary sulci. We chose to focus on relational reasoning, which is the ability to extract common features across objects and conceptualize them in terms of their relation to each other[6,7], because humans consistently outperform other species in relational reasoning tasks[6,8], and developmental improvements in this domain are strongly associated with anatomical changes in LPFC[7,9,10]. We chose to focus on tertiary sulci, which are the shallowest class of cortical folds, because they have been largely overlooked due to methodological difficulties in their identification (which we expand on further below)[11–13]. Due to these difficulties, very little is known regarding the role of tertiary sulci in human cognition despite the fact that many tertiary sulci are evolutionarily new structures that are uniquely human[14].

Additionally, as both reasoning skills and tertiary sulci exhibit protracted developmental trajectories in childhood, they serve as ideal targets to test a classic, yet largely unconsidered theory. Specifically, Sanides proposed that morphological changes in tertiary sulci would likely be associated with the protracted development of higher-order thinking and cognitive skills[15,16]. Fitting these criteria, relational reasoning skills continue to develop throughout childhood, while tertiary sulci emerge last (after primary and secondary sulci) in gestation and also show a protracted development that continues after birth for a still undetermined period of time[14,16–22]. A relationship between relational reasoning and tertiary sulcal morphology would build on previous findings relating the development of relational reasoning to changes in LPFC cortical thickness and structural connectivity[23,24]. Furthermore, relational reasoning supports complex problem solving and scaffolds the acquisition of additional cognitive skills in children[25,26]. Thus, exploring if or how tertiary sulci contribute to the development of this cognitive skill may not only provide insight to a classic theory, but also advance understanding of the anatomical features underlying variability in the development of a wide range of other cognitive skills.

To date, no study (to our knowledge) has tested the role of tertiary LPFC sulci in cognitive development. This gap likely persists for three key reasons. First, previous studies examining individual differences in the development of reasoning and anatomical variability in human LPFC[7] implemented analyses that were averaged across individuals on standard neuroanatomical templates, which obscure tertiary sulci in LPFC[13] (Supplementary Fig. 2). Therefore, to precisely characterize the relationship between tertiary sulcal morphology in LPFC and reasoning performance, it is necessary to consider cortical anatomy at the level of the individual. Second, the shallowness of tertiary sulci makes them hard to reliably identify in post-mortem tissue—typically considered the gold standard for anatomical analyses—because they are easily confused with

shallow indentations produced by veins and arteries on the outer surface of the cerebrum[11]. Consequently, neuroanatomical atlases and neuroimaging software packages largely exclude tertiary sulci. In turn, tertiary sulci in LPFC are excluded from most developmental cognitive neuroscience studies until the present study. Third, a common misconception is that macro-anatomical structures such as sulci and gyri are functionally and cognitively relevant in primary, but not association, cortices. For example, the calcarine sulcus predicts the location of the primary visual cortex[27] and a "knob" in the pre-central gyrus predicts the motor hand area in the primary motor cortex[28,29]. However, there is increasing evidence that tertiary sulci are functionally relevant in association cortices such as ventral temporal cortex[30] (VTC), medial PFC[21], and LPFC[13] in adults, as well as behaviorally meaningful in medial PFC[31,32]. Despite this mounting evidence that tertiary sulci are functionally and behaviorally relevant in association cortices within adults, it is largely unknown whether morphological features of tertiary sulci will predict individual differences in behavior and cognition in a developmental cohort.

To address this gap in knowledge, we characterized LPFC tertiary sulci for the first time in individual subjects in a broad developmental sample including participants between 6 and 18 years old. We leveraged the neuroanatomical and cognitive variability intrinsically present in the sample to ask whether variability in tertiary sulcal morphology predicts individual differences in relational reasoning. As sulcal depth is a defining feature of tertiary sulci, which are shallower than primary and secondary sulci[1,14,16–20,30,33], we hypothesized a relationship between the depth of tertiary sulci and reasoning skills.

To characterize this relationship, we developed a novel pipeline that combines the most recent anatomical definition of LPFC tertiary sulci[14] with data-driven analyses to model sulcal morphological features and reasoning performance. Our approach addresses four main questions:

1) Are LPFC tertiary sulci identifiable in a developmental sample, and are they smaller and shallower than primary and secondary LPFC sulci as in adults[13,14]? 2) Is there a relationship between the depth of LPFC tertiary sulci and reasoning performance across individuals? 3) If so, can we construct a neuroanatomical-behavioral model to predict an individual's reasoning score from tertiary sulcal depth and age in an independent sample? 4) If successful, does a neuroanatomical-behavioral model extend to other sulcal features or cognitive tasks? Answering these questions offers the first link between tertiary LPFC sulcal morphology and reasoning, as well as provides novel cognitive insights from evolutionary new brain structures in LPFC.

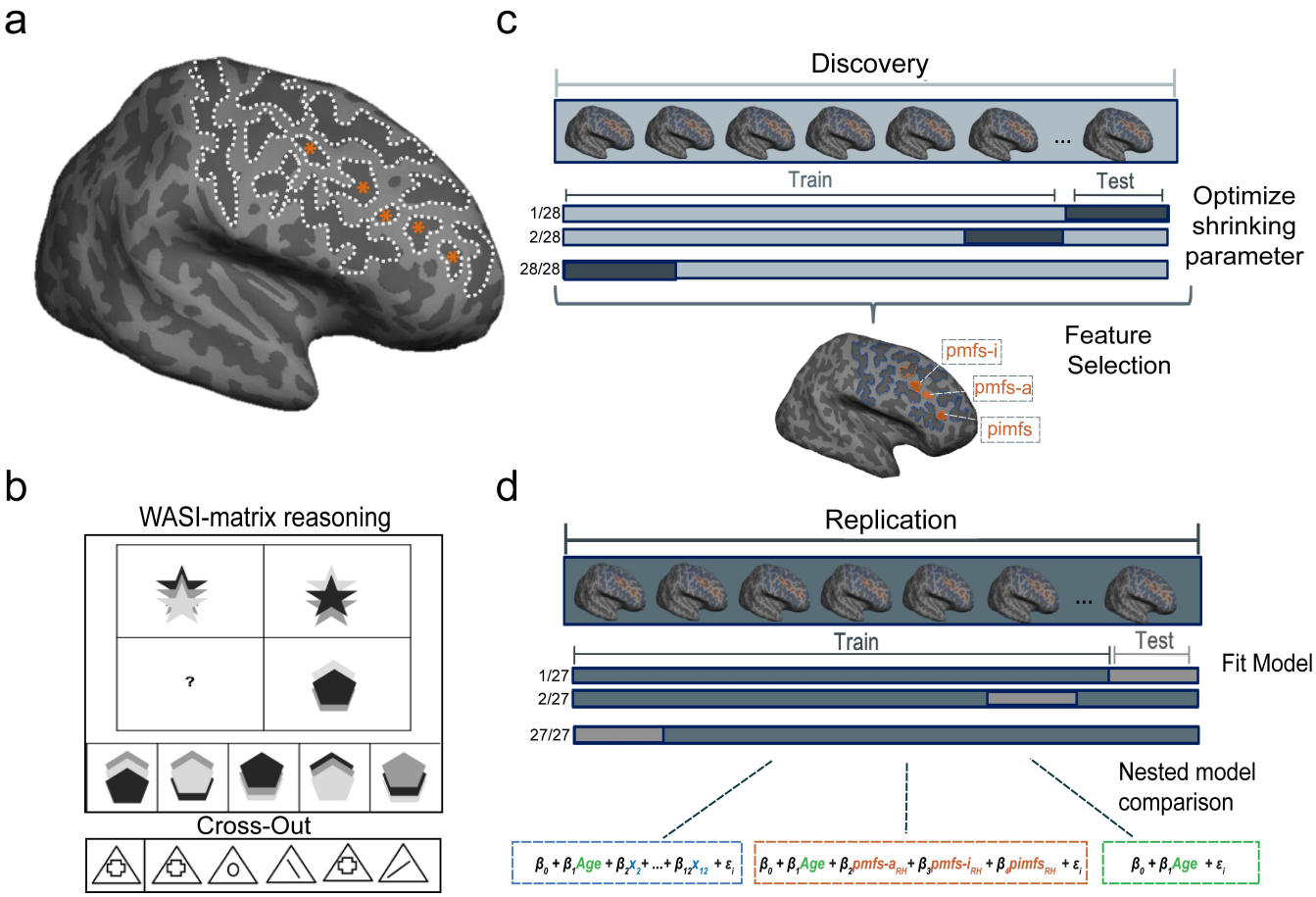


Fig. 1. ***A novel, data-driven analysis pipeline with Discovery and Replication samples that models the relationship between LPFC sulcal morphological features and reasoning performance.*** **a.** An inflated cortical surface reconstruction of a right hemisphere from one example subject. Dotted white outlines show manually labeled sulci. Asterisks indicate the frequently omitted or misclassified tertiary sulci (Supplementary Fig. 1 for all subjects). **b.** *Top:* Example from the standardized test used to assess relational reasoning in this study (WISC-IV, Matrix Reasoning task). In this task, participants are instructed to complete the matrix so that the relation between the two bottom shapes mirrors the relation between the two top shapes. In this example, option 4 completes the pattern. *Bottom:* Example from the processing speed task (WJ-R, Cross-Out test), which serves as a behavioral control. In this task, participants are instructed to cross out all objects that match the object on the left as quickly as possible. **c.** *Feature selection – Discovery sample.* A LASSO regression was performed in the *Discovery* sample to determine which sulci, if any, were associated with Matrix Reasoning performance. The model parameters were fit iteratively using a leave-one-out cross-validation procedure (Online Methods). **d.** *Model evaluation – Replication sample.* The sulci selected from the LASSO regression (orange; *pmfs*-i_{RH}, *pmfs*-a_{RH}, and *pimfs*$_{RH}$) were included along with age in a model to predict task performance in the *Replication* sample. In order to assess the unique contribution of the selected sulci to task performance, this model *(orange)* was compared to two nested alternate models: (1) age alone *(green)* and (2) age in addition to all 12 LPFC sulci *(blue)*. All models were fit with a leave-one-out cross-validation procedure.

Results

Tertiary sulci are consistently identifiable in the LPFC of 6-18 year-olds

Our sample consisted of 61 typically developing children and adolescents ages 6-18 years old. Participants were randomly assigned to *Discovery* (N = 33) and *Replication* (N = 28) samples with comparable age distributions (*Discovery*: mean(sd) = 12.0 (3.70); *Replication*: mean(sd) = 12.32 (3.53); $p = 0.81$). For each subject, we generated cortical surface reconstructions in FreeSurfer[34–36] from high-resolution T1-weighted anatomical scans. As current automated methods do not define LPFC tertiary sulci and often include gyral components in sulcal definitions (Supplementary Fig. 2), all sulci were manually defined on the native cortical surface for each subject according to the most recent and comprehensive atlas of LPFC sulcal definitions[14] (Fig. 2). We focused our analyses on the region commonly referred to as the dorsal LPFC, which is bounded posteriorly by the central sulcus (*cs*), anteriorly by the horizontal (*imfs-h*) and ventral (*imfs-v*) components of the intermediate frontal sulcus, superiorly by the two components of the superior frontal sulcus (*sfs-p* and *sfs-a*), and inferiorly by the inferior frontal sulcus (*ifs*). Throughout the paper, we refer to this region as the LPFC (Fig. 2a). Studies in adults report as many as five tertiary sulci within these anatomical boundaries[14]: the three components of the posterior middle frontal sulcus (posterior: *pmfs-p*; intermediate: *pmfs-i*; anterior: *pmfs-a*) and the two components of the para-intermediate

frontal sulcus (ventral: *pimfs-v*; dorsal: *pimfs-d*). We defined sulci on the inflated and pial cortical surfaces of each hemisphere for each participant (Online Methods).

Discovery sample: All primary and secondary sulci (*cs*; the superior (*sprs*) and inferior (*iprs*) portions of the precentral sulcus; *sfs-p*; *sfs-a*; *ifs*; *imfs-h*; *imfs-v*) were identifiable in both hemispheres of each individual subject. For the first time, we demonstrate that tertiary sulci in LPFC are also consistently identifiable within the hemispheres of pediatric participants as young as 6 years old (Fig. 2b). Qualitatively, we observed extensive variability in the precise location of LPFC tertiary sulci (Supplementary Fig. 1 for sulcal definitions in all 122 hemispheres from both samples). The three components of the posterior middle frontal sulcus (*pmfs-p; pmfs-i; pmfs-a*) were identifiable in all subjects in every hemisphere. However, the most anterior LPFC tertiary sulcus, the *para-intermediate frontal sulcus* (*pimfs*), was consistently variable across individuals (Supplementary Table 1). Specifically, while almost all subjects had at least one identifiable component of the *pimfs* (right hemisphere: 30/33; left hemisphere: 31/33), we were only able to identify both dorsal and ventral *pimfs* components in 48.3% of all subjects (right hemisphere: 12/33; left hemisphere: 16/33).

Replication sample: Consistent with the *Discovery* Sample, all primary and secondary sulci (numbered 1-8 in Fig. 2b) could be identified in both hemispheres of each individual subject. In terms of tertiary sulci, the *pmfs-p*, *pmfs-i*, and *pmfs-a* (numbered 9-11 in Fig. 2b) were also identifiable in each hemisphere of every individual. Once again, the *pimfs* was the most variable across individuals (Supplementary Fig. 1b; Supplementary Table 1). We were able to identify at least one *pimfs* component in almost every subject (right hemisphere: 28/28; left hemisphere: 27/28). Both the dorsal and ventral *pimfs* components were identifiable in 76.8% of hemispheres (right hemisphere: 19/28 subjects; left hemisphere: 24/28; Supplementary Table 1).

In sum, we could identify LPFC tertiary sulci in both *Discovery* and *Replication* samples. However, we could not identify both dorsal and ventral *pimfs* components in each hemisphere. Thus, our inclusion criteria for all subsequent analyses was to include subjects who had at least one *pimfs* component in each hemisphere (*Discovery*: 28/33, *Replication*: 27/28), which assures that all repeated measures statistics are balanced for effects of sulcus and hemisphere.

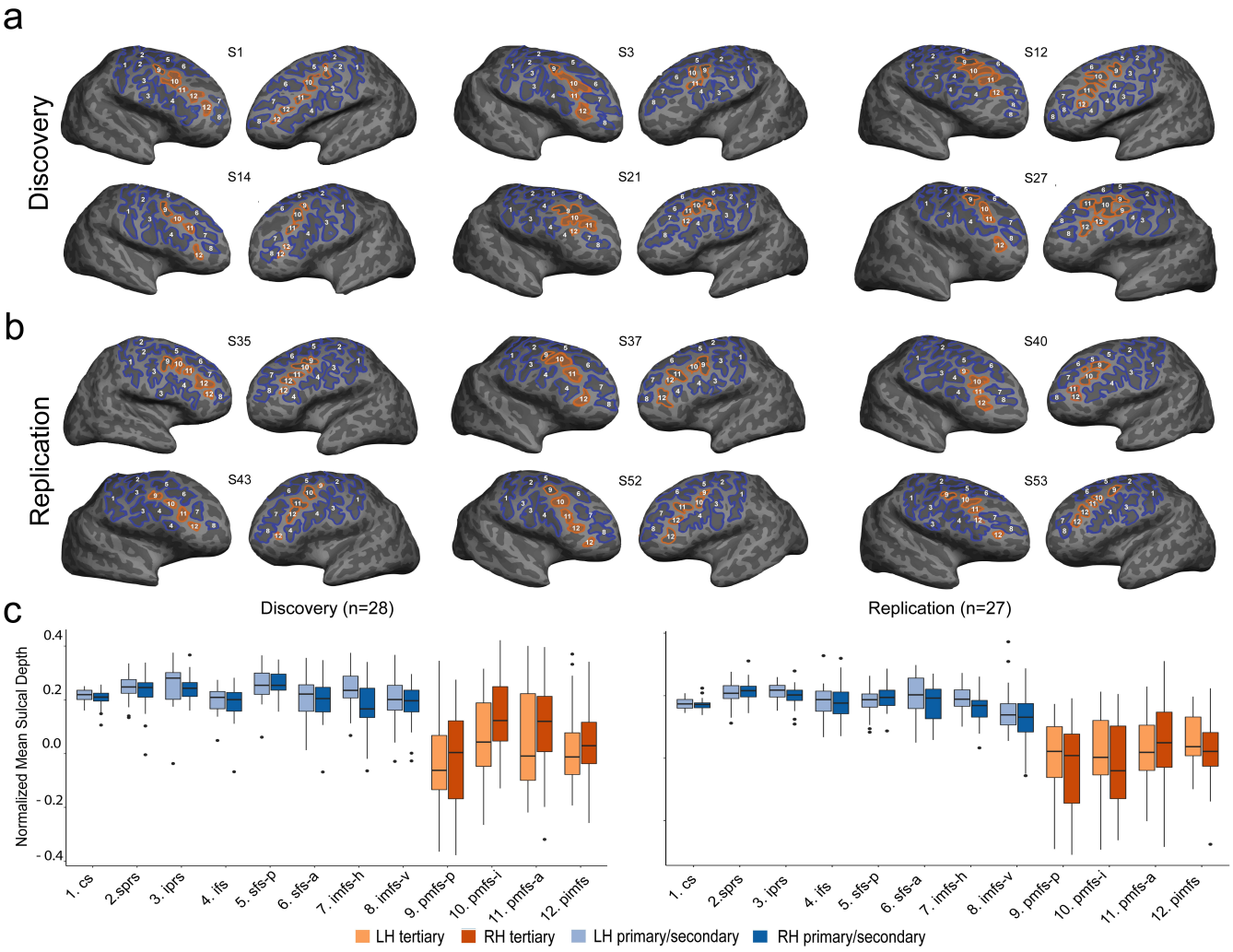


Fig. 2. ***LPFC tertiary sulci in 6-18 year-olds are more shallow and more variable than primary/secondary sulci.*** a. LPFC sulcal definitions on inflated cortical surface reconstructions from six example subjects in the Discovery sample. Sulci were identified based on the most recent neuroanatomical atlas to consider tertiary sulci14. Primary and secondary sulci (1-8) are in blue, while tertiary sulci (9-12) are in orange. The three tertiary sulci (pimfs-iRH (10), pimfs-aRH (11), and pimfsRH (12)) identified by our model-based approach with cross-validation (Fig. 3) are filled in. We emphasize that 1,320 manual labels were created in total to examine the relationship between LPFC sulcal depth and reasoning performance (Supplementary Fig. 1 for sulcal definitions in all 122 hemispheres). b. Same layout as in a., but for six example subjects in the Replication sample. c. Normalized sulcal depth for each of the 12 LPFC sulci in the Discovery (left) and Replication (right) samples. Tertiary sulci (orange) were shallower and more variable than primary/secondary sulci (blue) in both samples.

LPFC tertiary sulci are shallower and more variable than primary/secondary sulci in children

As depth is the key morphological feature differentiating tertiary from primary and secondary sulci[16–19], and recent findings show that LPFC tertiary sulci are shallower and more variable than primary/secondary sulci in adults[37], we next sought to compare the depth and variability of tertiary and primary/secondary sulci in children. Sulcal depth was normalized to the maximum depth value within each individual hemisphere in order to account for differences in brain size across individuals and hemispheres (Online Methods). From these normalized measures, we conducted a 2-way repeated-measures analysis of variance (rm-ANOVA) to statistically test for differences between sulcal type (*tertiary vs. primary/secondary*) and hemisphere (*left vs. right*).

Discovery sample: Consistent with findings in adults[37], we observed a main-effect of sulcal type ($F(1,27)= 95.63$, $p<10^{-3}$, $\eta^2_G = 0.35$) in which tertiary sulci were significantly more shallow than non-tertiary sulci ($\text{Mean(sd)}_{\text{Tertiary}}= 0.04(0.17)$; $\text{Mean(sd)}_{\text{Primary/secondary}}= 0.23(0.07)$). We also observed an interaction between sulcal type and hemisphere ($F(1,27) = 5.67$, $p<0.02$, $\eta^2_G= 0.01$) in which tertiary sulci were significantly deeper in the right hemisphere than in the left hemisphere ($\text{Mean(sd)}_{RH} = 0.06(0.17)$; $\text{Mean(sd)}_{LH}= 0.02(0.1)$). In contrast, the depth of primary/secondary sulci did not differ between hemispheres ($\text{Mean(sd)}_{RH}= 0.21(0.07)$; $\text{Mean(sd)}_{LH} = 0.23(0.07)$); Fig. 2c).

Replication sample: We observed the same main effect of sulcal type in the *Replication* sample. Tertiary sulci were more shallow than primary/secondary sulci ($F(1,26) = 136.5$, $p<10^{-3}$, $\eta^2_G = 0.46$; $\text{Mean(sd)}_{\text{Tertiary}}= 0.02(0.16)$; $\text{Mean(sd)}_{\text{Primary/secondary}} = 0.23(0.07)$). We did not observe an interaction with hemisphere in this sample ($F(1,26) = 0.26$, $p = 0.62$); Fig. 2c).

Additionally, while age was correlated with reasoning performance in both *Discovery* ($r = 0.58$, $p<10^{-3}$) and *Replication* samples ($r = 0.73$, $p<10^{-3}$), there was an inconsistent relationship between sulcal depth and age in either sample (Supplemental Fig. 3). Thus, we next implemented

a two-pronged, model-based approach to test if including sulcal depth predicted reasoning skills above and beyond age.

A model-based approach with nested cross-validation reveals that including the depth of three LPFC tertiary sulci predicts individual variability in reasoning skills above and beyond age alone

To examine the relationship between LPFC sulcal depth and reasoning skills, we first implemented a feature selection technique in the *Discovery* sample (Fig. 1c) to determine if the depths of any LPFC sulci are associated with reasoning performance (to remind the reader, we use depth in the model as depth is the main morphological feature differentiating tertiary from primary/secondary sulci). To do so, we submitted sulcal depth values for all 12 LPFC sulci in the *Discovery* sample to a LASSO-regression model, which provides an automated method for feature selection by shrinking model coefficients and removing sulci with very low coefficients from the model (Fig. 1c; Online Methods). This approach allowed us to determine, in a data-driven manner, which sulci are the strongest predictors of reasoning performance. To determine the value of the shrinking parameter (α), we iteratively fit the model with a range of α-values using cross-validation. By convention[41], we selected the α that minimized the cross-validated Mean Squared Error (MSE_{CV}; Fig. 3a). Although both tertiary and primary/secondary sulci were initially included as predictors, after implementing the LASSO regression, only three tertiary sulci (*pmfs-i*, *pmfs-a*, and *pimfs*) in the right hemisphere were found to be associated with reasoning performance (MSE_{CV}=21.84, $\alpha = 0.1$; $\beta_{pmfs\text{-}i}$= 4.50, $\beta_{pmfs\text{-}a}$= 1.78, β_{pimfs}= 11.88; Fig. 3a).

To test the generalizability of the sulcal-behavioral relationship identified in the *Discovery* sample, we constructed a linear model to predict reasoning from sulcal depth and age in our *Replication* sample. As age is highly predictive of reasoning ability (Supplemental Fig. 3), including age in this model allows us to compare performance of this tertiary sulci + age model to

a model with age alone in order to determine the unique contribution of LPFC tertiary sulcal depth to reasoning performance above and beyond age. This model (and all subsequent models) were fit using a leave-one-out cross-validation (looCV) procedure. LooCV, while appropriate for smaller sample sizes, can result in models with high variance, which can lead to overfitting. To address this concern, we estimated empirical MSE confidence intervals using a bootstrapping procedure (Online Methods; Fig. 3d). The mean depths of the *pmfs-i*$_{RH}$, *pmfs-a*$_{RH}$, and *pimfs*$_{RH}$, as well as age, were included as predictors in the model as they were the only three sulci identified in the sulcal-behavioral model in the *Discovery* sample. This model was highly predictive of reasoning score in the *Replication* sample (R^2_{cv}= 0.52, MSE_{CV} = 9.66; Bootstrapped 95% CI_{MSE}: 3.12-13.69, $median_{MSE}$ = 8.14). Additionally, we observed a high correspondence (Spearman's *rho* = 0.70) between predicted and actual measured reasoning scores (Supplementary Fig. 4). Furthermore, if we consider just the two LPFC tertiary sulci that are the strongest predictors of reasoning performance as identified in the *Discovery* sample (*pmfs-i*$_{RH}$: $\beta_{\text{pmfs-i}}$ = 4.50; *pimfs*$_{RH}$: β_{pimfs} = 11.88), the predictions of reasoning performance and model fits improved even further in the *Replication* sample: R^2_{cv}= 0.58; MSE_{CV} = 8.52; Bootstrapped 95% CI_{MSE} = 3.21-12.37, $median_{MSE}$ = 7.47; Spearman's *rho* = 0.73; Fig. 3b).

We used cross-validation to evaluate the fit of this model relative to two alternative models considering either 1) age alone or 2) sulcal depth from all right hemisphere LPFC sulci and age together in the *Replication* sample (Fig. 1d). This nested model comparison allowed us to determine the unique contribution of tertiary sulcal depth (from the sulci identified using the LASSO regression in the *Discovery* sample) while still accounting for the effects of age and primary/secondary LPFC sulcal depth on reasoning performance. Removing the *pmfs-i*$_{RH}$, *pmfs-a*$_{RH}$, and *pimfs*$_{RH}$ from the model decreased prediction accuracy and increased the MSE_{cv} (R^2_{cv} = 0.48, MSE_{CV} = 10.50; Bootstrapped 95% CI_{MSE} = 4.69-15.67, $median_{MSE}$ = 9.66), indicating that

the depths of these right hemisphere tertiary sulci uniquely contribute to the prediction of reasoning scores above and beyond age (Fig. 3c). Additionally, considering age and the depths of all RH LPFC sulci weakened the model prediction and increased MSE_{cv} (R^2_{cv} = 0.14, MSE_{CV} = 17.47, Bootstrapped 95% CI_{MSE} = 2.79-306.25, $median_{MSE}$ = 19.70). The bootstrapped CI_{MSE} showed that this model also suffered from very high variance (Fig. 3d). Taken together, our cross-validated, nested model comparison empirically supports that the depth of only a subset of LPFC tertiary sulci explain unique variance in reasoning performance that is not accounted for by age or the depths of all LPFC sulci.

Finally, to test whether our findings generalized to other neuroanatomical features or cognitive domains commonly explored in the field, we repeated our procedure with 1) a model in which we replaced sulcal depth with cortical thickness[42–45] and 2) a model in which we replaced reasoning performance with processing speed (as measured by cross-out score[38]; Fig. 1b). To provide a quantitative comparison of the models, we used the Akaike Information Criterion (AIC). If the ΔAIC is greater than 2, it suggests an interpretable difference between models. If the ΔAIC is greater than 10, it suggests a strong difference between models, with the lower AIC value indicating the preferred model[46,47] (Online Methods). This approach revealed that a model with cortical thickness and age was predictive of reasoning (R^2_{cv} = 0.33; MSE_{CV} = 13.54), but much less than age alone (R^2_{cv} = 0.48; MSE_{CV} = 10.50). Additionally, sulcal depth of the three critical LPFC tertiary sulci (*pmfs-*i_{RH}, *pmfs-*a_{RH}, and *pimfs*$_{RH}$) and age was predictive of processing speed (R^2_{cv}= .45; MSE_{CV} =20.53), but not much more than age alone (R^2 = 0.42; MSE_{CV} = 21.82). Finally, the AICs for both models ($AIC_{Thickness}$ = 78.58; $AIC_{Cross\text{-}Out}$ = 89.59) were much higher than the AIC for the tertiary sulci + age model ($AIC_{sulcalDepth}$ = 63.85, $\Delta AIC_{Thickness\text{-}Depth}$ = 14.73, $\Delta AIC_{Cross\text{-}out\text{-}MatrixReasoning}$ = 25.74). This indicates that sulcal depth is strongly preferred as a predictor over cortical thickness (Supplementary Fig. 6a), and reasoning is strongly preferred over processing speed, respectively (Supplementary Fig. 6b), in our models.

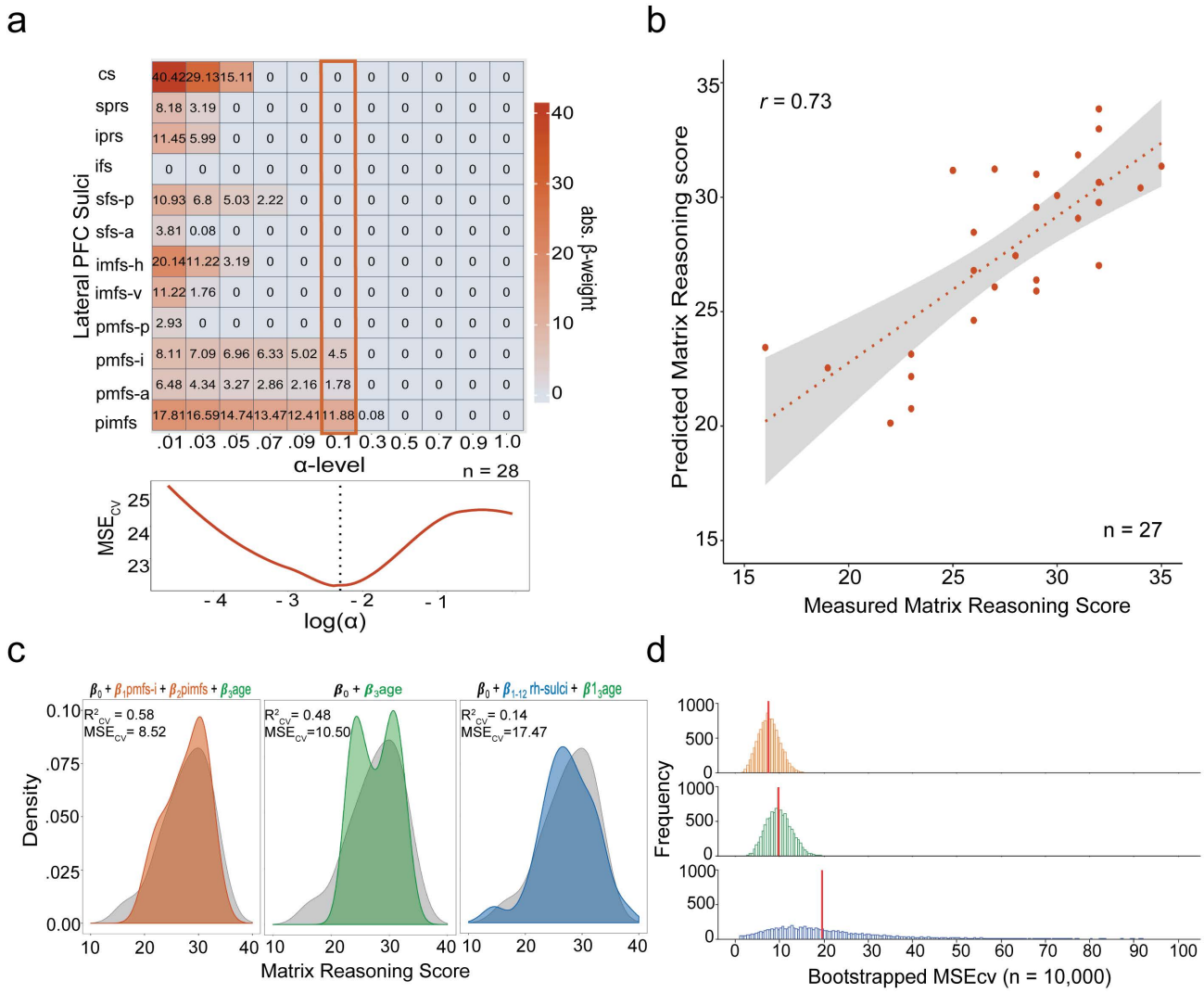


Fig. 3. ***A model-based approach with nested cross-validation reveals that the depth of a subset of LPFC tertiary sulci predicts individual variability in reasoning above and beyond age.*** **a**. Results from the LASSO regression predicting Matrix Reasoning score from sulcal depth in the *Discovery* sample. *Top:* Beta-coefficients for each sulcus at a range of shrinking parameter (alpha) values. Highlighted box indicates coefficients at the chosen alpha-level. *Bottom:* Cross-validated MSE at each alpha-level. By convention[41], we selected the α that minimized the cross-validated Mean Squared Error (MSE_{CV}; dotted line). **b.** Spearman's correlation between measured and predicted Matrix Reasoning scores in the *Replication* sample for the best tertiary sulci + age model, which includes the depths of the two most predictive sulci (pmfs-i_{RH} + $pimfs_{RH}$) from the *Discovery* sample, as well as age (Supplementary Fig. 4 for a model with all three tertiary sulci selected from the *Discovery* sample). **c**. Density plots comparing model fits for three nested models evaluated in the *Replication* sample. The measured Matrix Reasoning score is shown for the distribution of the predicted scores for each of the three models. *orange*: pmfs-i_{RH} + $pimfs_{RH}$ + age. *green*: age only. *blue*: all RH LPFC sulci + age. Each of the three model fits are overlaid on the distribution of measured Matrix Reasoning scores *(gray*). The pmfs-i_{RH} + $pimfs_{RH}$ + age model (*orange*) produced the best fit. **d.** Empirical MSE for each of the three models estimated with a bootstrapping procedure ($n_{iterations}$ = 10,000) to address the potential for looCV to result in high variance and overfitting. The all LPFC RH sulci + age model (*blue*) exhibited notably high variance in error estimation. The red vertical line indicates the estimated median MSE.

Discussion

Recent studies examining sulcal morphology in humans and other species continue to improve our understanding of the development and evolution of association cortices. They also provide anatomical insights into cognitive skills that set humans apart from other species[13,33,48,49]. A consistent finding from these previous studies is that developmentally and evolutionarily meaningful changes in sulcal morphology are not homogeneous within association cortices; instead, such changes are focal and related to different aspects of neuroanatomical and functional specialization that are behaviorally meaningful[13,21,30,33,50–53,54,55]. After manually defining 1,320 sulci in individual subjects and implementing a data-driven approach with nested cross-validation in both *Discovery* and *Replication* samples, our results are consistent with and extend these previous findings by showing for the first time that the sulcal depth of three LPFC tertiary sulci predict behavioral performance on a reasoning task in a developmental cohort above and beyond age. In the sections below, we discuss 1) potential underlying mechanisms that likely contribute to the relationship between tertiary sulcal depth and cognitive performance, 2) how the present findings provide a foundation for future studies attempting to link the morphology of brain structures to behavior and functional brain representations, and 3) how our novel, model-based approach can be applied to study other association cortices across the lifespan.

An immediate question generated from our findings is: *What underlying mechanisms could explain why the depths of LPFC tertiary sulci and age reliably predict reasoning performance on a complex behavioral task?* To answer this question, we integrate recent anatomical findings[13,56] with a classic theory[16] and propose a hypothesis linking sulcal depth to short-range anatomical connections, and in turn, to cognitive performance. Specifically, in the 1960s, Sanides[15,16] proposed that morphological changes in tertiary sulci would likely be associated with the development of higher order processing and cognitive skills. The logic of Sanides' hypothesis

extends from the fact that tertiary sulci emerge last in gestation and have a protracted development after birth, while complex cognitive skills such as reasoning ability also have a protracted development in childhood.

Two recent empirical findings provide underlying anatomical mechanisms that could support this relationship between tertiary sulci and cognition. First, there is a relationship between human LPFC tertiary sulcal morphology and myelination[13], which is critical for short- and long-range connectivity, as well as the efficiency of communicating neural signals among regions within cortical networks[57]. Second, anatomical work in non-human primates has shown that long-range white matter fiber tracts have a bias for terminating in gyri, while additional short-range white matter fibers commonly project from the deepest points (*fundi*) of sulci[56], which we refer to as *fundal fibers*. These previous and present findings serve as the foundation for the following novel mechanistic hypothesis linking tertiary sulcal depth to anatomical connections and neural efficiency: deeper tertiary sulci likely reflect shorter fundal fibers, which in turn, reduce the length of short-range anatomical connections between cortical regions, and thus, increase neural efficiency. This hypothesis is similar in logic to the tension-based theory of cortical folding[83] and also feasible as short-range structural connectivity increases and sulci deepen during development[58,59]. This increase in neural efficiency could underlie variability in cognitive performance, which can be tested in future studies incorporating anatomical, functional, and behavioral measures, as well as computational modeling.

In addition to this mechanistic hypothesis, our present findings improve the spatial scale of previous studies attempting to link cortical morphology to behavior associated with LPFC. For example, previous studies identified an association between cognitive skills and cortical thickness of LPFC in its entirety[42–45]. While we find an association between reasoning and cortical thickness

when considering individual tertiary sulci, our analyses indicate that the depths of tertiary sulci and age together are much stronger predictors of reasoning than the cortical thickness of these sulci and age together. In fact, when including the cortical thickness of sulci in the model, performance is worse than age alone (Supplementary Fig. 6a). The combination of these findings across studies suggests that neuroanatomical-behavioral relationships can exist at multiple spatial scales in the same macro-anatomical expanse such as LPFC: cortical thickness at the macroanatomical scale and tertiary sulcal depth at the meso-scale. We also emphasize that, though our model-driven approach identified a subset of LPFC tertiary sulci, morphological features of these sulci are likely correlated with other cognitive tasks, and it is highly probable that other LPFC tertiary sulci play critical roles in other tasks beyond reasoning. Future studies exploring the relationship between sulcal morphology and behavioral performance in additional cognitive tasks at the individual subject level will begin to generate a complete sulcal-behavioral map in LPFC.

In addition to this sulcal-behavioral map in LPFC, two recent lines of work show feasibility for future studies attempting to link tertiary sulcal morphology to brain function, especially for functional activity related to reasoning: one related to tertiary sulci as a meso-scale link between microstructural and functional properties of LPFC and the other identifying functional representations related to reasoning. In terms of the former, a series of recent studies have shown that tertiary sulci are critical functional landmarks in different association cortices[21,30,66]. Additionally, in LPFC, Miller and colleagues[13] showed that the different *pmfs* components explored here were functionally distinct in adults with respect to resting-state connectivity profiles. In terms of the latter, numerous functional neuroimaging studies show that LPFC is central for reasoning performance[67,68]. More explicitly, several studies also indicate that the middle frontal gyrus, the gyrus in which the three sulci (*pmfs-i, pmfs-a,* and *pimfs*) identified by our model are located, plays an important role in cognitive processes that are integral for reasoning, such as maintaining representations and forming associations[3,4]. Thus, future investigations of functional

connectivity, as well as functional representations, relative to tertiary sulci in future studies in children and adults will likely bring us closer to understanding the complex relationship between the development of LPFC anatomical organization, functional organization, and behavior.

While we focus on LPFC in the present study, the novel, data-driven pipeline introduced here can be applied to any cortical expanse. For example, lateral parietal cortex is also critical for relational reasoning, is expanded in humans compared to non-human primates[2,69], and also contains tertiary sulci[14]. Thus, future studies can explore how morphological features of tertiary sulci in a) LPFC and lateral parietal cortex contribute to reasoning performance and b) different association cortices contribute to performance on cognitive tasks, as well as functional representations in each cortical expanse. Our modeling approach can also be applied to data across the lifespan – either cross-sectionally or longitudinally. While it is known that tertiary sulci are shallow indentations in cortex that emerge last in gestation (relative to primary and secondary sulci) and have a protracted development after birth[14,16–20,30,33,49], the developmental timeline of tertiary sulci relative to the development of functional representations and cognitive skills is unknown. Future studies implementing and improving our model-based approach can begin to fill in these gaps in the developmental timeline of tertiary sulci anatomically, behaviorally, and functionally.

Despite the many positive applications of our model-based approach and the many future studies that will likely build on the foundation of the present novel findings, there are also limitations. The main drawback of the precise, single-subject approach implemented here is that it relies on manual sulcal definitions, which are time-consuming and require anatomical expertise. This limits sample sizes and the expanse of cortex that can be feasibly explored in a given study. Additionally, while there is "no one-size-fits-all sample size for neuroimaging studies"[84] and we had a large N (>1000) in terms of sulci explored in the present study, new methods and tools will

need to be developed to increase the number of subjects in futures studies. Ongoing work is already underway to develop deep learning algorithms to accurately define tertiary sulci automatically in individual subjects, and initial results are promising.[70,71]

In summary, using a data-driven, model-based approach, we provide cognitive insights from evolutionarily new brain structures in human LPFC for the first time. After manually defining 1,320 LPFC sulci, our approach revealed that the depths of tertiary sulci reliably predicted reasoning skills above and beyond age. Methodologically, our study opens the door for future studies examining these evolutionarily new tertiary sulci in other association cortices, as well as improves the spatial scale of understanding for future studies interested in linking cortical morphology to behavior. Theoretically, the present results support an untested anatomical theory proposed over 55 years ago[72]. Mechanistically, we outline a novel hypothesis linking tertiary sulcal depth to short-range white matter fibers, neural efficiency, and cognitive performance. Together, the methodological, theoretical, and mechanistic insights regarding whether, or how, tertiary sulci contribute to the development of higher-level cognition in the present study serve as a foundation for future studies examining the relationship between the development of cognitive skills and the morphology of tertiary sulci in association cortices more broadly.

Online Methods

Participants: The present study consisted of *Discovery* (N=33; males=16, females=17) and *Replication* (N=28; males=20, females=8) samples randomly selected from the Neurodevelopment of Reasoning Ability (NORA) dataset[7,10,38]. The terms *male* and *female* are used to denote parent reported gender identity. The *Discovery* Sample included cross-sectional data from 33 typically developing individuals between the ages of 6 and 18. The *Replication* Sample included data from an additional 28 aged-matched subjects. All participants were screened for neurological impairments, psychiatric illness, history of learning disability, and developmental delay. All participants and their parents gave their informed assent and/or consent to participate in the study, which was approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley.

Data Acquisition

Imaging data: Brain imaging data were collected on a Siemens 3T Trio system at the University of California Berkeley Brain Imaging Center. High-resolution T1-weighted MPRAGE anatomical scans (TR=2300ms, TE=2.98ms, 1×1×1mm voxels) were acquired for cortical morphometric analyses.

Behavioral data: Behavioral metrics are only reported for the subjects included in the morphology-behavior analyses (*Discovery*: n = 28, *Replication*: n = 27). Reasoning performance was measured as a total raw score from the WISC-IV Matrix Reasoning task[39] (Fig. 1b; *Discovery:* mean(sd) = 24.28 (4.86); *Replication:* mean(sd) = 27.64 (4.52)). Matrix Reasoning is an untimed subtest of the WISC-IV in which participants are shown colored matrices with one missing quadrant. The participant is asked to "complete" the matrix by selecting the appropriate quadrant from an array of options (Fig. 1b). Matrix Reasoning score was selected as it is a widely used measure of non-verbal reasoning[10,38] and it was the most consistently available reasoning measure for the subjects in this study. Matrix Reasoning has previously been examined in relation to white matter and

functional connectivity in a large dataset that included these participants[10] and a previous factor analysis in this dataset showed that the Matrix Reasoning score loaded strongly onto a reasoning factor that included three other standard reasoning assessments[38].

Processing speed was computed from raw scores on the Cross-Out task from the Woodcock-Johnson Psychoeducational Battery-Revised[73] (WJ-R; Fig. 1b). In this task, the participant is presented with a geometric figure on the left followed by 19 similar figures. The subject places a line through each figure that is identical to the figure on the left of the row (Fig. 1b). Performance is indexed by the number of rows (out of 30 total rows) completed in 3 minutes. Cross-Out scores are frequently used to estimate processing speed in developmental populations.[74,75] Cross-Out score was selected as a control measure in our *Replication* sample in order to assess the generalizability of the findings to another aspect of cognition that does not tax relational reasoning (*Replication:* Mean(sd) = 22.19 (6.26)).

Morphological Analyses

Cortical surface reconstruction: All T1-weighted images were visually inspected for scanner artifacts. FreeSurfer's automated segmentation tools[34,36] (FreeSurfer 6.0.0) were used to generate cortical surface reconstructions. Each anatomical T1-weighted image was segmented to separate gray from white matter, and the resulting boundary was used to reconstruct the cortical surface for each subject[34,76]. Each reconstruction was visually inspected for segmentation errors, and these were manually corrected.

Manual labeling of LPFC sulci: Sulci were manually defined separately in the *Discovery* and *Replication* samples according to the most recent atlas proposed by Petrides (2019)[14]. This atlas offers a comprehensive schematization of sulcal patterns in the cortex. The LPFC definitions have recently been validated in adults[37], but to our knowledge, these sulci have never been defined in a

developmental sample. 12 LPFC sulci were manually defined within each individual hemisphere in tksurfer[13] (Fig. 2; Supplementary Fig. 1 for all manually defined sulci in 122 hemispheres). The curvature metric in Freesurfer determined the boundaries between sulcal and gyral regions. Manual lines were drawn on the inflated cortical surface to define sulci based on the proposal by Petrides[14], as well as guided by the *pial* and *smoothwm* surfaces of each individual[13]. In some cases, the precise start or end point of a sulcus can be difficult to determine on one surface[71]. Thus, using the *inflated*, *pial*, and *smoothwm* surfaces of each individual to inform our labeling allowed us to form a consensus across surfaces and clearly determine each sulcal boundary. Our cortical expanse of interest was bounded by the following sulci: (1) the *anterior* and *posterior* components of the *superior frontal sulcus* (sfs) served as the superior boundary, (2) the *inferior frontal sulcus* (ifs) served as the inferior boundary, (3) the *central sulcus* served as the posterior boundary, and (4) the *vertical and horizontal* components of the *intermediate fronto-marginal sulcus* (imfs) served as the anterior boundary. We also considered the following tertiary sulci: *anterior* (pmfs-a), *intermediate* (pmfs-i), and *posterior* (pmfs-p) components of the *posterior middle frontal sulcus* (pmfs), and the *para-intermediate frontal sulcus* (pimfs)[13,14]. Please refer to Fig. 2a for the location of each of these sulci on example hemispheres and Supplementary Fig. 1 for the location of all 1,320 sulci in all 122 hemispheres.

For each hemisphere, the location of each sulcus was confirmed by two trained independent raters (W.V. and J.Y.) and finalized by the senior author (K.S.W). The surface vertices for each sulcus were then manually selected using tools in *tksurfer* and saved as surface labels for vertex-level analyses of morphological statistics. All anatomical labels for a given hemisphere were fully defined before any morphological or behavioral analyses were performed.

While we could not identify the dorsal and ventral components of the *pimfs* in every hemisphere (Results; Supplementary Table 1), we could identify at least one component of the *pimfs* in each hemisphere in nearly all participants in the *Discovery* (28/33) and *Replication* (27/28) samples. Thus, our inclusion criteria for all subsequent analyses was to include subjects who had at least one *pimfs* component in each hemisphere, which assures that all repeated measures statistics are balanced for effects of sulcus and hemisphere. For those subjects who had identifiable dorsal and ventral *pimfs* components, we merged the components into one label, using the FreeSurfer function *mris_mergelabels* and all findings are reported for the merged label[34].

Characterization of Sulcal morphology: As the most salient morphological feature of tertiary sulci is their shallowness compared to primary and secondary sulci[11-19,21], we focused morphological analyses on measures of sulcal depth. Raw depth metrics were computed in native space from the .sulc file generated in FreeSurfer 6.0.0[34]. We normalized sulcal depth to the maximum depth value within each individual hemisphere in order to account for differences in brain size across individuals and hemispheres. All depth analyses were conducted for normalized mean sulcal depth. As cortical thickness is a commonly used metric in developmental studies, we also considered the mean cortical thickness (mm) for each sulcus. Mean cortical thickness for each sulcal label was extracted using the *mris_anatomical_stats* function that is included in FreeSurfer[35].

Comparison between tertiary and primary/secondary sulci: We compared sulcal depth of tertiary and primary/secondary sulci with a 2-way (hemisphere x sulcal type) repeated measures analysis of variance (rm-ANOVA; Fig. 2c). We conducted the same repeated measures analyses with cortical thickness between tertiary and primary/secondary sulci in both samples (Supplementary Fig. 5; see Supplementary results). All ANOVAs were computed in R with the *aov* function, imported in python via *rpy2*. Effect sizes are reported with the *generalized* eta-squared (η^2) metric.

Assessing the relationship between sulcal depth and reasoning performance

Model selection - Discovery sample: We applied a least absolute shrinkage and selection operator (LASSO) regression model to determine which sulci, if any, were associated with Matrix Reasoning[39]. The depth of all 12 LPFC sulci were included as predictors in the regression model. This analysis was performed separately for each hemisphere. LASSO performs L1-regularization by applying a penalty, or shrinking parameter(α), to the absolute magnitude of the coefficients such that:

$$\min_{\beta \in \mathbb{R}^p} \left(||y - x\beta||_2^2 + \alpha ||\beta||_1 \right)$$

In a LASSO regression, low coefficients are set to zero and eliminated from the model. In this way, LASSO can facilitate variable selection, leading to simplified models with increased interpretability and prediction accuracy[41]. In our case, the LASSO regression algorithm shrinks the coefficients of each of the sulci until only the sulci most predictive of reasoning remain in the model. The LASSO regression model was conducted separately for left and right hemispheres. By convention, we used cross-validation to select the shrinking parameter (α). We used the SciKit-learn GridSearchCV package[77], to perform an exhaustive search across a range of α-values (0.01-10.0), and selected the value that minimized cross-validated Mean Squared Error (MSE_{CV}).

Model evaluation - Replication sample: To further characterize the relationship between sulcal depth and reasoning performance, we used the predictors identified by the LASSO-regression in the *Discovery* sample to predict Matrix Reasoning score in the *Replication* Sample. As age is correlated with Matrix Reasoning score, we included age as an additional covariate in the model [1a]. We fit this model as well as alternate nested models with leave-one-out cross validation

(looCV). We used nested model comparison to assess the unique variance explained by sulcal depth, while accounting for age-related effects on reasoning.

$$y_i = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{pmfsi} + \beta_3 \text{pmfs_a} + \beta_4 \text{pimfs} + \epsilon_{I.} \quad [1a]$$

Additionally, we conducted this analysis with only the two most predictive sulci (*pmfs-i, pimfs*) from the *Discovery* sample.

$$y_i = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{pmfsi} + \beta_3 \text{pimfs} + \epsilon_{I.} \quad [1b]$$

To assess the unique variance explained by tertiary sulcal depth, we compared the MSE_{CV} of this model to the MSE_{CV} of a model with age as the sole predictor [2].

$$y_i = \beta_0 + \beta_1 \text{Age} + \epsilon_{i.} \quad [2]$$

As these models are nested (all predictors in the smaller model [2] are also included in the larger models [1a-b]), we are able to directly compare the prediction error in these two models. Finally, to assess the specificity of the relationship to tertiary sulci in our *Replication* sample, we assessed the fit of model [1] to a full model that included all identified LPFC sulci within a hemisphere [3]. The full model is as follows:

$$y_i = \beta_0 + \beta_1 \text{Age} + \beta_2 x_{2\ldots} + \beta_{12} x_{12} + \epsilon_i \quad [3]$$

where $x_2 \ldots x_{12}$ represent the sulcal depth of each identified sulcus within a hemisphere.

Empirical MSE confidence intervals: The size (n = 27) of the *Replication* sample makes looCV suitable. However, models that are fit with looCV can have high variance. Thus, to assess the potential variance in our estimations, we performed a bootstrapping procedure to empirically estimate the distribution of possible MSE_{cv} predictions for models 1b, 2, and 3. For each model, data were randomly selected with replacement 10,000 times and MSE_{cv} was computed for each

iteration. From this process, we estimate Median MSE and 95% confidence intervals for each model (shown in Fig. 3d). All analyses were conducted with SciKit-Learn package in Python[77].

Assessing morphological and behavioral specificity of the model

Cortical thickness: To assess whether our findings generalized to other anatomical features, we considered cortical thickness, which is an anatomical feature commonly explored in developmental cognitive neuroscience studies[45,78–80]. To do so, we replaced sulcal depth with mean cortical thickness as the predictive metric in our best performing model in the *Replication* sample [Model 1b]. As with depth, the model was fit to the data with looCV. To compare the thickness model to the depth model, we used the Akaike Information Critertion (AIC) which provides an estimate of in-sample prediction error and is suitable for non-nested model comparison. AIC is given by:

$$AIC_i = -2logL_i + 2K_i$$

Where L_i is the likelihood for the model (i) and K_i is the number of parameters. By comparing AIC scores, we are able to assess the relative performance of the two models. If the ΔAIC is greater than 2, it suggests an interpretable difference between models. If the ΔAIC is greater than 10, it suggests a strong difference between models, with the lower AIC value indicating the preferred model[46,47].

Processing Speed: To ascertain whether the relationship between sulcal depth and cognition is specific to reasoning performance, or transferable to other general measures of cognitive processing[75], we investigated the generalizability of the sulcal-behavior relationship to another widely used measure of cognitive functioning: Processing speed. Specifically, we used looCV to predict processing speed (as indexed by Cross-Out score)[73] instead of Matrix Reasoning score. We used AIC to compare Processing speed predictions to Matrix Reasoning predictions.

Data Availability

The data analysis pipeline and code will be released upon publication.

Author Contributions

Initial data collection: S.A.B. Manual anatomical labeling: W.I.V, J.Y, K.S.W. Data analysis and interpretation: W.I.V, J.A.M, K.S.W., S.A.B. Drafting the manuscript: W.I.V, J.A.M, K.S.W., S.A.B. Study conceptualization and supervision: W.I.V, K.S.W., S.A.B

References

1. Zilles, K., Palomero-Gallagher, N. & Amunts, K. Development of cortical folding during evolution and ontogeny. *Trends Neurosci* **36**, 275–284 (2013).
2. Donahue, C. J., Glasser, M. F., Preuss, T. M., Rilling, J. K. & Van Essen, D. C. Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5183–E5192 (2018).
3. Fuster, J. M. The prefrontal cortex - An update: Time is of the essence. *Neuron* vol. 30 319–333 (2001).
4. Badre, D. & Nee, D. E. Frontal Cortex and the Hierarchical Control of Behavior. *Trends in Cognitive Sciences* vol. 22 170–188 (2018).
5. Szczepanski, S. M. & Knight, R. T. *Insights into Human Behavior from Lesions to the Prefrontal Cortex*. *Neuron* vol. 83 (2014).
6. Vendetti, M. S. & Bunge, S. A. Evolutionary and Developmental Changes in the Lateral Frontoparietal Network: A Little Goes a Long Way for Higher-Level Cognition. *Neuron* **84**, 906–917 (2014).
7. Wendelken, C. *et al.* Frontoparietal structural connectivity in childhood predicts development of functional connectivity and reasoning ability: A large-scale longitudinal investigation. *J. Neurosci.* **37**, 8549–8558 (2017).
8. Penn, D. C., Holyoak, K. J. & Povinelli, D. J. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* **31**, 109–178 (2008).
9. Wendelken, C., O'Hare, E. D., Whitaker, K. J., Ferrer, E. & Bunge, S. A. Increased functional selectivity over development in rostrolateral prefrontal cortex. *J. Neurosci.* **31**, 17260–17268 (2011).
10. Wendelken, C., Ferrer, E., Whitaker, K. J. & Bunge, S. A. Fronto-Parietal Network ReconFig. uration Supports the Development of Reasoning Ability. *Cereb. Cortex* **26**, 2178–2190 (2016).
11. Weiner, K. S. The Mid-Fusiform Sulcus (sulcus sagittalis gyri fusiformis). (2018) doi:10.1002/ar.24041.
12. Weiner, K. S., Natu, V. S. & Grill-Spector, K. On object selectivity and the anatomy of the human fusiform gyrus. (2018) doi:10.1016/j.neuroimage.2018.02.040.
13. Miller, J. A., Voorhies, W. I., Lurie, D. J., D'Esposito, M. & Weiner, K. S. A new sulcal landmark identifying anatomical and functional gradients in human lateral prefrontal cortex. *bioRxiv* 2020.03.24.006577 (2020) doi:10.1101/2020.03.24.006577.
14. Petrides, M. *Atlas of the morphology of the human cerebral cortex on the average MNI brain*. 1 Edition. London, UK: Elsevier (2019).
15. SANIDES, F. [Architectonics of the human frontal lobe of the brain. With a demonstration of the principles of its formation as a reflection of phylogenetic differentiation of the cerebral cortex.]. *Monogr. Gesamtgeb. Neurol. Psychiatr.* **98**, 1–201 (1962).
16. Sanides, F. Structure and function of the human frontal lobe. *Neuropsychologia* **2**, 209–219 (1964).
17. Chi, J. G., Dooling, E. C. & Gilles, F. H. Gyral development of the human brain. *Ann. Neurol.* (1977) doi:10.1002/ana.410010109.
18. Welker. Why does cerebral cortex fissure and fold? A review determinants of gyri and sulci. *Cereb. cortex* **8b**, 3--136 (1990).
19. Armstrong, E., Schleicher, A., Omran, H., Curtis, M. & Zilles, K. The ontogeny of human gyrification. *Cereb. Cortex* **5**, 56–63 (1995).
20. Essen, D. C. Van. Cerebral Cortical Folding Patterns in Primates: Why They Vary and What They Signify. *Evol. Nerv. Syst.* 267–276 (2007).
21. Lopez-Persem, A., Verhagen, L., Amiez, C., Petrides, M. & Sallet, J. The human ventromedial prefrontal cortex: Sulcal morphology and its influence on functional organization. *J. Neurosci.* **39**, 3627–3639 (2019).
22. Weiner, K. S. The Mid-Fusiform Sulcus (sulcus sagittalis gyri fusiformis). (2018) doi:10.1002/ar.24041.
23. Dumontheil, I. Development of abstract thinking during childhood and adolescence: The role of rostrolateral prefrontal cortex. *Developmental Cognitive Neuroscience* vol. 10 57–76 (2014).
24. Dumontheil, I., Houlton, R., Christoff, K. & Blakemore, S. J. Development of relational reasoning during adolescence. *Dev. Sci.* **13**, (2010).
25. Blair, C. *How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability*.
26. Cattell, R. B. (Raymond B. & Cattell, R. B. (Raymond B. *Intelligence : its structure, growth, and action*. (North-Holland, 1987).
27. Benson, N. C. *et al.* The retinotopic organization of striate cortex is well predicted by surface topology. *Curr. Biol.* **22**, 2081–2085 (2012).
28. Witthoft, N. *et al.* Where is human V4? Predicting the location of hV4 and VO1 from cortical folding.

Cereb Cortex **24**, 2401–2408 (2014).
29. Yousry, T. A. *et al. Localization of the motor hand area to a knob on the precentral gyrus A new landmark*. vol. 120 (1997).
30. Weiner, K. S. The Mid-Fusiform Sulcus (*sulcus sagittalis gyri fusiformis*). *Anat. Rec.* **302**, 1491–1503 (2019).
31. Amiez, C., Wilson, C. R. E. & Procyk, E. Variations of cingulate sulcal organization and link with cognitive performance. *Sci. Rep.* **8**, 13988 (2018).
32. Garrison, J. R. *et al.* Paracingulate sulcus morphology is associated with hallucinations in the human brain. *Nat. Commun.* **6**, 1–6 (2015).
33. Amiez, C. *et al.* Sulcal organization in the medial frontal cortex provides insights into primate brain evolution. *Nat. Commun.* **10**, 3437 (2019).
34. Dale, A. M., Fischl, B. & Sereno, M. I. *Cortical Surface-Based Analysis I. Segmentation and Surface Reconstruction*. http://www.idealibrary.com (1999).
35. Fischl, B., Dale, A. M. & Raichle, M. E. *Measuring the thickness of the human cerebral cortex from magnetic resonance images*. www.pnas.org.
36. Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
37. Miller, J., Voorhies, W., Lurie, D., D'Esposito, M. & Weiner, K. A new sulcal landmark identifying anatomical and functional gradients in human lateral prefrontal cortex. *bioRxiv* 2020.03.24.006577 (2020) doi:10.1101/2020.03.24.006577.
38. Ferrer, E. *et al.* White matter maturation supports the development of reasoning ability through its influence on processing speed. *Dev. Sci.* **16**, 941–951 (2013).
39. Wechsler, D. *Wechsler intelligence scale for children–Fourth Edition (WISC-IV).* (The Psychological Corporation).
40. Dumas, D., Alexander, P. & Peterson, E. G. Relational Reasoning and Its Manifestations in the Educational Context: A Systematic Review of the Literature Reading and digital listening habits of college students View project Dynamic Measurement Modeling View project. *Artic. Educ. Psychol. Rev.* (2013) doi:10.1007/s10648-013-9224-4.
41. Heinze, G., Wallisch, C. & Dunkler, D. Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal* vol. 60 431–449 (2018).
42. Burgaleta, M., Johnson, W., Waber, D. P., Colom, R. & Karama, S. Cognitive ability changes and dynamics of cortical thickness development in healthy children and adolescents. *Neuroimage* **84**, 810–819 (2014).
43. Dickerson, B. C. *et al.* Detection of cortical thickness correlates of cognitive performance: Reliability across MRI scan sessions, scanners, and field strengths. *Neuroimage* **39**, 10–18 (2008).
44. Østby, Y., Tamnes, C. K., Fjell, A. M. & Walhovd, K. B. Morphometry and connectivity of the fronto-parietal verbal working memory network in development. *Neuropsychologia* **49**, 3854–3862 (2011).
45. Gogtay, N. *et al.* Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8174–8179 (2004).
46. Wagenmakers, E.-J., Farrell, S. & Wagenmakers, -J. *AIC model selection using Akaike weights*. (2004).
47. Burnham, K. P. & Anderson, D. R. Multimodel Inference: Understanding AIC and BIC in Model Selection Multimodel Inference Understanding AIC and BIC in Model Selection. **33**, (2004).
48. Rogers, J. *et al.* On the genetic architecture of cortical folding and brain volume in primates. *Neuroimage* **53**, 1103–1108 (2010).
49. Zilles, K., Palomero-Gallagher, N. & Amunts, K. Development of cortical folding during evolution and ontogeny. *Trends in Neurosciences* vol. 36 275–284 (2013).
50. Amiez, C. & Petrides, M. Functional rostro-caudal gradient in the human posterior lateral frontal cortex. *Brain Struct. Funct.* **223**, 1487–1499 (2018).
51. Brun, L. *et al.* Localized Misfolding Within Broca's Area as a Distinctive Feature of Autistic Disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 160–168 (2016).
52. Leroy, F. *et al.* New human-specific brain landmark: The depth asymmetry of superior temporal sulcus. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1208–1213 (2015).
53. Im, K. *et al.* Spatial distribution of deep sulcal landmarks and hemispherical asymmetry on the cortical surface. *Cereb. Cortex* **20**, 602–611 (2010).
54. Liu, T. *et al.* The relationship between cortical sulcal variability and cognitive performance in the elderly. *Neuroimage* **56**, 865–873 (2011).
55. Borst, G. *et al.* Folding of the anterior cingulate cortex partially explains inhibitory control during

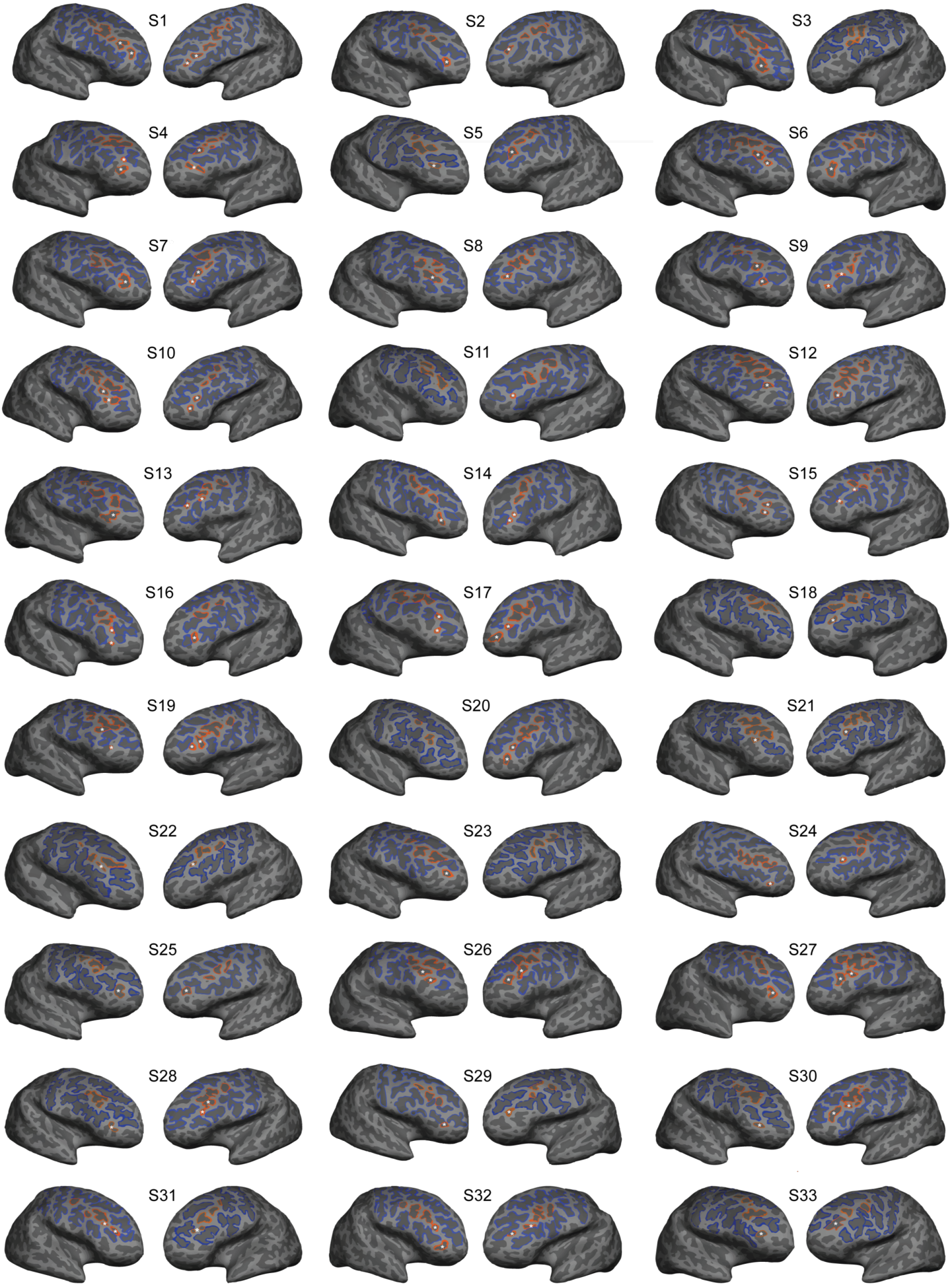
childhood: A longitudinal study. *Dev. Cogn. Neurosci.* **9**, 126–135 (2014).
56. Reveley, C. *et al.* Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion MR tractography. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E2820–E2828 (2015).
57. Turner, R. Myelin and modeling: Bootstrapping cortical microcircuits. *Frontiers in Neural Circuits* vol. 13 34 (2019).
58. Oyefiade, A. A. *et al.* Development of short-range white matter in healthy children and adolescents. *Hum. Brain Mapp.* **39**, 204–217 (2018).
59. Natu, V. S. *et al.* Sulcal Depth in the Medial Ventral Temporal Cortex Predicts the Location of a Place-Selective Region in Macaques, Children, and Adults. *Cereb. Cortex* **00**, 1–14 (2020).
60. Badre, D. & D'Esposito, M. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience* vol. 10 659–669 (2009).
61. Demirtaş, M. *et al.* Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics. *Neuron* **101**, 1181-1194.e13 (2019).
62. Koechlin, E., Ody, C. & Kouneiher, F. The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science (80-.).* **302**, 1181–1185 (2003).
63. Koechlin, E. & Summerfield, C. An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* **11**, 229–235 (2007).
64. Nee, D. E. & D'Esposito, M. The hierarchical organization of the lateral prefrontal cortex. *Elife* **5**, (2016).
65. Reid, A. T. *et al.* Multimodal connectivity mapping of the human left anterior and posterior lateral prefrontal cortex. *Brain Struct. Funct.* **221**, 2589–2605 (2016).
66. Weiner, K. S. *et al.* The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage* **84**, 453–465 (2014).
67. Crone, E. A. *et al.* Neurocognitive development of relational reasoning. *Dev. Sci.* **12**, 55–66 (2009).
68. Watson, C. E. & Chatterjee, A. A bilateral frontoparietal network underlies visuospatial analogical reasoning. *Neuroimage* **59**, 2831–2838 (2012).
69. Van Essen, D. C., Donahue, C. J. & Glasser, M. F. Development and evolution of cerebral and cerebellar cortex. *Brain. Behav. Evol.* **91**, 158–169 (2018).
70. Hao, L. *et al.* Automatic Labeling of Cortical Sulci Using Spherical Convolutional Neural Networks in a Developmental Cohort. in *Proceedings - International Symposium on Biomedical Imaging* vols 2020-April 412–415 (IEEE Computer Society, 2020).
71. Borne, L., Rivière, D., Mancip, M. & Mangin, J. F. Automatic labeling of cortical sulci using patch- or CNN-based segmentation techniques combined with bottom-up geometric constraints. *Med. Image Anal.* **62**, 101651 (2020).
72. Sanides, F. *STRUCTURE AND FUNCTION OF THE HUMAN FRONTAL LOBE**. vol. 2 (1964).
73. Woodcock, R., Mather, N., McGrew, K. & Wendling, B. *Woodcock-Johnson III tests of cognitive abilities*. (Riverside Publishing Company, 2001).
74. McBride-Chang, C. & Kail, R. V. Cross-cultural similarities in the predictors of reading acquisition. *Child Dev.* **73**, 1392–1407 (2002).
75. Kail, R. V. & Ferrer, E. Processing speed in childhood and adolescence: Longitudinal models for examining developmental change. *Child Dev.* **78**, 1760–1770 (2007).
76. Wandell, B. A., Chial, S. & Backus, B. T. Visualization and measurement of the cortical surface. *J. Cogn. Neurosci.* **12**, 739–752 (2000).
77. Pedregosa FABIANPEDREGOSA, F. *et al. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. Journal of Machine Learning Research* vol. 12 http://scikit-learn.sourceforge.net. (2011).
78. Tamnes, C. K. *et al.* Longitudinal working memory development is related to structural maturation of frontal and parietal cortices. *J. Cogn. Neurosci.* **25**, 1611–1623 (2013).
79. Brown, T. T. *et al.* Neuroanatomical assessment of biological maturity. *Curr. Biol.* **22**, 1693–1698 (2012).
80. Vijayakumar, N. *et al.* Brain development during adolescence: A mixed-longitudinal investigation of cortical thickness, surface area, and volume. *Hum. Brain Mapp.* **37**, 2027–2038 (2016).
81. Fischl, B. *et al.* Automatically Parcellating the Human Cerebral Cortex. *Cortex* **14**, 11–22 (2004).
82. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* (2010) doi:10.1016/j.neuroimage.2010.06.010.
83. Van Essen D. A tension based theory of morphogenesis and compact wiring in the central nervous system. *Nature* **385** (1997).
84. Marek, S. *et al.* Towards Reproducible Brain-Wide Association Studies. *bioRxiv* 2020.08.21.257758 (2020) doi:10.1101/2020.08.21.257758.

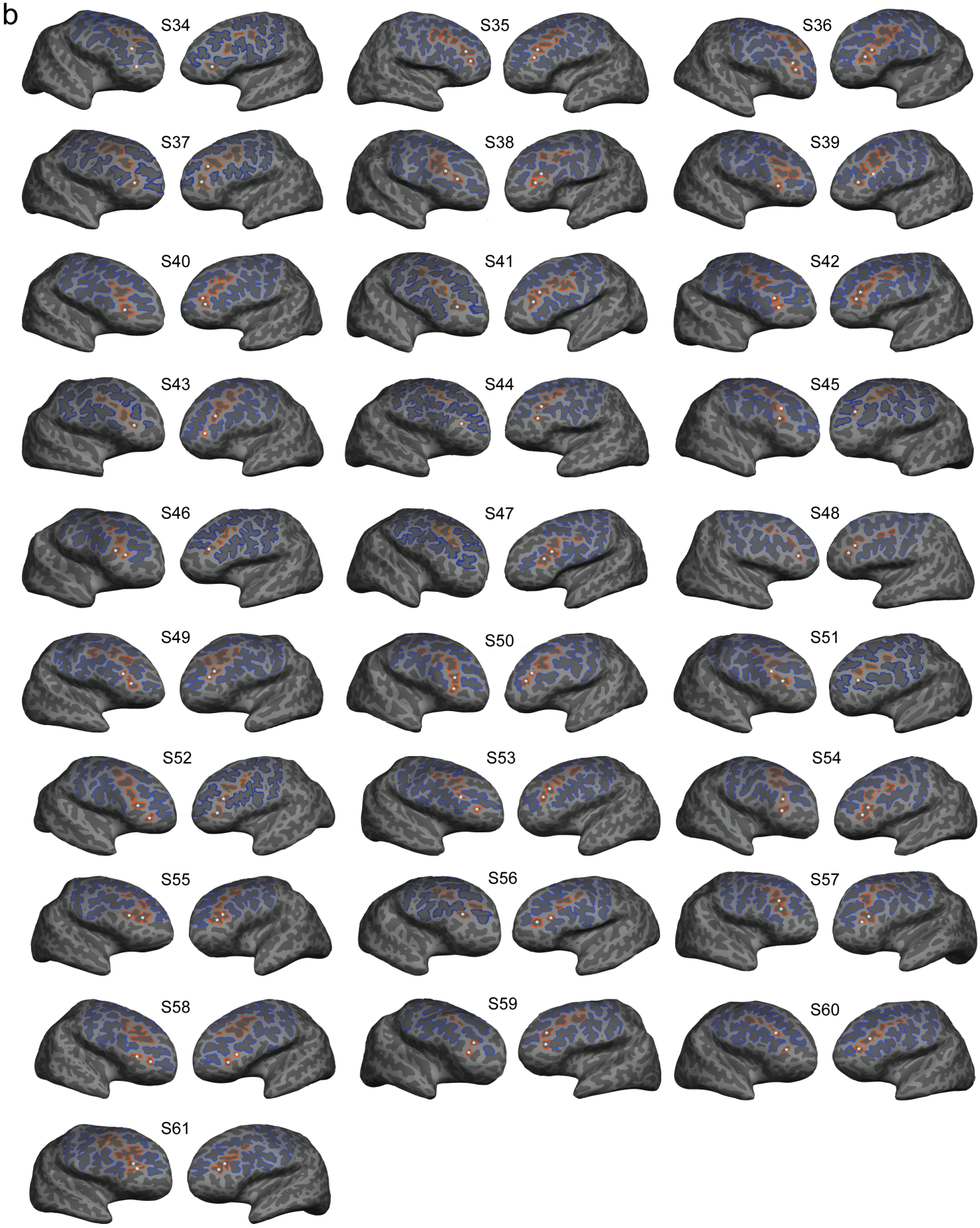
Supplementary Information

No differences in cortical thickness among primary, secondary, and tertiary sulci in LPFC

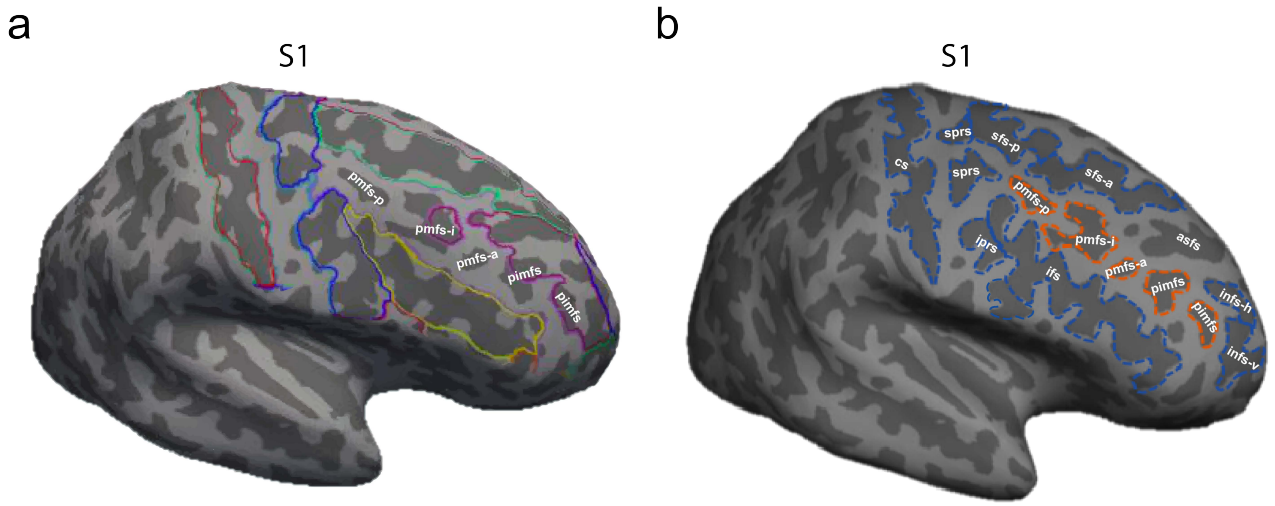
As discussed throughout this paper, sulcal depth is the main morphological feature differentiating tertiary from primary and secondary sulci[11-19,21]. Nevertheless, previous developmental work on structural variability in PFC has frequently focused on cortical thickness[45,78–80]. Thus, we also investigated variability in cortical thickness in both the *Discovery* and *Replication* samples as a function of sulcal type (*tertiary* vs. *primary/secondary*) and hemisphere (*left* vs. *right*). There were no significant differences in cortical thickness between primary/ secondary sulci and tertiary sulci in the *Discovery* ($F(1,27) = 2.44$, $p = 0.13$; $\text{Mean(sd)})_{\text{Tertiary}} = 2.41(0.36)$; $\text{Mean(sd)}_{\text{Non- tertiary}} = 2.37\ (0.26)$; Supplementary Fig. 5a) or *Replication* ($F(1,26) = 2.31$, $p = 0.14$; $\text{Mean(sd)})_{\text{Tertiary}} = 2.31(0.41)$, $\text{Mean(sd)}_{\text{Primary/secondary}} = 2.38(0.30)$; Supplementary Fig. 5b) samples. Interestingly, the rm-ANOVA revealed a main effect of hemisphere in both samples in which right hemisphere sulci were cortically thinner than left hemisphere sulci (*Discovery:* ($F(1,27) = 123.1$, $p<10^{-3}$, $\eta^2_G = 0.09$; Mean(sd)_{RH}; $= 2.30\ (0.28)$; $\text{Mean(sd)}_{LH} = 2.47(0.27)$; *Replication:* ($F(1, 26) = 42.91$, $p<10^{-3}$, $\eta^2_G = 0.06$; $\text{Mean(sd)}_{RH} = 2.20(0.36)$, $\text{Mean(sd)}_{LH} = 2.41(0.29)$; Supplementary Fig. 5b). Thus, while previous developmental work on structural variability in PFC has focused on cortical thickness[45,78–80], when considering tertiary sulci, the present analyses emphasize the utility of sulcal depth, not cortical thickness, for differentiating tertiary from primary/secondary sulci.

a


S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
S25
S26
S27
S28
S29
S30
S31
S32
S33



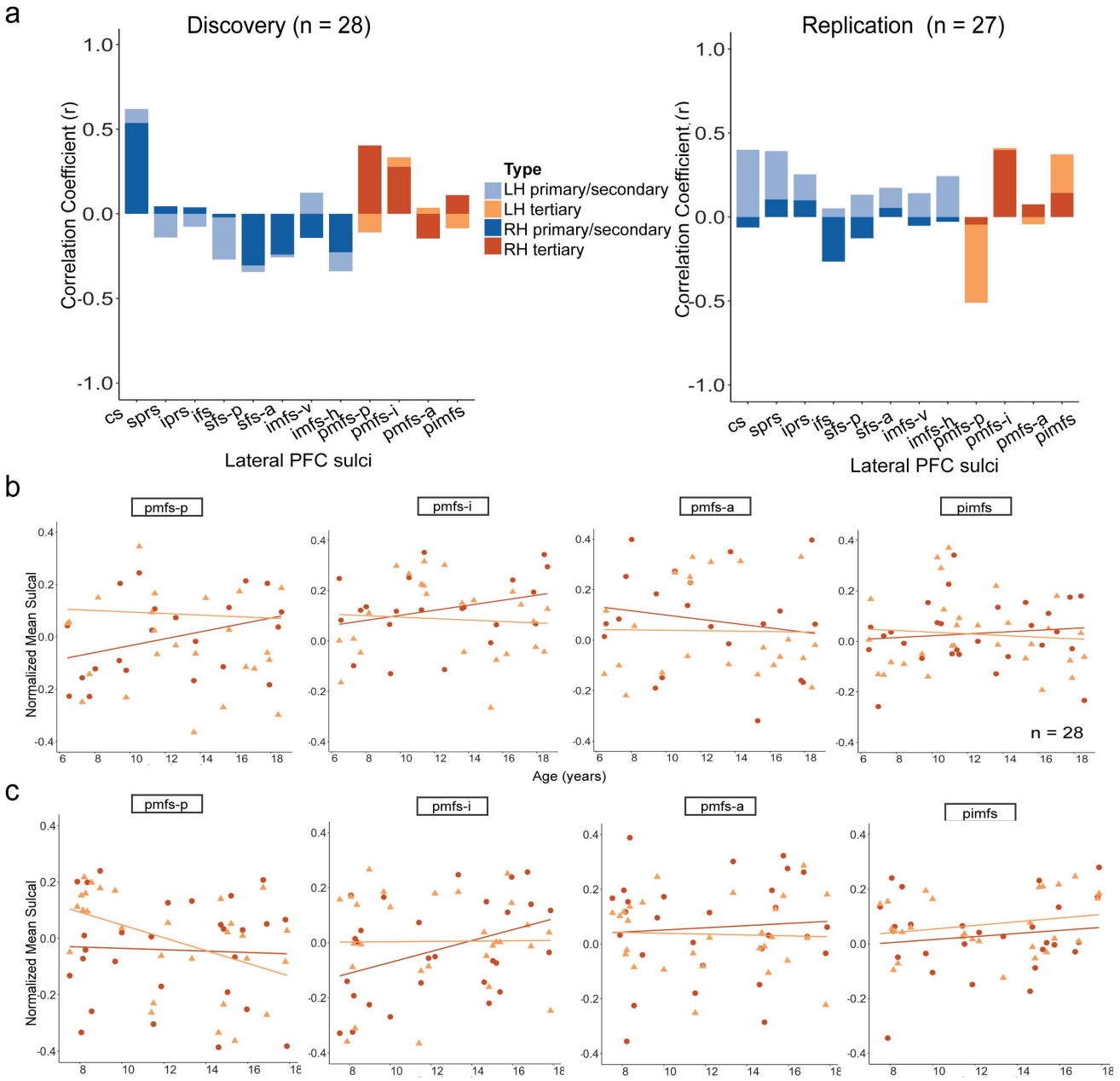
Supplementary Fig. 1. ***Manual labels in the left and right hemispheres of every subject displayed on the inflated cortical surface in FreeSurfer 6.0.0.*** a. Manually labeled sulci on the inflated cortical surface in the left and right hemisphere for every subject in the Discovery Sample. Tertiary (orange) and primary/secondary (blue) sulci are identifiable in every subject. The pimfs (*) can contain two components, one component, or be absent altogether in a given hemisphere (Supplementary Table 1). b. Same layout as in a., but for the Replication Sample. As in the Discovery Sample, tertiary (orange) and primary/secondary (blue) sulci are identifiable in every subject in which the pimfs could contain 0, 1, or 2 components. When present, the pimfs is indicated by *.



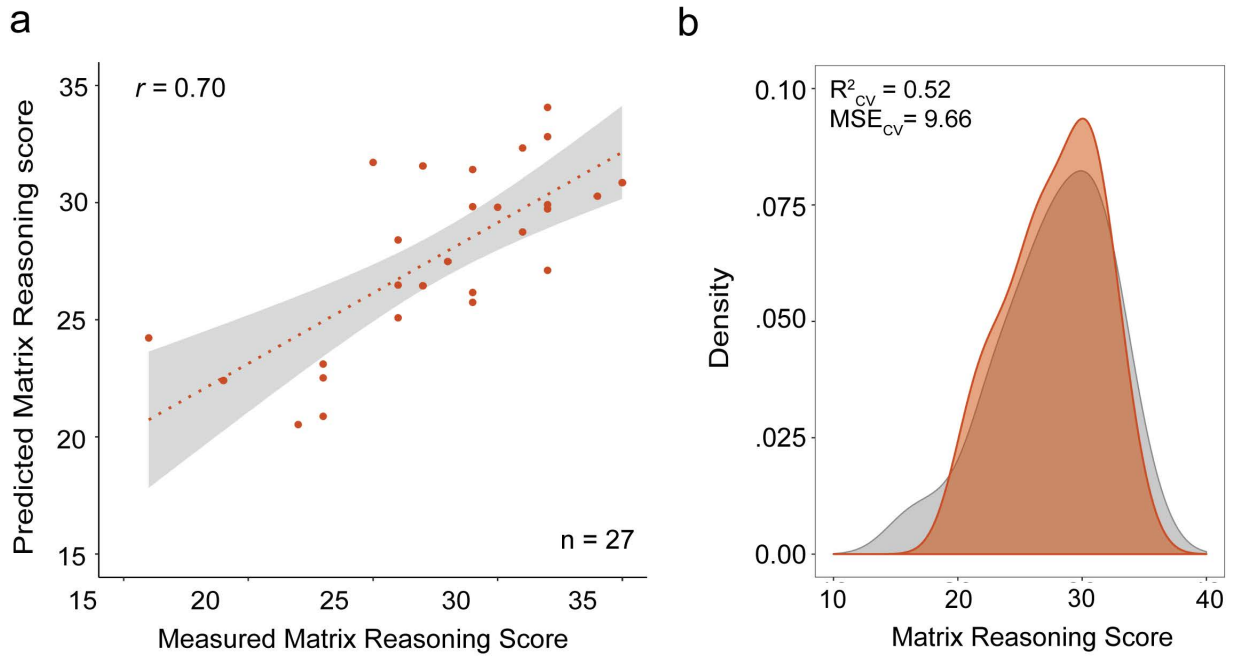
Supplementary Fig. 2. ***LPFC tertiary sulci are often omitted in commonly used atlases.*** **a.** Example inflated cortical surface reconstruction of a right hemisphere. Colors indicate sulcal and gyral definitions provided by automated methods[34,36]. The omitted tertiary sulci explored in the present study are labeled in white. While the automated approach is useful for many studies, we manually defined sulci for our study as present approaches do not yet include tertiary sulci (labeled in white: *pmfs-p, pmfs-i, pmfs-a, pimfs*), and automated methods often include gyral components in the sulcal definitions. Red: central sulcus. Blue: pre-central sulcus. Yellow: inferior frontal sulcus. Turquoise: superior frontal sulcus. Magenta: fronto-marginal sulcus (which includes the horizontal and ventral components of the intermediate frontal sulcus, as well as portions of the *pimfs-i*, and what Petrides[14] refers to as the *accessory superior frontal sulcus* (*asfs*; not examined in the present study). **b.** Example of manual sulcal definitions in the same subject as in **a**. Manual definitions capture both tertiary (*orange*) and primary/secondary (*blue*) sulci.

	Discovery (n = 33)		Replication (n = 28)	
Components	***left***	***right***	***left***	***right***
0	2/33	3/33	1/28	0/28
1	15/33	18/33	3/28	9/28
2	16/33	12/33	24/28	19/28

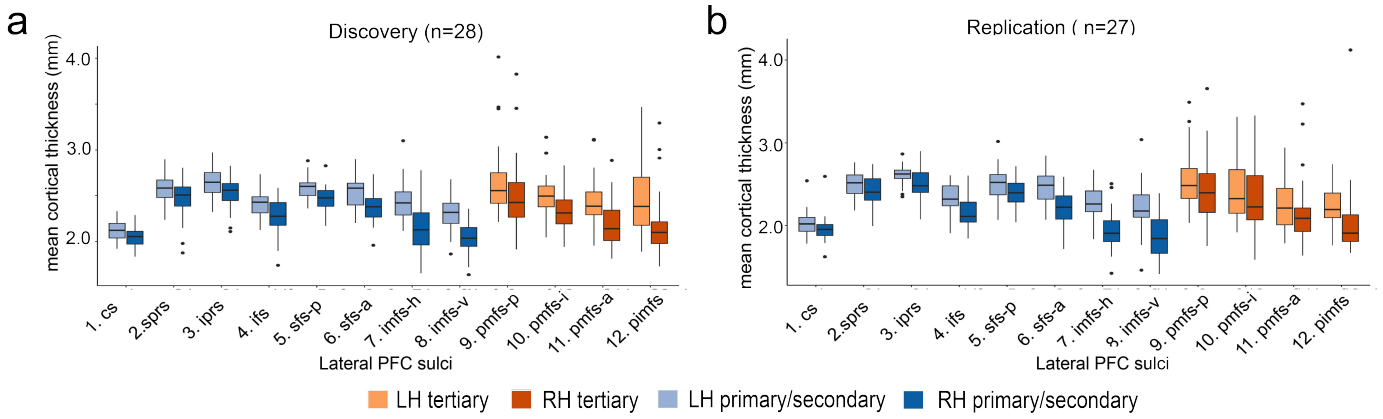
Supplementary Table 1. ***Variability in the number of pimfs components across individuals.*** Subjects had 0,1, or 2 *pimfs* components in each hemisphere. As a majority of subjects in both samples had at least one *pimfs* component, our inclusion criteria was to include subjects who had at least one *pimfs* component in each hemisphere (*Discovery*: 28/33, *Replication*: 27/28), which assures that all repeated measures statistics are balanced for effects of sulcus and hemisphere



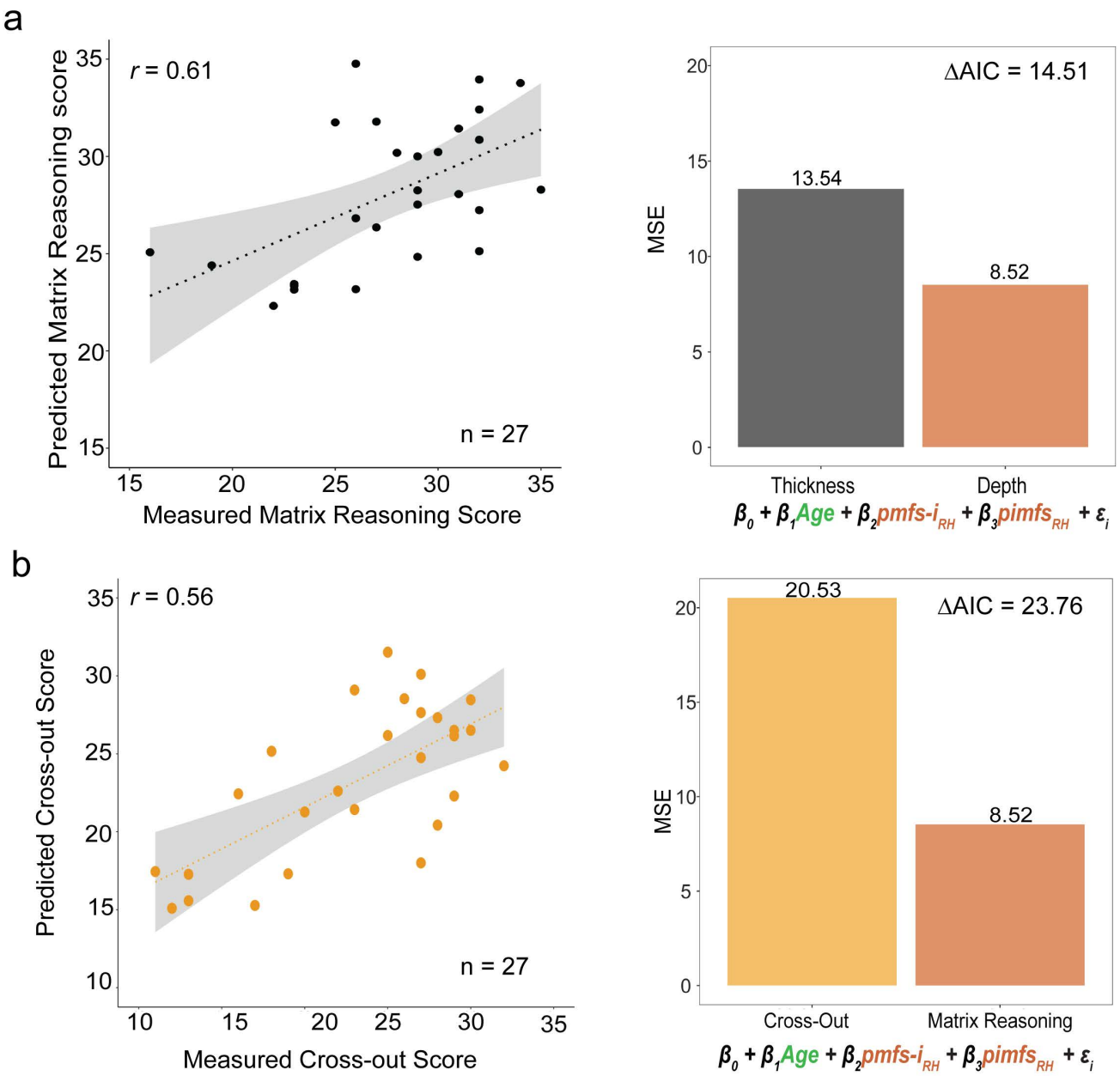
Supplementary Fig. 3. ***Morphological and behavioral associations with age in both Discovery and Replication samples.*** **a.** Correlation between age and sulcal depth in the *Discovery* (*left*) and *Replication* (*right*) samples. Each bar represents the correlation coefficient (Pearson's *r*) between sulcal depth and age for each sulcus (*orange:* tertiary; *blue:* primary/secondary) in the left (lighter shades) and right (darker shades) hemispheres. There is not a clear relationship between sulcal depth and age that is generalizable among LPFC sulci. **b-c.** Scatterplots showing the association between age and sulcal depth for each of the 4 tertiary sulci explored in the present study in each hemisphere (left: lighter shades; right: darker shades) for individual subjects in the *Discovery* (**b**) and *Replication* (**c**) samples. Age does not account well for individual variability in LPFC tertiary sulcal depth.



Supplementary Fig. 4. ***Predicted matrix reasoning score in the Replication sample from three tertiary sulci (pmfs-i, pmfs-a, pimfs).*** **a.** Spearman's correlation between measured and predicted Matrix Reasoning scores in the *Replication* sample for the model including all three tertiary sulci identified in the *Discovery* sample (*pmfs-i*$_{RH}$, *pmfs-a*$_{RH}$, *pimfs*$_{RH}$). **b**. Density plot showing model fit. *orange*: The distribution of predicted scores from this model. *gray:* the distribution of measured Matrix Reasoning scores.



Supplementary Fig. 5. ***No difference in cortical thickness between tertiary and primary/secondary sulci in LPFC.*** **a.** Mean cortical thickness for each of the 12 LPFC sulci in the *Discovery* sample. **b**. Mean cortical thickness for each of the 12 LPFC sulci in the *Replication* sample. Tertiary sulci *(orange)* and primary/secondary sulci *(blue)* do not significantly differ in cortical thickness in either sample.



Supplementary Fig. 6. ***Tertiary sulcal depth more strongly relates to reasoning than cortical thickness and this relationship shows behavioral specificity over other general processing speed measures.*** **a.** Thickness was used in place of depth to predict Matrix Reasoning in the *Replication* sample. The model was fit with looCV. *left:* Spearman's correlation between measured and predicted Matrix Reasoning scores in the *Replication* sample for the best model (pmfs-i_{RH} + $pimfs_{RH}$. + age). *right:* MSE_{cv} for the thickness model compared to the analogous depth model. Difference in AIC was used to compare model performance. Tertiary sulcal depth offered substantially better predictions than cortical thickness. **b.** The same depth model was used to predict Cross-Out score instead of Matrix Reasoning in the *Replication* sample. *left:* Spearman's correlation between measured and predicted Cross-Out scores in the *Replication* sample using the best performing depth model (pmfs-i_{RH} + $pimfs_{RH}$ + age). *right:* MSE_{cv} for the Cross-Out score predictions compared to MSE_{cv} for Matrix Reasoning predictions using the same model. Difference in AIC was used to compare model performance. As the MSE_{cv} for Matrix Reasoning was much lower than the MSE_{cv} for Processing speed, our model shows a degree of behavioral specificity for Matrix Reasoning predictions over general processing speed measures.