Stability and similarity of the pediatric connectome as developmental outcomes.

Tamara Vanderwal,[1,2] Jeffrey Eilbott,[2] Clare Kelly,[3,4,5] Todd S. Woodward,[6,1]
Michael P. Milham,[7,8] F. Xavier Castellanos[8,9]

1. Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada
2. BC Children's Hospital Research Institute, Vancouver, BC, Canada
3. Department of Psychology, Trinity College Dublin, Dublin, Ireland
4. Department of Psychiatry at the School of Medicine, Trinity College Dublin, Dublin, Ireland
5. Trinity College Institute of Neuroscience, Trinity College, Dublin, Ireland
6. Behavioral & Cognitive Neuroscience Institute, Vancouver, BC, Canada
7. Center for the Developing Brain, Child Mind Institute, New York, NY, USA
8. Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, New Orangeburg, NY, USA
9. Hassenfeld Children's Hospital at NYU Langone, New York, NY, USA

Abstract:
Patterns of functional connectivity are unique at the individual level, enabling test-retest matching algorithms to identify a subject from among a group using only their functional connectome. Recent findings show that accuracies of these algorithms in children increase with age. Relatedly, the persistence of functional connectivity (FC) patterns across tasks and rest also increases with age. This study investigated the hypothesis that within-subject stability and between-subject similarity of the whole-brain pediatric connectome are developmentally relevant outcomes. Using data from 210 help-seeking children and adolescents, ages 6-21 years (Healthy Brain Network Biobank), we computed whole-brain FC matrices for each participant during two different movies (MovieDM and MovieTP) and two runs of task-free rest (all from a single scan session) and fed these matrices to a test-retest matching algorithm. We replicated the finding that matching accuracies for children are low (18-44%) relative to previous work in adults and found that cross-state and cross-movie accuracies for children were the lowest. Next, we calculated three measures of connectome stability for each subject: cross-rest (Rest1-Rest2), cross-state (Rest1-MovieDM), and cross-movie (MovieDM-MovieTP), and three measures of within-state between-subject connectome similarity for Rest1, MovieDM, and MovieTP. We show that all six measures of stability and similarity are correlated, but that these measures are not related to age. A principal component analysis of these measures yielded two components that were used to test for brain-behavior correlations with IQ, general psychopathology, and a social skills measure (n=119). The first component was significantly correlated with the social skills measure (r=-0.27, p=0.003) and overall psychiatric scores (r=-0.239, p=0.009). Post hoc correlations showed that cross-rest stability was correlated with both social skills (p=0.0007) and general psychopathology (p=0.001). Connectome similarity during MovieDM was also correlated with social skills (p=0.0026). These findings suggest that the stability and similarity of the whole-brain connectome relate to overall brain development, and in particular, to those regions that support social cognition. We infer that the development of the functional connectome simultaneously achieves patterns of FC that are distinct at the individual subject level, that are shared across individuals, and that are persistent across states and across runs—features which presumably combine to optimize neural processing during development. Future longitudinal work could reveal the developmental trajectories of stability and similarity of the connectome.

Keywords:
movies, naturalistic, development, functional connectivity, social cognition, fMRI

Highlights:
- Identification algorithms yielded low accuracies in this developmental sample.
- Individual differences in FC were not as persistent across states or movies.
- Connectome within-subject stability and between-subject similarity were interrelated.
- Stability during rest and similarity during a movie correlate with social skills scores.

Introduction

Individually distinct patterns of functional connectivity
Whole-brain functional connectivity (FC), also known as the "functional connectome," comprises the complete set of correlations amongst BOLD-signal time-courses in the brain. Aspects of FC are highly reproducible at the group level (Damoiseaux et al., 2006; Yeo et al., 2011), but FC matrices also contain patterns that are distinct at the individual subject level (Finn et al., 2015). This has been most clearly demonstrated by studies applying test-retest matching algorithms, which show that an individual subject's FC matrix is more strongly correlated with another of their own matrices than with an FC matrix from any other subject. Most typically, this is determined by the greatest Pearson's correlation between pairs of FC matrices, making it possible to identify an individual subject from among a group based only on their connectome (Finn et al., 2017, 2015).

Matching algorithms can be run using data collected under a variety of acquisition states, most commonly resting state and conventional task conditions. In 2017, Vanderwal et al. showed that using movies as an acquisition state for FC measures yielded matching accuracies up to 100% in a small sample of adults (Vanderwal et al., 2017). Movies were shown to preserve and possibly enhance the identification of individually distinct patterns of FC in a sample of healthy adults when compared to task-free rest. A remarkable aspect of the adult human connectome is that FC matrices of individuals can be matched successfully even if those matrices comprise data collected during markedly different acquisition states. For example, matching across a social action movie (*Ocean's Eleven)* and a nonverbal movie of abstract shapes and music (*Inscapes)* yielded 94% accuracy, and matching across rest and the action movie achieved 90%. In a larger sample from the Human Connectome Project, matches across conventional tasks (e.g., working memory task, language task) and rest yielded accuracies of 82-92% (Finn et al., 2015). These results highlight the cross-state persistence of individually distinct patterns in FC in adults.

Individualization as a maturational process
Recently, Kaufmann et al. applied this type of identification algorithm to a large pediatric data set (Kaufmann et al., 2017). From a sample of 797 individuals (range 8-22 years, mean 15 ± 3.3), they showed that cross-state identification accuracies across resting state, a working memory task, and an emotion recognition task ranged from 37 to 60%. Using an amalgam measure of identification accuracies across rest and tasks to calculate each participant's "connectome distinctiveness," they further showed a robust positive relationship between age and connectome distinctiveness. The authors posited that individualization of the functional connectome is a maturational process that occurs during adolescence, and that this individualization emerges as the connectome becomes stable across conditions.

Connectome stability has typically been investigated as test-retest reliability, and unsurprisingly, pediatric samples have been shown to have lower reliability for multiple whole-brain FC measures relative to adults (Somandepalli et al., 2015; Thomason et al., 2011). Typically, lower reliabilities in pediatric data are thought to reflect variability in data quality or compliance, but here, we build on the findings from Kauffman et al., to ask whether or not connectome stability might be a developmental outcome in and of itself. We hypothesize that within-state (i.e., across different runs of the same condition, such as rest-rest) and cross-state stability are developmental outcomes, and that the lower reliabilities of FC measures in children might, at least in part, be capturing a developmental process rather than simply noise or poor compliance.

Connectome similarity as a potential developmental measure
We were also interested in asking whether within-state between-subject connectome similarity might be a useful marker of functional brain development. This may seem counter-intuitive in the context of individual differences, but the conceptual model was that if the changes occurring in the developing connectome (e.g., Fair et al., 2008, 2007) are optimizing overall efficiency, a majority of those pathways

and connections might be similar, and between-subject similarity of the connectome would increase. Individually distinct differences in FC might then be expected to emerge as a necessary byproduct of that process. Here, as an exploratory first step and using cross-sectional data, we thus wanted to investigate the relationship of connectome similarity with both connectome stability and behavioral measures of interest.

The present study uses naturalistic conditions to investigate the within-subject stability and the between-subject similarity of the functional connectome in a psychiatric help-seeking pediatric sample (age range 6-21 years). We used data from two runs of task-free rest and two different movie-watching conditions acquired within a single imaging session from the publicly available *Healthy Brain Network Biobank* (Alexander et al., 2017) to calculate FC matrices for each participant. The matrices were then fed to a test-retest matching algorithm. We had three main predictions. First, we expected to replicate the finding that accuracies for this developmental sample would be lower than those obtained for adults across all conditions, especially for cross-state (rest-movie) matches. Second, we hypothesized that secondary whole-brain measures of within-subject connectome stability (cross-rest, cross-movie, and cross-state) might be correlated with each other, such that an individual participant would be more likely to have higher cross-state stability if they had high cross-rest stability. Third, we predicted that within-subject connectome stability and between-subject connectome similarity would be positively associated with age, and possibly related to behavioral measures of interest. In particular, based on the finding that the organization of the pediatric connectome does not persist across states as it does for adults, we thought that cross-state (rest-movie) stability would be likely to change the most from childhood to adulthood. We thus hypothesized that cross-state stability would be the most developmentally relevant measure, i.e., that it would relate to behaviors more strongly than the other brain-based measures.

Methods

Sample and exclusion criteria. Data from the publicly available Healthy Brain Network (HBN) were used for all analyses (Alexander et al., 2017). This database includes fMRI scans and a battery of cognitive and clinical behavioral scores (as well as data from multiple other modalities including electroencephalography, voice samples, saliva samples, and actigraphy) from a community-based sample from the greater New York area. All participants were seeking psychiatric help. The Chesapeake Institutional Review Board approved all study procedures. Written consent was obtained from all participants 18 years of age and older, and parental consent and participant assent were obtained from all other participants. For this study, data from 568 participants were accessed and selected as having at least one T1-HCP anatomical scan, as well as complete data for all four functional runs of interest. Three-hundred and fifty-eight participants were excluded for having a mean framewise displacement of greater than 0.3mm in any functional run, leaving 210 remaining participants (69 females, age range 6-21 years, mean age 12.9 ± 3.6, referred to as HBN-210). See Figure 1 for age distribution, head motion statistics, and distribution of CBCL (the Child Behavior Checklist) scores to reflect the level of psychiatric symptomatology in the sample.

MRI Acquisition. MRI data were collected at Rutgers University using a Siemens 3-Tesla Tim Trio and at the Cornell Brain Imaging Center using a Siemens 3-Tesla Prisma. A 3D T1-weighted structural MP-RAGE structural volume was acquired (TR=2500ms, TE=3.15ms, FA=8°, slice thickness=0.8mm, slices=224). Echo-planar imaging sequences for the functional runs used TR=800ms, TE=30ms, FA=31°, slice thickness=2.4mm, FOV=204mm, a multiband factor of 6, and voxel size=2.4x2.4x2.4mm.

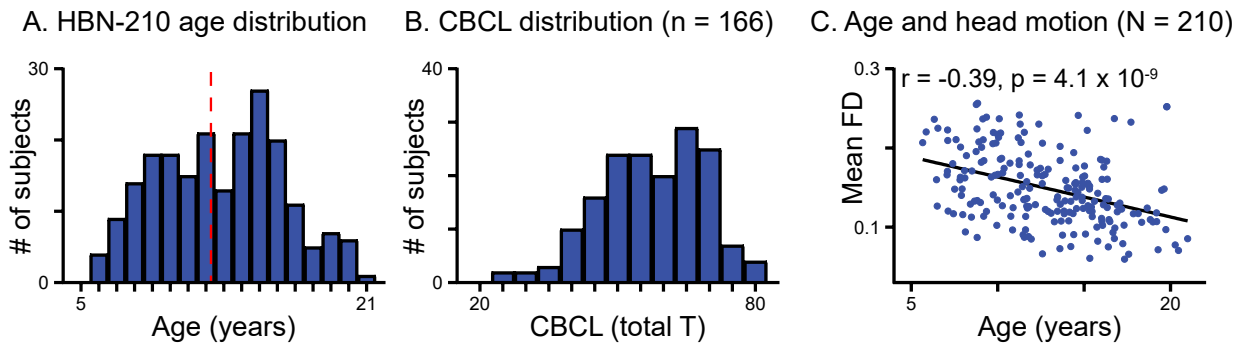


Figure 1. Sample demographics, HBN-210. From the 210 participants (69 females), mean age 12.9 ± 3.6 years. Total T-scores for the Child Behavior Checklist (CBCL), a widely used questionnaire to assess child psychiatric and behavioral problems in which higher scores reflect higher degrees of clinical concern, were available for 166 participants. The right-skew of these data indicates the psychiatric enrichment of the sample, as all participants were seeking psychiatric help (mean total CBCL = 57.2 ± 11.4, cutoff for clinical severity is 65). As expected, data also show a strong relationship between increasing age and decreasing head motion. Mean framewise displacement was averaged across the full duration of the four conditions of interest (two resting state runs, and two movie runs) ($r = -0.39$, $p = 4.1 \times 10^{-9}$) for visualization.

Experimental runs. The functional runs used in this study were two 5-minute runs of eyes-open rest during which a fixation cross was displayed (Rest1 and Rest2), a 10-minute run during which a clip (with sound) from the movie *Despicable Me* was shown (MovieDM), and a 3.5-minute run during which a short film called *The Present* (also with sound) was shown (MovieTP). The MovieDM clip shows three children without parents asking their begrudging temporary guardian to read them a bedtime story (Coffin and Renaud, 2010). MovieTP is a short animation that tells the story of an unhappy boy who receives a puppy for a gift that has only three legs (Frey, 2014). Both stories are highly social, feature strong attachment themes, include a mix of poignant and humorous scenes, and were selected to be emotionally provocative. To match the number of volumes across conditions, all functional runs were truncated at 250 volumes. Where indicated, some analyses were conducted with all volumes for comparison. Technical descriptions of the movie clips are available at the HBN study site (http://dx.doi.org/10.15387/CMI_HBN).

Phenotypic Assessment

Behavioral measures used in these analyses were collected at visits one and two, with MRI scanning occurring at visit three. Three behavioral measures of interest were selected. To capture global functioning, we used the full-scale intelligence quotient (FSIQ) from the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V), and the age-adjusted total T-score from the CBCL (Achenbach, 1999). The CBCL consists of 113 items and is filled out by a caregiver. Higher scores indicate greater clinical concern across the full range of psychiatric symptom domains. Third, we selected the total age-adjusted score from the Social Responsiveness Scale-2 (here, SRS) (Constantino et al., 2003). The SRS is a 65-item questionnaire that measures social skills, with higher scores indicating greater impairment. It is classically used to measure autistic traits but is also used to capture social problems in other disorders. SRS scores have been shown to be normally distributed in the general population. We focused on a social skills measure because of the highly social nature of the movies used in this experiment, and because of the dynamic social development that is known to occur in the studied age range (Blakemore, 2008; Blakemore and Frith, 2004; Braams and Crone, 2017; Mills et al., 2014). Of the 210 participants, all three behavioral measures were available for 119 participants. Demographics for this subset (HBN-119), along with motion statistics and relevant correlations, are shown in Supplementary Figure 1s.

Data preprocessing
Standard preprocessing was performed using fMRIPrep version 1.0.15 (Esteban et al., 2019) including slice time correction, motion realignment, and transformation into Montreal Neurological Institute space. Non-aggressive denoising was performed filtering out the 6 rigid-body motion correction estimates, framewise displacement, 6 anatomical CompCor components (Behzadi et al., 2007), and ICA-AROMA motion artifacts. In addition to excluding all subjects with a mean framewise displacement (FD) of greater than 0.3mm in any functional run, ICA-AROMA was used to mitigate motion artifact, as it has been shown to remove spurious noise more effectively than spike regression (Pruim et al., 2015). Motion censoring was not performed to preserve the continuous time-courses and temporal structure of the movie-watching data. Global signal regression was not performed. Data were not spatially smoothed.

Whole-brain FC matrices
The primary analyses were based on FC connectivity matrices generated using 200 functionally defined regions-of-interest (ROIs) (Craddock et al., 2012). For each subject, the mean time series of each ROI was extracted and Pearson's correlation coefficients were calculated between all ROI pairs creating a 200x200 whole brain connectivity matrix. Correlation coefficients were then Fisher z-transformed for statistical inference. Duplicate ROI pairs were excluded, leaving 19,900 edges in subsequent analyses. Each subject had an FC matrix for each of the four conditions (Rest1, Rest2, MovieDM and MovieTP).

Test-retest matching algorithm
The prediction procedure closely followed methods described previously (Finn et al., 2015). Four databases of 210 matrices were created, one per subject for each condition (Rest1, Rest2, MovieDM, MovieTP). The matching algorithm was run between two databases at a time. A subject's FC matrix was selected from one database, and the Pearson's correlation coefficient was calculated between that matrix and every other matrix in the other database. If the matrix with the highest correlation was from the same subject, the pairing was considered an accurate match. Group accuracy of the condition pairing was determined by calculating the proportion of subjects for whom an accurate match was obtained. This analysis was performed across all twelve possible state pairings (e.g., Rest1-Rest2, Rest2-Rest1, Rest1-MovieDM, Rest1-MovieTP, etc.).

To assess the significance of the accuracies produced by the algorithm, we performed nonparametric permutation testing using randomly assigned false identities. The algorithm was run 10,000 times to determine how many times the falsely assigned pair was identified as being the most strongly correlated (though see Wang et al., 2018 for an argument that this is not necessary). To assess the degree to which head motion might be contributing to successful matching, we performed the identification algorithm using only the distribution of framewise displacement values in each run (also from Finn & Shen, 2015). We computed motion distribution vectors for each participant from the framewise displacement time courses of each condition. Sixty bins were set to capture the grand mean ± 3 standard deviations, and the 1x60 vectors were then used (in place of the FC matrices above) to run the matching algorithm. In this dataset, the grand mean was 0.149mm ± 0.20mm, creating bins between -0.44 and 0.74mm. Limiting this to positive numbers resulted in 38 bins being used. To check for motion effect that might occur due to the small but statistically significant differences in motion across conditions, we created a motion-matched subset of subjects, selecting subjects who had no significant difference in head motion across conditions, and reran the matching algorithm as above using only those subjects' FC matrices. Finally, to evaluate the effect of condition duration on matching accuracies, we created matrices for Rest1, Rest2 and MovieDM for all subjects, this time using the full duration of each condition (instead of truncating at 250 volumes), and re-ran the matching algorithm with these matrices and the previously used MovieTP matrices.

To explore the network location of edges that contributed to successful matches, we computed the differential power (DP) of each edge. DP is an empirical probability that indicates the proportion of times a subject is matched to itself rather than to another subject based on that edge. The measure therefore

captures how much a particular edge contributes to a successful match (see Finn and Shen 2015 for more details). We then extracted the top 5% of edges for the Rest1-Rest2, Rest1-MovieDM, and MovieDM-MovieTP pairings, and calculated the percentage of edges within each pair of a priori networks that met that threshold. We used the Yeo 7-network parcellation (Yeo et al., 2011), and the network membership of each Crad-200 ROI was assigned to whichever network contained the majority of its voxels.

Parcellation resolution
In previous work, we have shown that increasing parcellation resolution can improve identification accuracies, but that it does so somewhat equally across conditions (Vanderwal et al., 2017). To test that finding here, we parcellated the data at all of the 43 resolutions defined in a publicly available atlas (Craddock et al., 2012). The number of clusters ranged from 10-950. We then ran the matching algorithm at each resolution, using the Rest1-Rest2, Rest1-MovieDM, and MovieDM-MovieTP pairings. All other analyses used the Crad-200 parcellation.

Scan duration
To test the effect of scan duration on cross-state and cross-run matching, we ran the identification algorithm while varying the amount of data used, from 2-250 volumes, adding sequential TRs one at a time.

Number of edges
To test if one condition required fewer edges to facilitate successful identification of individual subjects, we ran the algorithm using a sequentially increasing number of edges. Edges were rank-ordered from least contributory (lowest DP) to most contributory (highest DP). We then ran the algorithm using the lowest 0.5% of edges, and successively repeated the procedure adding an additional 0.5% of edges until all edges were included.

Connectome stability
We define stability as the correlation between a subject's whole brain FC matrix collected during two separate functional runs. These basic whole-brain correlations were calculated for each subject for cross-run (Rest1-Rest2), cross-movie (MovieDM-MovieTP), and cross-state stability (Rest1-MovieDM).

Connectome similarity
Between-subject similarity is defined as the average of the correlations between a subject's FC matrix and that of every other subject within a given condition (Rest1, MovieDM, MovieTP). We note that this is an FC-based measure, and therefore can be computed for both movies and rest, as opposed to a time-course based correlation, such as Intersubject Correlations (Hasson et al., 2004)).

Stability-similarity relationships
We then tested for linear relationships between each subject's stability and similarity measures. Due to the possibility that head motion artifact might still be present in these whole-brain measures of stability and similarity for some participants, we performed follow-up multiple linear regression including each subject's framewise displacement from the relevant condition(s) as a covariate.

Brain-behavior correlations
To test the hypothesis that whole-brain stability and/or similarity is developmentally relevant, we first aimed to reduce the dimensionality of the six measures of connectome stability and similarity by conducting a principal component analysis on them. The top components explaining the majority of the variance were selected, and correlations between those components and the three behavioral measures of interest (IQ, CBCL score, and SRS score) were calculated.

Results

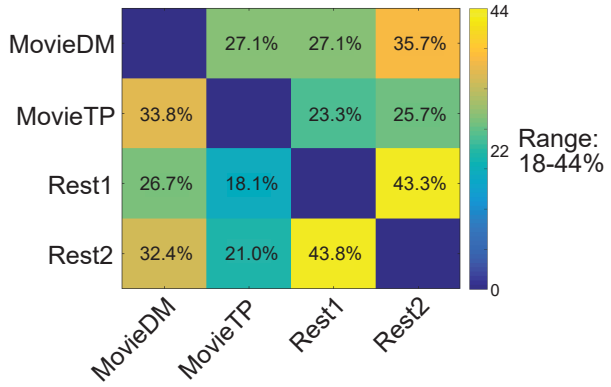
Head motion: As expected, mean framewise displacement was strongly correlated with age (r=-0.39, $p=4.1x10^{-09}$, see Figure 1). By condition, mean FD showed an order effect, and was Rest1=0.141mm; MovieDM=0.147mm; Rest2=0.15mm; MovieTP=0.16mm). A repeated-measures ANOVA of mean FD across conditions was statistically significant (F(3,209)=6.74, p=0.0002). Differences were driven by greater head motion during MovieTP relative to each of the other conditions (p<0.01). These patterns also held true in the behavioral subsample of 119 participants; head motion was not correlated with any of the three behavioral measures of interest (see Supplementary Figure 1s).

Accuracies of identification algorithm

For matches across all conditions, matching accuracies ranged from 18-44% (Figure 2A), whereas permutation testing using falsely assigned subject pairs yielded accuracies of 0.3-0.4% (Supplementary Figure 2s). The only pairing of a truly repeated condition (Rest1 and Rest2) had the highest accuracies (43% and 44%), while the four cross-state (Rest-Movie) pairings had the lowest (18-36%). The two cross-movie accuracies were comparable (27% and 34%). When the algorithm was re-run using the full duration of each scan to compute FC matrices for each condition (i.e., not limited by the 250 volumes of MovieTP), the accuracies did shift, ranging from 10-49%. In most cases, pairings of mismatched durations got worse (e.g., MovieDM-MovieTP went from 27 to 21%), and pairings with longer durations improved (e.g. Rest1-Rest2 accuracies went from 43 to 48%). When using motion distribution parameters alone to match subjects, accuracies ranged from 2-8%, and when using motion-matched groups (n=190), percentages ranged from 20-46% and followed the same patterns observed in the full sample. Results of these supporting analyses are shown in Supplementary Figure 2s.

When assessing the network distribution of the top 5% of edges, we found that the default, frontoparietal and dorsal attention networks were prominent across all condition pairings (Figure 2B). Qualitatively, the default-frontoparietal network connections and dorsal attention-visual network connections appear more contributory in pairings involving movies, while default-ventral attention network connections were more prominent in the Rest1-Rest2 pairing.

A. Identification accuracies of test-retest matching algorithm



B. Network distribution of most contributory edges

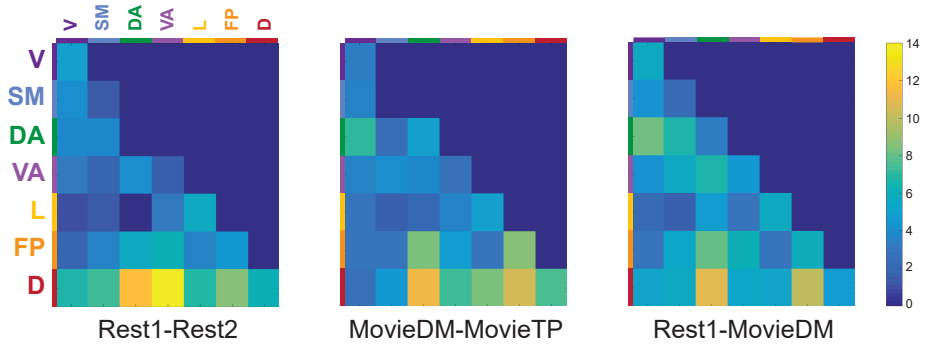


Figure 2. Test-retest matching accuracies, HBN-210. A) Matching accuracies of the identification algorithm ranged from 18-44%, with the highest accuracies attained by the only true condition pairing, Rest1-Rest2 (and vice versa). Movie-TP pairings yielded the lowest accuracies (18-23%), and cross-movie and Rest-MovieDM pairings were in the middle (26-36%). **B)** When examining the distribution of the most contributory edges, default, frontoparietal, and ventral attention networks were prominent across pairings, and the cross-state pairing (Rest1-MovieDM) appears to have the most diffuse distribution of contributory edges.

Parcellation resolution

As expected, when testing with different parcellation resolutions, identification accuracies increased as parcellation resolution increased for all pairings (cross-run, cross movie, and cross-state) (see Figure 3A). The highest accuracy attained was for Rest1-Rest2 pairings, which was 70% at around 600 clusters. All other analyses presented here used the Crad-200 parcellation.

Scan duration

Also as expected, longer scan durations improved matching accuracies (Figure 3B). Here, the effect of condition was more pronounced, as Rest1-Rest2 pairings continued to improve incrementally up to the maximum of 250 volumes, whereas both the cross-state and cross-movie pairings reached ceiling accuracy at around 100 volumes.

Number of edges used

Including more edges yielded higher accuracies overall. Again, the trajectory for the Rest1-Rest2 pairing separates out from both the cross-state and cross-movie pairings (Figure 3C).

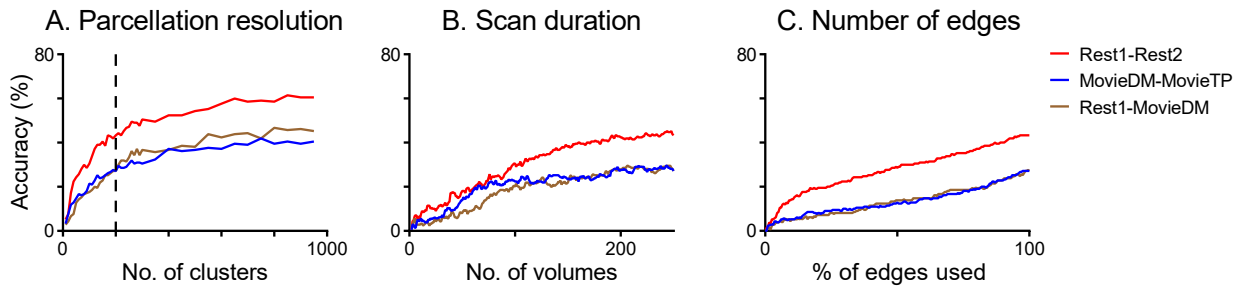


Figure 3. Effects of processing parameters on identification accuracies across different pairings. Altering preprocessing parameters did not change the overall pattern of results. The only properly matched pairing, the Rest1-Rest2 pairing (red line), achieves higher accuracies than either the cross-movie or the cross-state pairings, regardless of the parcellation resolution, the number of volumes, or the number of edges used. The dashed line in A denotes the Crad-200 parcellation, which was used for all other analyses.

Connectome stability and similarity

Figure 4 illustrates how the stability and similarity measures were calculated, and how they relate. Cross-rest and cross-state stability were correlated ($r=0.64$, $p=9.0\times10^{-26}$). Between-subject similarity during Rest1 and MovieDM were also correlated ($r=0.38$, $p=1.5\times10^{-8}$). When each subject's three stability scores and their three similarity scores were averaged, mean stability and mean similarity were also strongly correlated ($r=0.68$, $p=2.5\times10^{-29}$), showing that participants with higher within-subject stability also demonstrated higher between-subject similarity. As expected, all six measures were significantly correlated with head motion, but contrary to predictions, no significant relationships with age were found. Further, neither mean stability nor mean similarity were correlated with age (stability: $r=-0.0001$, $p=0.99$; similarity: $r=0.0106$, $p=0.91$). For completeness, all pair-wise correlations for the six secondary brain measures with each other, and with head motion and age, are shown in Supplementary Figure 3s.

We were concerned that the apparent interrelated nature of these measures could be driven by head motion, such that a participant with low stability due to noise from head motion would also be expected to be less similar to the other participants because of that same noisiness. We entered mean framewise displacement across all runs into a GLM relating stability and similarity. Even when covarying for motion, the stability-similarity relationship remained highly significant ($p=6.5\times10^{-27}$). As a second test, we binned participants by mean FD, and selected bins to get the most subjects with the least spread in FD. Using FD of 0.13-0.15mm, we had 33 participants. Neither stability nor similarity in this subsample was significantly correlated with FD, but the robust correlation between stability and similarity persisted ($r=0.81$, $p=1.17\times10^{-8}$).

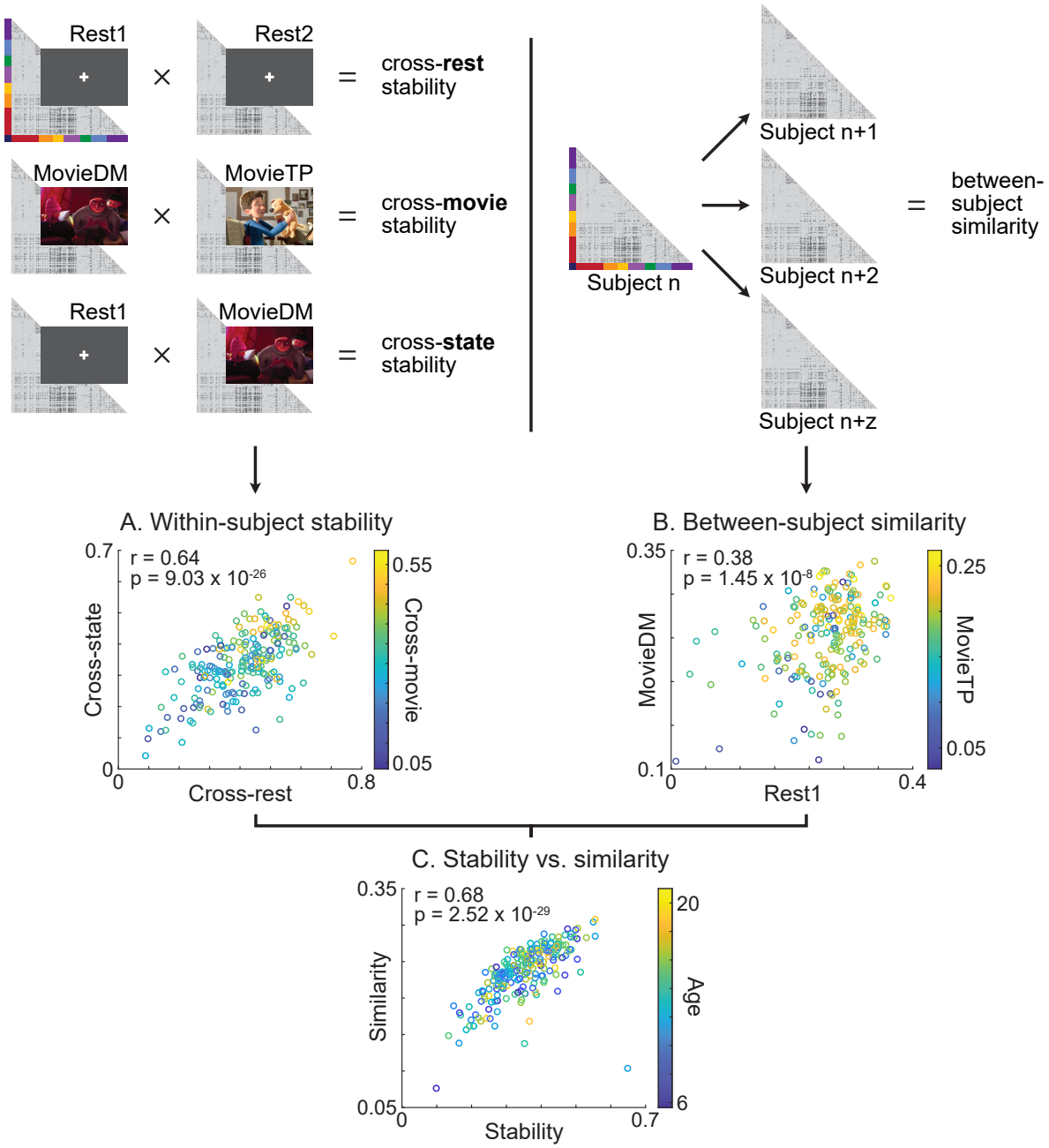
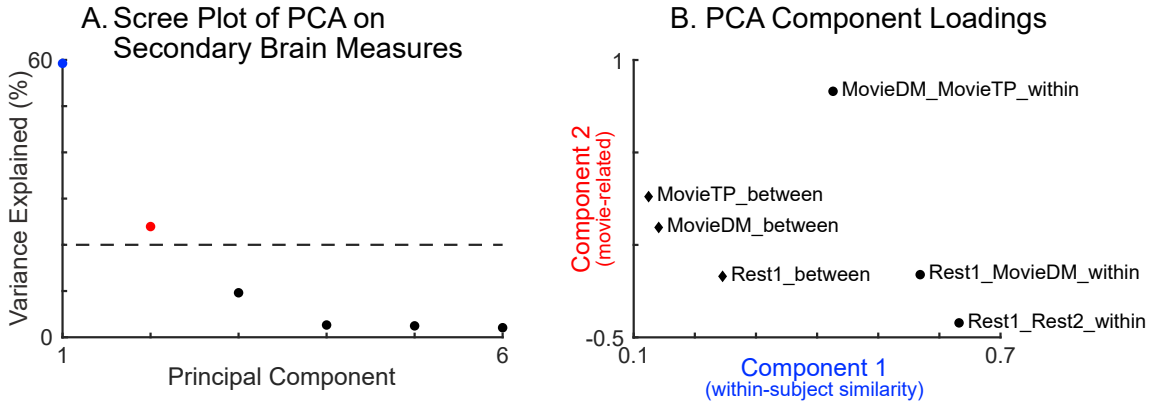


Figure 4. Whole-brain within- and between-subject measures of FC correlations, N=210. Six measures of connectome stability and similarity were calculated for each subject. Plot A shows the three within-subject measures of connectome stability. For between-subject similarity, we calculated the correlation between a subject's matrix with all other subjects' matrices for each condition and took the average of those between-subject correlations. These three similarity measures are plotted in B. For C, each subjects' three measures of stability and similarity were averaged. R-values for each scatterplot report the correlation between the x and y axes, with the third measure mapped onto that relationship using color. Correlations between all six measures, and head motion and age, are shown in supplementary Figure 3s.

Brain Behavior Correlations

A principal component analysis (PCA) was used to reduce the dimensionality of the six measures of connectome stability and similarity. The PCA yielded components that explained 3-60% of the variance. The top two components explained 59% and 24% of the variance and were selected for subsequent analyses (see Figure 5 for scree plot and component loadings). Component 1 was weighted towards the within-subject stability measures, and Component 2 was weighted towards the movie-related measures, particularly cross-movie stability.

When correlated with age-adjusted measures of psychopathology and general behavioral problems (CBCL), full-scale IQ (FSIQ), and social responsiveness score (SRS), Component 1 was significantly correlated with both SRS (r=-0.270, p=0.003) and CBCL (r=-0.239, p=0.009) (Figure 5C). None of the other correlations were significant. After accounting for mean FD within a linear model, the relationships between component 1 and SRS and CBCL remained significant (p=0.003, p=0.012, respectively, not shown). We then conducted post hoc multiple linear regressions to assess for relationships between SRS, CBCL and the six brain measures of stability and similarity. We report both the correlations (r and p-values) of these pair-wise correlations, as well as the p-value of a GLM with head motion from the relevant conditions as covariates (Figure 5D). We found that within-subject stability for rest was significantly correlated with both SRS (p=0.0007) and CBCL (p=0.0011). Between-subject similarity during MovieDM was also significantly correlated with SRS (p=0.0026) and was just under significance for CBCL (p=0.0044) (Bonferroni corrected to p<0.00416 for twelve comparisons). CBCL and SRS were subsequently found to be strongly correlated (r=0.71, p=3.8×10^{-19}), and all CBCL subsets (e.g., Social Problems, Attention Problems, Somatic Complaints, etc.) were significantly correlated with SRS.



C. Brain-behavior correlations

	CBCL	WISC FSIQ	SRS
Component 1: r (p)	**-0.239 (0.009)**	0.142 (0.124)	**-0.27 (0.003)**
Component 2: r (p)	0.163 (0.077)	0.095 (0.306)	0.093 (0.317)

D. Post hoc correlations

		CBCL		SRS	
		r (p)	GLM (p)	r (p)	GLM (p)
Stability (within subjects)	Rest1_Rest2	**-0.30 (0.0011)**	**(0.0017)**	**-0.31 (0.0007)**	**(0.0007)**
	MovieDM_MovieTP	-0.03 (0.7374)	(0.9182)	-0.10 (0.2662)	(0.3348)
	Rest1_MovieDM	-0.18 (0.0494)	(0.0462)	-0.19 (0.0342)	(0.0401)
Similarity (between subjects)	Rest1	-0.23 (0.0113)	(0.0137)	-0.20 (0.0310)	(0.0305)
	MovieDM	-0.26 (0.0044)	**(0.0040)**	**-0.27 (0.0026)**	**(0.0021)**
	MovieTP	0.06 (0.5300)	(0.4236)	-0.02 (0.8053)	(0.9136)

Figure 5. Brain-Behavior Correlations, n=119. 5A) A Principal Component Analysis (PCA) performed on the six secondary brain measures (from Figure 4) yielded the scree plot shown here. The top two components were selected for further analyses**. 5B)** Component 1 explained 59% of the variance and was weighted towards the within-subject stability measures. Component 2 explained 24% of the variance, and was weighted towards movie-related measures, especially cross-movie stability. We then tested the two components for correlations with the three developmental measures of interest, CBCL (general psychopathology and behavioral problems), full-scale IQ, and a social responsiveness score (SRS). **5C)** A significant relationship was identified between component 1 (FC stability) and social skills and general behavioral problems. **5D)** Post-hoc statistical tests between both SRS and CBCL and the 6 secondary brain measures showed significant relationships between SRS and cross-rest stability, and between SRS and MovieDM similarity (Bonferroni corrected). To a somewhat lesser degree, CBCL followed the same pattern, and when included in a GLM with head motion as a covariate, all relationships remained significant. Directionality of the brain-behavior correlations indicate that higher stability or similarity predict lower SRS scores (better social skills) and lower CBCL scores (lower clinical concern).

Discussion

This study investigated within-subject stability and between-subject similarity of the whole-brain connectome during movies and rest in a psychiatrically enriched sample of 210 children and adolescents, ages 6-21 years. As expected, we replicated the finding that accuracies for this developmental sample were low (Kaufmann et al., 2017). We also extended the finding of low cross-state stability in children to include naturalistic conditions, showing that both cross-movie and cross-state (movie-rest) matching accuracies were particularly low in this sample. Also as hypothesized, we showed that within-subject connectome stability (cross-rest, cross-movie, and cross-state) were correlated with each other, meaning that an individual participant was more likely to have higher cross-state stability if they had high cross-rest stability. The same was also true for between-subject connectome similarity. Further, connectome stability and similarity were shown to be largely interrelated measures, even when accounting for head motion. Contrary to our hypotheses, we found no relationship between either connectome stability or connectome similarity with age. In a subset of 119 participants, we identified significant brain-behavior correlations using the stability and similarity measures: cross-rest stability was significantly correlated with both social skills and overall psychiatric problems (CBCL scores). Further, connectome similarity during the movie *Despicable Me* was significantly correlated with social skills. Overall, we conclude that stability and similarity of the functional connectome may be meaningful developmental markers, but multiple methodological questions and caveats remain that can be especially influential when conducting these types of analyses in children and adolescents.

Fingerprinting in developmental samples

We showed that test-retest matching algorithms of whole-brain functional connectivity matrices in a pediatric sample are lower than were previously reported in adults, and more pointedly, that accuracies for cross-state and cross-movie matches are remarkably low. For example, matching accuracies across two different cartoons from the same scan session in this study were 27% and 34%. In a previous study in adults (with a smaller sample), we found cross-movie matching accuracies to be between 80-94%, even across markedly different movies and with one-week interscan intervals (Vanderwal et al., 2017). The finding here suggests that within-subject stability and between-subject similarity of the whole brain connectome are substantially different in children than in adults. In other words, patterns of functional connectivity (FC) that are distinct at the individual subject level appear to persist (or are identifiable) across states for adults, but not for children and adolescents.

A potential caveat to this finding relates to scan duration, or the number of volumes used to construct each matrix. Most of the analyses we reported were truncated at 250 volumes to match the shortest movie

clip (MovieTP). When we re-ran the matching algorithm using the full duration of available volumes for Rest1, Rest2 (375 volumes), and MovieDM (750 volumes), accuracies shifted. For example, Rest1-Rest2 accuracies went from 43% to 48%, whereas Rest1-MovieDM accuracies went from 27% to 21%. In most cases, pairings involving markedly mismatched numbers of volumes got worse, e.g. MovieDM_MovieTP went from 27% to 20%. These shifts indicate that the identification of individually distinct patterns of FC depends in part on the amount of data used to construct the FC matrices, which is consistent with previous findings (Finn et al., 2017, 2015; Vanderwal et al., 2017). It would be useful in future work to determine whether there are particular thresholds for data quantity for fingerprinting algorithms, and whether these differ in children versus adults, and across states (or even different movies). Given the limitations of the data used here, it is possible that individually distinct patterns of FC are present in children as they are in adults, but that more data would be required to identify them in children.

A second potential issue related to individual differences in the pediatric connectome has to do with parcellating the cortex. Here, we used an adult-based resting-state parcellation scheme (Craddock et al., 2012). Increasing the resolution of this parcellation did increase matching accuracies, for example, from 44% to a ceiling of 70% for Rest-Rest matches, indicating that some of the patterns of FC that drive successful matches improve with finer parcels (Figure 3A). In a developmental sample, this could be a significant issue, and using better parcellations that are defined within the study sample (Shen et al., 2013), or that are condition-specific (Salehi et al., 2019) may reveal that individually distinct patterns in FC are robustly present during childhood if small differences in topography are better accounted for. Another potential avenue to fully optimize this approach would be to use hyperalignment (Conroy et al., 2013; Guntupalli et al., 2018; Haxby et al., 2011; Nastase et al., 2019) in a developmental sample, though the data duration (upwards of 20 minutes) required for this approach—even with movie-watching—remains a challenge.

A third limitation related to this finding is that the only true matched pairing available in the algorithm was between Rest1 and Rest2, as the data set does not include two runs of the same movie. One of the hypotheses that originally motivated this study was that individually distinct patterns of functional connectivity may exist in children when the brain is being driven under naturalistic conditions, but that those connectivity patterns do not yet (either via Hebbian learning or repeated use, etc. (Keysers and Perrett, 2004)) persist as part of the intrinsic functional connectome. Whether the pediatric connectome is markedly more stable across runs of the same movie relative to rest, or relatedly, whether matching accuracies would be higher for pairings using the same movie in children, remain open questions.

Finally, head motion artifact is a major issue in any developmental study of functional connectivity (Grayson and Fair, 2017; Power et al., 2012), and because fingerprinting algorithms rely on a ratio of within and between subject measures of FC, head motion may be confounding in complex ways. The head motion treatment in these analyses was quite strict: we excluded any participant with greater than 0.3mm mean framewise displacement, even if they only exceeded the threshold in a single run. We also used ICA-AROMA to remove motion-related artifact (Pruim et al., 2015), and tested for residual motion effects by running the algorithm with just motion-based outputs (Finn et al., 2015), and in a subsample where motion across conditions was matched. Consequently, we feel confident that head motion is not playing a major role in our findings. However, the robust and intractable relationship between age and head motion remains a concern. Future studies in high-motion groups that use movie-watching could help determine whether motion correction strategies that leave the time-course intact (as opposed to motion censoring in which volumes are removed) are superior and could further investigate the role of motion artifact in these types of algorithms.

Connectome stability and similarity as developmental measures

We calculated six secondary measures of whole-brain correlations: three within-subject measures of FC stability (cross-rest, cross-state, and cross-movie), and three between-subject measures of FC similarity

during Rest1, MovieDM, and MovieTP. Analysis of these measures revealed that the different types of stability are correlated, as are similarity under different conditions. Further, if a participant has high within-subject stability, they generally also exhibit a higher degree of similarity with the connectomes of the other participants (Figure 4C). These results suggest that these organizational characteristics are interrelated, and possibly interdependent, in the maturation of the pediatric connectome. A remaining consideration is that some form of noise beyond head motion could explain the interrelatedness of these measures.

Aspects of neural stability, or conversely, within-subject measures of variability, have long been of interest in developmental research. McIntosh et al. argue that neural variability provides options and efficiency to the brain (McIntosh et al., 2008). They showed that variability of evoked potentials during task-based EEG increased with age, and that increasing neural variability accompanies the decreasing behavioral variability and improved task performance. Somewhat aligned with this idea but using fMRI, Hutchison et al. investigated the dynamics of FC organization during task and rest in participants ages 9-30 years old (Hutchison and Morton, 2015). During rest, adults and children expressed the same number of network configurations or states, but during task, adults exhibited a larger repertoire of states, as well as more (and faster) transitions between states. However, in children ages 9-32 years, Marusak et al. found a positive association between age and the temporal variability between functional connectivity networks during rest (Marusak et al., 2017). In one of the early pediatric movie-watching studies, Emerson et al. compared FC during rest and movie-watching in healthy six-year old children (Emerson et al., 2015). Using both independent component and seed-based analyses, they showed that three networks (frontal control, default and dorsal attention) reconfigure during movie-watching such that frontal control-default network connectivity is upregulated, and frontal control to dorsal attention network connectivity is downregulated. They observed the same pattern in adults but noted that the amount of change from resting state to movie-watching (i.e., the cross-state shift in FC) is less pronounced for children relative to adults. While these findings are somewhat contradictory to the idea that cross-state differences are more pronounced for children, they do suggest that age-based differences in FC organization are more prominent during task/movie than resting state in normal development. They also underscore the importance of investigating cross-state differences in FC across development.

Studies of neural similarity during child development are also emerging. When studying between-subject measures of EEG signal during naturalistic viewing, Petroni et al. showed that intersubject correlations (ISCs) of evoked-responses in participants ages 5-44 years decreased with age, and also, that males generally had lower ISCs relative to females (Petroni et al., 2018). In other words, neural variability across subjects in these longer time courses decreased with age. In fMRI, Cantlon and Li showed that ISCs of BOLD-signal responses during naturalistic viewing were generally lower for children relative to adults, but that children had higher ISCs in a particular region (superior temporal cortex) (Cantlon and Li, 2013). In an fMRI study at the other end of the developmental spectrum (ages 18-88 years), network-based ISCs during movie-watching were found to decrease with advancing age (Campbell et al., 2015). In psychiatric populations, adults with autism spectrum disorder exhibited higher ISCs of BOLD-signal time-courses during movie-watching relative to healthy controls (Hasson et al., 2009; Salmi et al., 2013), and similar patterns have been found in individuals with schizophrenia when watching humorous movie clips (Tu et al., 2019) and in major depressive disorder across positive, negative and neutral clips (Guo et al., 2015). Overall, these longer time-course analyses (ISCs) suggest that neural variability across participants is higher when the brain is immature or during advanced aging, as well as when individuals have psychiatric disorders. Some data indicate that regional differences exist within this overall pattern, and particularly in psychiatric disorders, the movie content may be an important factor in these measures. It is not yet known if increasing ISCs in development correspond with emerging stability or similarity of FC relationships.

Social scores and functional connectivity
Using a principal component analysis, we showed that the variance captured in these large-scale, interrelated secondary brain measures of stability and similarity relates to a measure of social skills. Post hoc analyses showed that better social skills correlate with greater stability across resting state runs ($p=0.0007$), and also with greater between-subject similarity of the connectome during the *Despicable Me* movie run ($p=0.0026$). Similar relationships were found with CBCL scores of general psychiatric problems as well. We did not find relationships with age or IQ, and we did not find significant relationships with the other movie data, from *The Present*.

The finding that two of the whole-brain secondary measures of stability and similarity correlate with social responsiveness scores meshes with the central role that social cognition plays in the developing brain. Marked improvement in social cognitive skills during childhood and adolescence is generally considered to be the dominant developmental focus of this highly dynamic age range (Blakemore, 2008). The brain regions implicated in "the social brain" are some of the most densely connected in the cortex (McCormick et al., 2018; Oldham and Fornito, 2019; Saxe et al., 2009), and previous studies have shown that significant changes in FC organization of these regions occurs during development (Richardson et al., 2018). Recently, Lake et al. used Connectome Predictive Modeling to identify a complex set of FC edges that predict SRS scores in a large sample of children and adolescents with and without autism spectrum disorder (Lake et al., 2019). The SRS network identified by the machine-learning algorithm was predictive of SRS scores of unseen participants, even when those participants were from a separate study and had attention-deficit hyperactivity disorder. These transdiagnostic results indicate a surprisingly tight coupling between intrinsic functional connectivity and social skills. Our findings here support the idea that SRS scores relate closely to large-scale functional connectivity relationships in the developing brain, both during rest and movie-watching conditions.

One complication of the brain-behavior correlations identified here is that, contrary to our hypothesis, stability and similarity were not correlated with age. Moreover, none of the behavioral measures tested were correlated with age, and the only robust relationship observed with age was head motion. The brain-behavior correlation analyses were conducted with only 119 participants, and given the clinical heterogeneity of the sample, it is possible that age-relationships would emerge with a larger sample. It is also possible that because the sample is psychiatrically enriched, it is maturationally heterogeneous to the degree that age is not a meaningful developmental marker. That is potentially a controversial suggestion, particularly as SRS could then be construed as being more meaningful than age in this regard. Either way, further studies with larger samples are needed to test the brain-behavior relationship observed here.

We also note that the between-subject similarity during movies predicted social and psychiatric scores only during one of the movies. Here, we point to differences with the stimuli as possible explanations for this finding. *The Present* was shown last during the scanning session, whereas *Despicable Me* was shown second after just one run of rest. It is possible that the data for MovieTP suffered because of this order effect even beyond the movie's elevated head motion. There are also relevant differences between the movies themselves. For example, *Despicable Me* is a social, highly verbal movie that features children of different ages and sexes. *The Present* is largely nonverbal and features a single young adolescent male. It is possible that *Despicable Me* may have been more engaging and more "whole-brain" overall, and that this combined with the order effect of MovieTP makes the FC data derived from *Despicable Me* more meaningful. A more sobering possibility is that the SRS relationship stems from some idiosyncratic, stimulus-specific aspect of MovieDM (Westfall et al., 2017), though this concern would not apply to the cross-rest stability correlation. Overall, these findings suggest that the processes that produce cross-rest stability and between-subject similarity of the connectome also support social cognitive development.

Conclusions:
Based on the fingerprinting findings here, and on the subsequent analyses using whole-brain measures of stability and similarity, we suggest that the developmental reorganization of the functional connectome during childhood and adolescence is accomplishing congruent trajectories, meaning it is becoming both more stable within-subjects (across resting state runs and across states) and more similar between-subjects in a potentially interrelated way. Moreover, it seems that a connectome that is both stable and similar also contains individually distinct patterns of FC—perhaps as a byproduct of attaining the rich neural variability that maximizes efficiency within an individual's brain. The strongest finding in these data was that the stability of the connectome across resting state runs relates to a measure of social skills, and we infer that the same processes that support the stabilization of the connectome also support social cognitive development. As above, significant caveats attend these findings. Particularly because this sample was both developmental and psychiatrically enriched, and because these findings did not relate to age, further investigation is needed. Ideally, longitudinal data would be used to characterize the trajectory of connectome stability both across rest runs and across different states—and to track how stability relates to both connectome similarity and behavioral measures of development—over time.

Works Cited

Achenbach, T.M., 1999. The Child Behavior Checklist and related instruments, in: The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, 2nd Ed. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 429–466.

Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Bronstein, B., Bui, A., Bushey, M., Castagna, V., Camacho, N., Chan, E., Citera, D., Clucas, J., Cohen, S., Eaves, M., Fradera, B., Grant-Villegas, N., Green, G., Gregory, C., Hart, E., Harris, S., Lord, C., Kahn, D., Kabotyanski, K., Kleinman, K., Koo, B., Kramer, E., Margolis, A., Merikangas, K., Milham, J., Minniti, G., Neuhaus, R., Nussbaum, A., Osman, Y., Parra, L., Pugh, K.R., Racanello, A., Restrepo, A., Saltzman, T., Septimus, B., Tobe, R., Waltz, R., Williams, A., Yeo, A., Castellanos, F.X., Klein, A., Paus, T., Leventhal, B.L., Craddock, C., Koplewicz, H.S., Milham, M.P., 2017. The Healthy Brain Network Biobank: An open resource for transdiagnostic research in pediatric mental health and learning disorders. bioRxiv. https://doi.org/10.1101/149369

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Blakemore, S.J., 2008. Development of the social brain during adolescence. Q J Exp Psychol Hove 61, 40–9. https://doi.org/10.1080/17470210701508715

Blakemore, S.-J., Frith, U., 2004. How does the brain deal with the social world? Neuroreport 15, 119–128. https://doi.org/10.1097/00001756-200401190-00024

Braams, B.R., Crone, E.A., 2017. Longitudinal Changes in Social Brain Development: Processing Outcomes for Friend and Self. Child Dev. 88, 1952–1965. https://doi.org/10.1111/cdev.12665

Campbell, K.L., Shafto, M.A., Wright, P., Tsvetanov, K.A., Geerligs, L., Cusack, R., Tyler, L.K., Brayne, C., Bullmore, E., Calder, A., Cusack, R., Dalgleish, T., Duncan, J., Henson, R., Matthews, F., Marslen-Wilson, W., Rowe, J., Shafto, M., Campbell, K., Cheung, T., Davis, S., Geerligs, L., Kievit, R., McCarrey, A., Price, D., Taylor, J., Tsvetanov, K.,

Williams, N., Bates, L., Emery, T., Erzinçlioglu, S., Gadie, A., Gerbase, S., Georgieva, S., Hanley, C., Parkin, B., Troy, D., Allen, J., Amery, G., Amunts, L., Barcroft, A., Castle, A., Dias, C., Dowrick, J., Fair, M., Fisher, H., Goulding, A., Grewal, A., Hale, G., Hilton, A., Johnson, F., Johnston, P., Kavanagh-Williamson, T., Kwasniewska, M., McMinn, A., Norman, K., Penrose, J., Roby, F., Rowland, D., Sargeant, J., Squire, M., Stevens, B., Stoddart, A., Stone, C., Thompson, T., Yazlik, O., Dixon, M., Barnes, D., Hillman, J., Mitchell, J., Villis, L., Tyler, L.K., 2015. Idiosyncratic responding during movie-watching predicted by age differences in attentional control. Neurobiol. Aging 36, 3045–3055. https://doi.org/10.1016/j.neurobiolaging.2015.07.028

Cantlon, J.F., Li, R., 2013. Neural Activity during Natural Viewing of Sesame Street Statistically Predicts Test Scores in Early Childhood. PLoS Biol. 11. https://doi.org/10.1371/journal.pbio.1001462

Coffin, P., Renaud, C., 2010. Despicable Me. Universal Pictures.

Conroy, B.R., Singer, B.D., Guntupalli, J.S., Ramadge, P.J., Haxby, J.V., 2013. Inter-subject alignment of human cortical anatomy using functional connectivity. NeuroImage 81, 400–411. https://doi.org/10.1016/j.neuroimage.2013.05.009

Constantino, J.N., Davis, S.A., Todd, R.D., Schindler, M.K., Gross, M.M., Brophy, S.L., Metzger, L.M., Shoushtari, C.S., Splinter, R., Reich, W., 2003. Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. J Autism Dev Disord 33, 427–33.

Craddock, R.C., James, G.A., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S., 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum Brain Mapp 33, 1914–28. https://doi.org/10.1002/hbm.21333

Damoiseaux, J.S., Rombouts, S.A., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks across healthy subjects. Proc Natl Acad Sci U A 103, 13848–53. https://doi.org/10.1073/pnas.0601417103

Emerson, R., Short, S., Lin, W., Gilmore, J., Gao, W., 2015. Network-Level Connectivity Dynamics of Movie Watching in 6-Year-Old Children. Front. Hum. Neurosci. 9. https://doi.org/10.3389/fnhum.2015.00631

Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J., Durnez, J., Poldrack, R.A., Gorgolewski, K.J., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat Methods 16, 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fair, D.A., Cohen, A.L., Dosenbach, N.U.F., Church, J.A., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., Schlaggar, B.L., 2008. The maturing architecture of the brain's default network. Proc. Natl. Acad. Sci. 105, 4028. https://doi.org/10.1073/pnas.0800376105

Fair, D.A., Dosenbach, N.U.F., Church, J.A., Cohen, A.L., Brahmbhatt, S., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., Schlaggar, B.L., 2007. Development of distinct control networks through segregation and integration. Proc. Natl. Acad. Sci. 104, 13507–13512. https://doi.org/10.1073/pnas.0705843104

Finn, E.S., Scheinost, D., Finn, D.M., Shen, X., Papademetris, X., Constable, R.T., 2017. Can brain state be manipulated to emphasize individual differences in functional connectivity? NeuroImage 160, 140–151. https://doi.org/10.1016/j.neuroimage.2017.03.064

Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci 18, 1664–71. https://doi.org/10.1038/nn.4135

Frey, J., 2014. The Present.

Grayson, D.S., Fair, D.A., 2017. Development of large-scale functional networks from birth to adulthood: A guide to the neuroimaging literature. NeuroImage 160, 15–31. https://doi.org/10.1016/j.neuroimage.2017.01.079

Guntupalli, J.S., Feilong, M., Haxby, J.V., 2018. A computational model of shared fine-scale structure in the human connectome. PLOS Comput. Biol. 14, e1006120. https://doi.org/10.1371/journal.pcbi.1006120

Guo, C.C., Nguyen, V.T., Hyett, M.P., Parker, G.B., Breakspear, M.J., 2015. Out-of-sync: disrupted neural activity in emotional circuitry during film viewing in melancholic depression. Sci. Rep. 5, 11605. https://doi.org/10.1038/srep11605

Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., Behrmann, M., 2009. Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. Autism Res. 2, 220–231. https://doi.org/10.1002/aur.89

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–40. https://doi.org/10.1126/science.1089506

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. Neuron 72, 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Hutchison, R.M., Morton, J.B., 2015. Tracking the Brain's Functional Coupling Dynamics over Development. J. Neurosci. Off. J. Soc. Neurosci. 35, 6849–6859. https://doi.org/10.1523/JNEUROSCI.4638-14.2015

Kaufmann, T., Alnaes, D., Doan, N.T., Brandt, C.L., Andreassen, O.A., Westlye, L.T., 2017. Delayed stabilization and individualization in connectome development are related to psychiatric disorders. Nat Neurosci 20, 513–515. https://doi.org/10.1038/nn.4511

Keysers, C., Perrett, D.I., 2004. Demystifying social cognition: a Hebbian perspective. Trends Cogn. Sci. 8, 501–507. https://doi.org/10.1016/j.tics.2004.09.005

Lake, E.M.R., Finn, E.S., Noble, S.M., Vanderwal, T., Shen, X., Rosenberg, M.D., Spann, M.N., Chun, M.M., Scheinost, D., Constable, R.T., 2019. The Functional Brain Organization of an Individual Allows Prediction of Measures of Social Abilities Transdiagnostically in Autism and Attention-Deficit/Hyperactivity Disorder. Biol Psychiatry 86, 315–326. https://doi.org/10.1016/j.biopsych.2019.02.019

Marusak, H.A., Calhoun, V.D., Brown, S., Crespo, L.M., Sala-Hamrick, K., Gotlib, I.H., Thomason, M.E., 2017. Dynamic functional connectivity of neurocognitive networks in children. Hum. Brain Mapp. 38, 97–108. https://doi.org/10.1002/hbm.23346

McCormick, E.M., van Hoorn, J., Cohen, J.R., Telzer, E.H., 2018. Functional connectivity in the social brain across childhood and adolescence. Soc Cogn Affect Neurosci 13, 819–830. https://doi.org/10.1093/scan/nsy064

McIntosh, A.R., Kovacevic, N., Itier, R.J., 2008. Increased Brain Signal Variability Accompanies Lower Behavioral Variability in Development. PLoS Comput. Biol. 4, e1000106. https://doi.org/10.1371/journal.pcbi.1000106

Mills, K.L., Lalonde, F., Clasen, L.S., Giedd, J.N., Blakemore, S.-J., 2014. Developmental changes in the structure of the social brain in late childhood and adolescence. Soc. Cogn. Affect. Neurosci. 9, 123–131. https://doi.org/10.1093/scan/nss113

Nastase, S.A., Liu, Y.-F., Hillman, H., Norman, K.A., Hasson, U., 2019. Leveraging shared connectivity to aggregate heterogeneous datasets into a common response space. bioRxiv 741975. https://doi.org/10.1101/741975

Oldham, S., Fornito, A., 2019. The development of brain network hubs. Dev Cogn Neurosci 36, 100607. https://doi.org/10.1016/j.dcn.2018.12.005

Petroni, A., Cohen, S.S., Ai, L., Langer, N., Henin, S., Vanderwal, T., Milham, M.P., Parra, L.C., 2018. The Variability of Neural Responses to Naturalistic Videos Change with Age and Sex. eNeuro 5. https://doi.org/10.1523/ENEURO.0244-17.2017

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59, 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Pruim, R.H.R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. Neuroimage 112, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., Saxe, R., 2018. Development of the social brain from age three to twelve years. Nat Commun 9, 1027. https://doi.org/10.1038/s41467-018-03399-2

Salehi, M., Greene, A.S., Karbasi, A., Shen, X., Scheinost, D., Constable, R.T., 2019. There is no single functional atlas even for a single individual: Parcellation of the human brain is state dependent. bioRxiv 431833. https://doi.org/10.1101/431833

Salmi, J., Roine, U., Glerean, E., Lahnakoski, J., Nieminen-von Wendt, T., Tani, P., Leppämäki, S., Nummenmaa, L., Jääskeläinen, I.P., Carlson, S., Rintahaka, P., Sams, M., 2013. The brains of high functioning autistic individuals do not synchronize with those of others. NeuroImage Clin. 3, 489–497. https://doi.org/10.1016/j.nicl.2013.10.011

Saxe, R.R., Whitfield-Gabrieli, S., Scholz, J., Pelphrey, K.A., 2009. Brain regions for perceiving and reasoning about other people in school-aged children. Child Dev 80, 1197–209. https://doi.org/10.1111/j.1467-8624.2009.01325.x

Shen, X., Tokoglu, F., Papademetris, X., Constable, R.T., 2013. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. NeuroImage 82, 403–415. https://doi.org/10.1016/j.neuroimage.2013.05.081

Somandepalli, K., Kelly, C., Reiss, P.T., Zuo, X.N., Craddock, R.C., Yan, C.G., Petkova, E., Castellanos, F.X., Milham, M.P., Di Martino, A., 2015. Short-term test-retest reliability of resting state fMRI metrics in children with and without attention-deficit/hyperactivity disorder. Dev Cogn Neurosci 15, 83–93. https://doi.org/10.1016/j.dcn.2015.08.003

Thomason, M.E., Dennis, E.L., Joshi, A.A., Joshi, S.H., Dinov, I.D., Chang, C., Henry, M.L., Johnson, R.F., Thompson, P.M., Toga, A.W., Glover, G.H., Van Horn, J.D., Gotlib, I.H., 2011. Resting-state fMRI can reliably map neural networks in children. Neuroimage

55, 165–75. https://doi.org/S1053-8119(10)01558-2 [pii] 10.1016/j.neuroimage.2010.11.080

Tu, P.-C., Su, T.-P., Lin, W.-C., Chang, W.-C., Bai, Y.-M., Li, C.-T., Lin, F.-H., 2019. Reduced synchronized brain activity in schizophrenia during viewing of comedy movies. Sci. Rep. 9, 1–11. https://doi.org/10.1038/s41598-019-48957-w

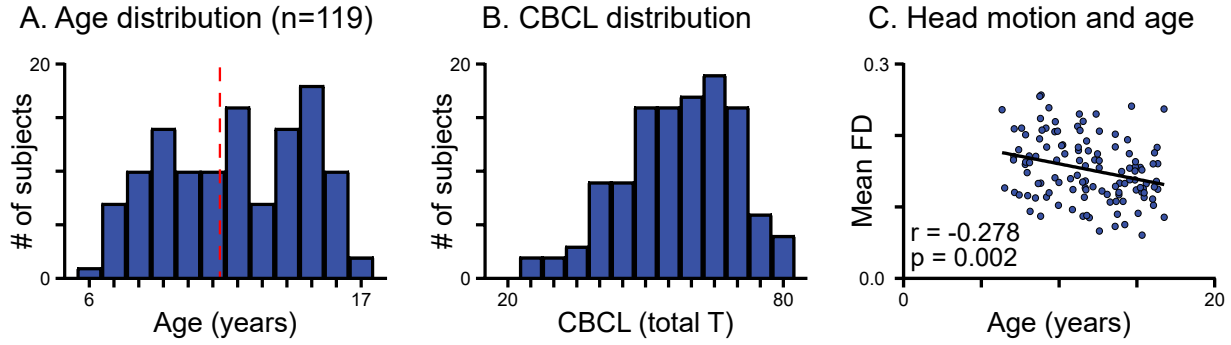
Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Turnbull, A., Castellanos, F.X., 2017. Individual differences in functional connectivity during naturalistic viewing conditions. Neuroimage 157, 521–530. https://doi.org/10.1016/j.neuroimage.2017.06.027

Wang, Z., Sair, H., Crainiceanu, C., Lindquist, M., Landman, B.A., Resnick, S., Vogelstein, J.T., Caffo, B., 2018. On statistical tests of functional connectome fingerprinting. bioRxiv.

Westfall, J., Nichols, T.E., Yarkoni, T., 2017. Fixing the stimulus-as-fixed-effect fallacy in task fMRI. Wellcome Open Res. 1. https://doi.org/10.12688/wellcomeopenres.10298.2

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106, 1125–65. https://doi.org/10.1152/jn.00338.2011

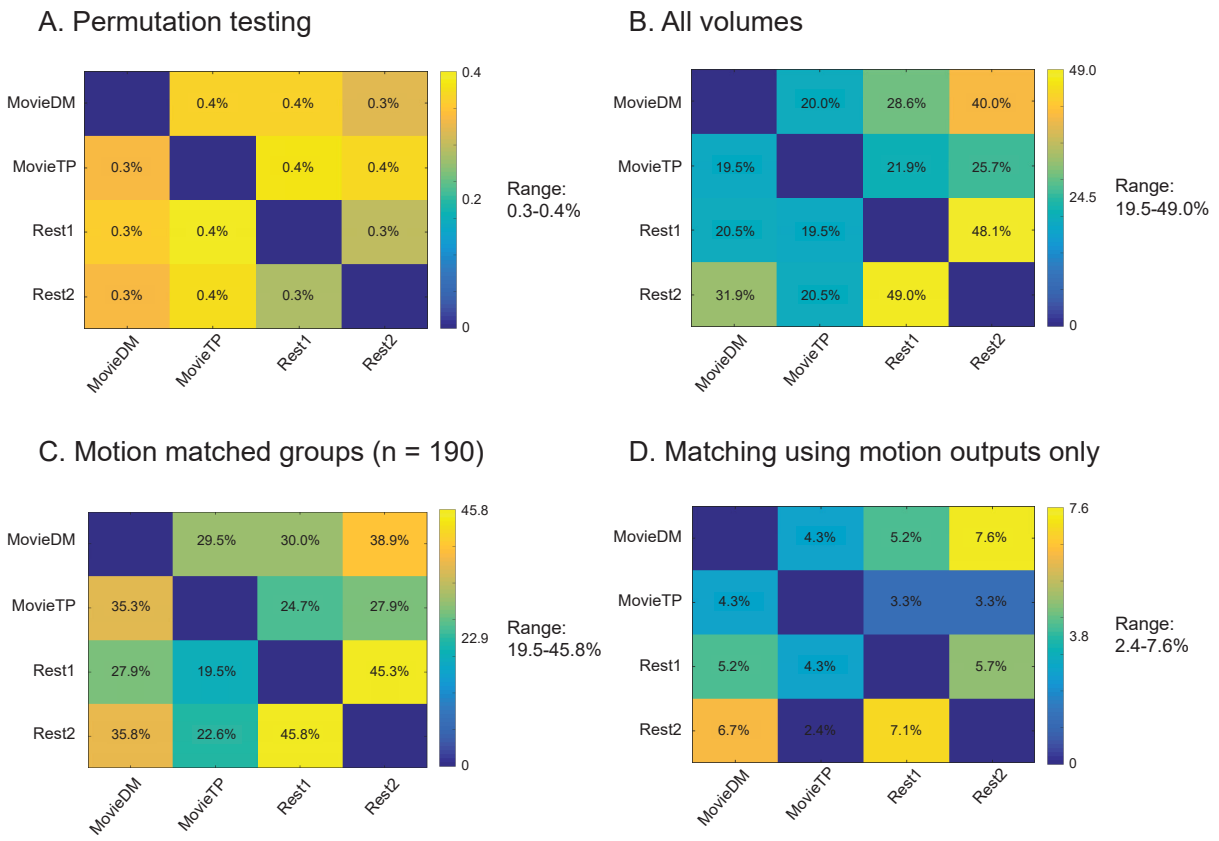
Supplementary Materials



D. Behavioral measures of interest

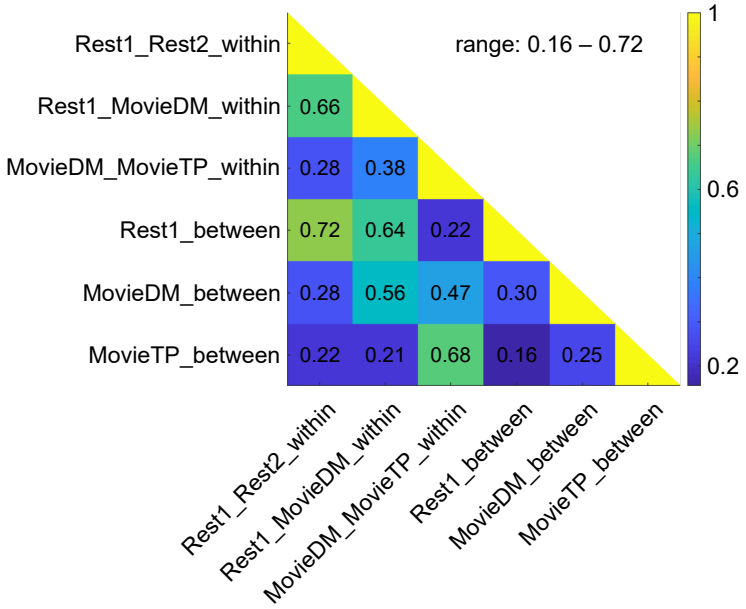
	CBCL total T	WISC-V FSIQ	SRS-2 total T
Age: r (p)	0.112 (0.224)	-0.067 (0.467)	0.071 (0.445)
Head motion: r (p)	0.065 (0.485)	-0.001 (0.990)	0.013 (0.891)

Supplementary Figure 1s: Demographics for brain-behavior subsample, n=119. Age distribution, the right-skew of CBCL T-scores (in which greater scores indicate greater overall behavioral problems and psychopathology), and the association between advancing age and decreasing head motion are similar to the full 210 sample shown in Figure 1. D shows that there are no significant correlations among the three behavioral measures and either age or mean framewise displacement.



Supplementary Figure 2s: Secondary test-retest matching algorithm analyses. To further interrogate the results for the main test-retest matching algorithm reported in Figure 2, we conducted a number of other fingerprinting algorithms. A) Results of running the algorithm with permuted using falsely assigned identity pairs. B) The main analyses used matrices truncated at 250 volumes. Here, we reran the algorithm using all available volumes, so 750 volumes for MovieDM, and 375 volumes for Rest1 and Rest2. C) There was a significant difference in mean framewise displacement across conditions, so here we reran the algorithm using a subset of subjects that created no difference in motion across conditions. D) Here we reran the algorithm using only motion parameters.

A. Stability and similarity: correlations of all pairings



B. Stability and similarity: age and head motion

		Age: r (p)	FD: r (p)
Connectome Stability	cross-rest	0.01 (0.91)	**-0.20 (4.13 x 10^{-3})**
	cross-movie	0.15 (0.03)	**-0.38 (2.00 x 10^{-8})**
	cross-state	0.07 (0.30)	**-0.29 (2.14 x 10^{-5})**
Connectome Similarity	Rest1	0.01 (0.89)	-0.15 (0.03)
	MovieDM	0.16 (0.02)	**-0.41 (6.96 x 10^{-10})**
	MovieTP	0.10 (0.15)	**-0.30 (8.53 x 10^{-6})**

Supplementary Figure 3s: All pairwise correlations among the six measures of connectome stability and similarity. Six secondary measures of whole-brain FC were calculated for each subject, three each of within-subject stability and between-subject similarity. As shown in Figure 4, cross-rest and cross-state stability are correlated, and similarity during Rest1 and MovieDM are correlated. Moreover, mean stability and mean similarity are correlated ($r = 0.68$, $p = 2.5 \times 10^{-29}$). For the sake of completion, we provide numerical values for all pairwise correlations of the six measures of stability and similarity here. Relationships with framewise displacement (FD) and age are shown in B.