The Healthy Brain Network Serial Scanning Initiative: A resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions

David O’Connor¹,², Natan Vega Potler¹, Meagan Kovacs¹, Ting Xu¹, Lei Ai¹, John Pellman¹,², Tamara Vanderwal³, Lucas Parra⁴, Samantha Cohen⁵, Satrajit Ghosh⁶, Jasmine Escalera¹, Natalie Grant-Villegas¹, Yael Osman¹, Anastasia Bui¹, R. Cameron Craddock¹,², Michael P. Milham¹,²*

2. Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York
3. Yale University, New Haven, Connecticut
4. City College of New York, New York, New York
5. The Graduate Center of the City University of New York, New York, New York
6. Massachusetts Institute of Technology, Cambridge, Massachusetts

*Correspondence:

Michael Peter Milham, MD, PhD
Center for Developing Brain
Child Mind Institute
New York, NY 10022, USA
Michael.Milham@childmind.org (M.P. Milham)
Abstract

Background
Although typically measured during the resting state, a growing literature is illustrating the ability to map intrinsic connectivity in task and naturalistic viewing fMRI paradigms. These paradigms are drawing excitement due to their greater tolerability in clinical and developing populations and because they enable a wider range of analyses (e.g. inter-subject correlations). To be clinically useful, the test-retest reliability of connectivity measured during these paradigms needs to be established. This resource provides data for evaluating test-retest reliability for full-brain connectivity patterns detected during each of four scan conditions that differ with respect to level of engagement (rest, abstract animations, movie clips, flanker task). Data is provided for thirteen participants, each scanned in twelve sessions with 10 minutes for each scan of the four conditions. Diffusion kurtosis imaging data was also obtained at each session.

Findings
Technical validation and demonstrative reliability analyses found that variation in intrinsic functional connectivity across sessions was greater than that attributable to scan condition. Between-condition reliability was generally high, particularly for the frontoparietal and default networks. Between-session reliabilities obtained separately for the different scan conditions were comparable, though notably lower than between-condition reliabilities.

Conclusions
The described resource provides a test-bed for quantifying the reliability of connectivity indices across conditions and time. The resource can be used to compare and optimize different frameworks for measuring connectivity and data collection parameters such as scan length. Additionally, investigators can explore the unique perspectives of the brain’s functional architecture offered by each of the scan conditions.

DATA NOTE

Purpose of Data Collection
An extensive literature has documented the utility of fMRI for mapping the brain’s functional interactions through the detection of temporally correlated patterns of spontaneous activity between spatially distinct brain areas [1]–[7]. Commonly referred to as intrinsic functional connectivity (iFC), these patterns are commonly studied during the ‘resting state’, which involves the participant quietly lying awake and not performing an externally driven task. Resting state fMRI (R-fMRI) has gained popularity in clinical neuroimaging due to its minimal task and participant compliance demands. R-fMRI has also demonstrated good test-retest reliability for commonly used measures [8]–[12], and utility in detecting brain differences associated with neuropsychiatric disorders [13], [14]. Despite these successes, a growing body of work is questioning the advantages of resting state, given reports of higher head motion, decreased tolerance of the scan environment (e.g. boredom, rumination), and increased likelihood of falling asleep compared to more engaging task-based fMRI paradigms [15]–[18]. This is particularly relevant for studies of pediatric, geriatric and clinical populations, all of which are characterized by lower tolerance of the scanner environment.

A number of less challenging scan conditions have been proposed as alternatives for estimating iFC. Particularly intriguing are “naturalistic viewing” paradigms [15], [19], [20]. It has been shown that the mental state (i.e., emotional state, performing a task, etc.) of the
participant during scanning can effect iFC patterns; recent work suggests that low engagement states (e.g., computer animations with limited cognitive content) may come close to mimicking rest from a neural perspective [21]. Several studies have illustrated the ability to relate trait phenotypic variables to inter-individual differences in iFC across conditions, even if extrinsically driven signals (i.e., task stimulus functions) are not removed [21]–[27]. However, comprehensive comparisons of the relative impact of scan condition on detection of inter-individual differences in intrinsic functional connectivity, and the test-retest reliability of these differences, are needed before these paradigms can fully supplant R-fMRI.

Here we describe a dataset that was generated as part of a pilot testing effort for the Child Mind Institute Healthy Brain Network – a large-scale data collection effort focused on the generation of an open resource for studying child and adolescent mental health. The primary goal of the data collection was to assess and compare test-retest reliability of full-brain connectivity patterns detected for each of four scan conditions that differed with respect to level of engagement. Specifically, 13 participants were scanned during each of the following four conditions on 12 different occasions: 1) rest, 2) free viewing of abstract computer graphics and sounds designed to have minimal cognitive or emotional content (i.e., "Inscape" [15]), 3) free viewing of highly engaging movies [19], and 4) performance of an active task (i.e., an Erickson flanker task [28], with no-Go trials included). For each of the non-rest conditions, three different stimuli were used, with each being repeated four times across the 12 sessions to enable the evaluation of repetition effects. Given the focus on naturalistic viewing, an additional scan session containing a full viewing of “Raiders of the Lost Ark” was included to facilitate interested parties in the exploration and evaluation of increasingly popular hyper alignment approaches, which offer unique solutions to matching brain function across individuals [29].

Although not a primary focus of the data collection, additional structural imaging data was collected, which are being shared as well: 1) MPRAGE [30], 2) diffusion kurtosis imaging [31], [32], 3) quantitative T1/T2 anatomical imaging (single session) [33], 4) magnetization transfer (single session) [34] (see Table 1). Functional MRI data from a single movie viewing session during which Raiders of the Lost Arc was viewed in its entirety, is included as well.

**METHODS**

*Participants and Procedures*

13 adults (ages 18-45 years; mean age: 30.3; 38.4% male) recruited from the community participated in the Healthy Brain Network’s Serial Scanning Initiative. Each participant attended 14 sessions over a period of 1-2 months; see Table 1 for the breakdown of data acquired across sessions. All imaging data were collected using a 1.5T Siemens Avanto equipped with a 32-channel head coil in a mobile trailer (Medical Coaches, Oneonta, NY). The scanner was selected as part of a pilot initiative being carried out to evaluate the capabilities of a 1.5T mobile scanner when equipped with a state-of-the-art head coil and imaging sequences. All research performed was approved by the Chesapeake Institutional Review Board, Columbia, MD (https://www.chesapeakeirb.com/).
**Experimental Design.** As outlined in Table 1, each participant attended a total of 14 separate imaging sessions; these included: 1) a baseline characterization session containing a variety of quantitative anatomical scans, 2) 12 serial scanning sessions, each using the same imaging protocol consisting of four functional MRI scan conditions (10 minutes per condition), diffusion kurtosis imaging and a reference MPRAGE anatomical scan, and 3) a ‘Raiders of the Lost Arc’ movie viewing session.

**Functional MRI Scan Conditions Included in Serial Scanning.** The following four functional scan conditions were selected to sample a range of levels of engagement, presented in ascending order of level of engagement (See Figure 1):

**Rest**
The participant was presented a white fixation cross in the center of a black screen and instructed to rest with eyes open. Specific instructions were as follows: “Please lie quietly with your eyes open, and direct your gaze towards the plus symbol. During this scan let your mind wander. If you notice yourself focusing on a particular stream of thoughts, let your mind wander away.”

**Inscape**
Inscape is a computer generated animation comprised of abstract, non-social, technological-looking 3D forms that transition in a continuous fashion without scene cuts. Visual stimulation is accompanied by repetitive, slow tempo (48 bpm) music based on the pentatonic scale, which was previously selected based on calming influences and to harmonize with the noise generated by EPI sequences [15]. Three unique 10 minute Inscape scenarios were presented across the 12 repeat scanning sessions.

**Movie**
Three unique 10-minute movie clips were presented across the 12 repeat scanning sessions. To ensure a high level of engagement, three Hollywood movie clips were selected, each representing a different movie genre. The specific clips selected were: Wall-E (time codes 00:02:03:13 to 00:12:11:05), The Matrix (00:25:23:10 to 00:35:19:20), and A Few Good Men (01:58:13:01 to 02:08:11:18).

**Flanker**
The Eriksen Flanker task consisted of presenting a series of images containing 5 arrows. For each image, the participant was asked to focus on the center arrow and indicate if it is pointing left or right by pushing a button with their left or right index finger. The flanking arrows could be pointing the same way (congruent) or the opposite way (incongruent). Also built into the task were a neutral stimulus and a go/no-go aspect. The neutral task would contain diamonds instead of flanking arrows, making the central arrow direction more obvious. The no-go stimuli contains x’s instead of flanking arrows, indicating the subject should not push either button. See Figure 1 for a visualization of the stimuli.

**Counter-Balancing**
Order effects are an obvious concern when comparing the four functional scan conditions. To minimize these effects, we ensured that for each participant: 1) each scan type occurred an equal number of times in each of the four scan slots across the 12 sessions, and that 2) each scan type had an equal frequency of being preceded by each of the other three scan types. We made use of 3 exemplars of each non-rest stimuli to enable the examination of repetition effects. For movies, this involved having three 10-minute clips, each from a different movie; for inscapes, this involved three different animation sequences and for the
flanker task, three different stimulus orderings were used. We guaranteed that across the 12 scan sessions, each exemplar occurred one time across every three scan sessions. Specific ordering of exemplars were varied across ‘odd’ and ‘even’ numbered participants. For each participant, individual-specific ordering information is provided in the release.

**Imaging Protocols (See Table 2 for scan protocol details).**

- **Functional MRI (sessions 1-14):** For all functional MRI scans, the multiband EPI sequence provided by CMRR [35] was employed to provide high spatial and temporal resolutions (multiband factor 3, voxel size: 2.46x2.46x2.5mm; TR: 1.46 seconds).
- **MEMPRAGE (sessions 1-7, 9-14):** Across all sessions (except the full-movie session), we obtained a multi-echo MPRAGE sequence for the purposes of anatomical registration [36]. Within a given scan, four echoes are collected per excitation and combined using root mean square average. This enables the images to be acquired with a higher bandwidth to reduce distortion, while recovering SNR through averaging. The added T2* weighting from the later echoes also helps differentiate dura from brain matter.
- **Diffusional kurtosis imaging (DKI):** Leveraging the capabilities of the CMRR multiband imaging sequence, we were able to acquire 64 directions at 2 b-values (1000 and 2000 s/mm²). This enables diffusion kurtosis specific metrics to be calculated from the data in addition to standard DTI metrics and can improve tractography [31].
- **Quantitative Relaxometry MRI (Quantitative T1, T2, and Myelin Water Fraction [MWF]):** DESPOT1 and DESPOT2 sequences were used to characterize microstructural properties of brain tissue. These innovative acquisition strategies enable quantitation of T1 and T2 relaxation constants, which can be combined to calculate myelin water fraction [37].
- **Magnetization Transfer:** High-resolution T1-weighted structural images were acquired with a FLASH sequence, with and without a saturation RF pulse. The magnetization transfer ratio is calculated from the resulting images, which is purportedly sensitive marker of myelination [34].

**DATA RECORDS**

**Data Privacy**

The HBN-SSI data are being shared via the 1000 Functional Connectomes Project and its International Neuroimaging Data-sharing Initiative (FCP/INDI) [38]. Prior to sharing, all imaging data were fully de-identified by removing all personally identifying information (as defined by the Health Insurance Portability and Accountability) from the data files, including facial features. All data were visually inspected before release to insure that these procedures worked as expected.

**Distribution for use**

**Imaging Data**

All MRI data can be accessed through the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC; http://fcon_1000.projects.nitrc.org/indi/hbn_ss) and FCP/INDI’s Amazon Web Services public Simple Storage Service (S3) bucket. In both locations, the imaging data is stored in a series of tar files that can be directly downloaded through a HTTP client (e.g., a web browser, Curl or wget). The data is additionally available on S3 as individual NifTI files for each scan, which can be downloaded using a HTTP client or S3 client software such as cyberduck (https://cyberduck.io).
All imaging data are released in the NIfTI file format; they are organized and named according to the brain imaging data structure (BIDS) format [39].

Phenotypic Data

Partial phenotypic data will be publically available without any requirements for a data usage agreement. This includes age, sex, handedness, the internal state questionnaire, and the New York Cognition Questionnaire [39]. These data are located in a comma separated value (.csv) file accessible via the HBN-SSI website and are included with the BIDS organized imaging data as tab separate values (TSV) files. The remainder of the phenotypic data (see Table 3), including the PANAS [40] and results from the ADHD Quotient system [41], will be made available to investigators following completion of the HBN Data Usage Agreement (DUA). The HBN DUA is modeled after that of the NKI-Rockland Sample and is intended to prevent against data re-identification; it does not place any constraints on the range of analyses that can be carried out using the shared data, or place requirements for co-authorship. Following submission and execution of the data usage agreement, users can access the phenotypic data through the COINS Data Exchange (an enhanced graphical query tool, which enables users to target and download files in accord with specific search criteria) [42].

TECHNICAL VALIDATION

Quality Assessment

Consistent with the established FCP/INDI policy, all completed datasets contributed to HBN-SSI are made available to users regardless of data quality. Justifications for this decision include the lack of consensus within the imaging community on what constitutes good or poor quality data, and the utility of ‘lower quality’ datasets for facilitating the development of artifact correction techniques. For HBN-SSI, the inclusion of datasets with significant artifacts related to factors such as motion are particularly valuable, as it facilitates the evaluation of the impact of such real-world confounds on reliability and reproducibility.

To help users assess data quality, we calculated a variety of quantitative quality metrics from the data using the Preprocessed Connectome Project quality assurance protocol (QAP; http://preprocessed-connectomes-project.org/quality-assessment-protocol). The QAP includes a broad range of quantitative metrics that have been proposed in the imaging literature for assessing data quality [43].

For the structural data, spatial measures include: Signal-to-Noise Ratio (SNR) [44], Contrast-to-Noise Ratio (CNR) [44], Foreground-to-Background Energy Ratio (FBER), Percent artifact voxels (QI1) [45], Spatial smoothness (FWHM) [46], Entropy focus criterion (EFC) [47]. These are shown for different participants in Figure 2. Spatial measures of fMRI data include (Figure 3): EFC, FBER, FWHM, and well as Ghost-to-Signal Ratio (GSR) [48]. Temporal measures of fMRI data include (Figure 4): Mean Frame-wise Displacement (Mean FD) [49], Median Distance Index (Quality) [50], Standardized DVARS (DVARS) [51], Outliers Detection [50], and Global correlation (GCOR) [52]. See Figures 2-4 for a subset of the metrics; the full set of measures are included on the HBN-SSI website in .csv format for download. Review of the QAP profiles led us to exclude 3 participants based on excessively high mean FD from the illustrative analyses presented in the next section. Although not a focus of the current work, visual inspection of the figures points to the potential value of this dataset for establishing the reliability of QAP measures. The impact of scan condition on each of the functional QAP measures was examined using a one-way ANOVA. No significant differences were found for any of the measures. In addition, the test-retest
reliability of each QAP measure, for each condition, was assessed using the intra-class correlation coefficient (ICC). The results are shown in Table 4.

**FMRI Analyses.** A broad range of analyses, including but not limited to evaluations of test-retest reliability, can be performed using the present HBN-SSI dataset. Here, we provide a few illustrative analyses to demonstrate the technical validity and utility of these data; they are not intended to be exhaustive.

**Data preprocessing.**

Prior to image processing, Freesurfer was used to combine the 12 available MPRAGE images into an MRI robust average image for each individual participant. A non-rigid registration between MPRAGE images and a 2mm MNI brain-only template (FSL’s MNI152_T1_2mm_brain.nii.gz, [53]) was calculated using ANTs [54]. Further anatomical processing included with skull stripping using AFNI’s 3dSkullstrip[55] (to include any voxels in the ventricles incorrectly removed by this utility, the brain mask was augmented using a ventricle mask that was generated by reverse transforming the ventricles included in the MNI atlas into native space for each participant). Next, data was processed using a development version of the open-source, Nipype-based [56]- Configurable Pipeline for the Analysis of Connectomes [1] (C-PAC version 0.4.0, [https://www.nitrc.org/frs/downloadlink.php/9275 for image preprocessing configuration file]).

Following resampling of the functional MRI data to RPI orientation, image preprocessing in C-PAC consisted of the following steps: 1) motion correction, 2) boundary-based registration [57], 3) nuisance variable regression (1st and 2nd order polynomial, 24-regressor model of motion [58], mean WM mask signal, mean CSF mask signal). We then extracted representative time series for each ROI in the CC200 atlas[59] (by averaging within-ROI voxel time series). All possible pairwise correlations were calculated amongst ROI time series to generate a ROI-to-ROI connectivity matrix for each scan in each session for each subject. To facilitate ease of presentation and interpretation for our findings, the connections were sorted by intrinsic connectivity network membership, as defined by Yeo et al. [60].

**Fingerprinting.** Prior work by Finn et al. [22] demonstrated the ability to “fingerprint” individuals based on their functional connectivity matrices. Specifically, they found that the level of correlation between connectivity matrices for data obtained from the same participant on different occasions was markedly higher than that observed for connectivity matrices obtained from different participants; this was true regardless of whether functional connectivity was based on resting state or task activation data. Consistent with their work, we found a dramatically higher degree of spatial correlation between connectivity matrices obtained from the same individual on differing sessions, when compared to differing individuals (Figure 5). Also consistent with their findings, we found this to be true regardless of the scan condition employed.

**Connection-Wise Reliability For the Four States.** A key question is how much variation among scan conditions (i.e., between-condition reliability) impacts reliability as opposed to between-session reliability (i.e., test-retest reliability). To address this question, we analyzed the 12 sessions obtained for the 10 participants with minimal head motion using a hierarchical Linear Mixed Model (note: three subjects were missing the flanker task from one session each; these were treated as missing values in our analyses). The hierarchical LMM allows for the estimation of reliability by providing estimates of variance between participants, across the four conditions (for the same participant) and between sessions within each condition.
For a given functional connectivity measurement \( \nu \), \( i{\text{FC}}_{ijk}(\nu) \) is the modeled intrinsic functional connectivity for the i-th session, for the j-th condition of the k-th participant, taking into account condition and session effects. The equation is composed of an intercept \( \mu_{000} \), a random effect between sessions for the j-th condition of k-th participant \( y_{ijk} \), a random effect for the k-th participant \( \delta_k \), and an error term \( \varepsilon_{ijk} \). All of \( y_{ijk}, \delta_k, \text{ and } \varepsilon_{ijk} \) are assumed to be independent, and follow a normal distribution with zero mean. The total variances of iFC can be decomposed into three parts, 1) variance between participants (\( \sigma_y^2 = \text{Var}[y] \)), 2) variance between conditions for the same participant (\( \sigma_A^2 = \text{Var}[y] \)), and 3) variance of the residual; indicating variance between sessions (\( \sigma_\varepsilon^2 = \text{Var}[\varepsilon] \)). The reliability of the iFC across conditions can be calculated as intra-class correlation coefficients as follows (Figure 6, left):

\[
\text{ICC}(\text{between - conditions}) = \frac{\sigma_y^2}{\sigma_y^2 + \sigma_A^2} \tag{2}
\]

and across sessions as follows Figure 6, right):

\[
\text{ICC}(\text{between - sessions, conditions}) = \frac{\sigma_y^2 + \sigma_A^2}{\sigma_y^2 + \sigma_A^2 + \sigma_\varepsilon^2} \tag{3}
\]

Findings revealed impressively high degree of between-condition reliability for most connections (percentiles: 50th: 0.854, 75th: 0.955, 95th: 1), as opposed to between-session (i.e., test-retest) reliability, which was notably lower (percentiles: 50th: 0.270, 75th: 0.355, 95th: 0.507). Of interest, between-condition reliability tended to be lowest in the visual and somatosensory networks – each of which would be expected to vary in a systematic way across conditions due to differences in visual stimulation (movie > inscapes > flanker > rest) and motor demands (flanker > all other conditions).

Regarding test-retest reliability, follow-up analyses also looked at connection-wise ICC for each of the stimulus/task conditions separately using a linear mixed model (as implemented in R) (see Figure 7), finding similar ranges of ICC scores across conditions, though with some notable differences (e.g., higher ICC for visual network in movies and inscapes; higher frontoparietal ICC’s in flanker task and rest). Additionally, we used image-wise correlation coefficient (I2C2) [61] to look at functional networks and their interactions from a multivariate perspective. As can be seen in Figure 7, a high degree of correspondence was noted between the strength of the reliability for a given network (i.e., I2C2) and the strengths of the reliabilities for the individual edges in the network (i.e., ICC).

Finally, to gain insights into the effects of scan duration on test-retest reliabilities, we repeated ICC and I2C2 analyses using 10, 20 and 30 minutes of scan data across 4 pseudo-sessions (i.e., for 20 minutes, we combined data from 2 sessions; for 30 minutes, we combined data from 3 sessions). Consistent with prior reports, our analyses revealed notable improvement of ICC and I2C2 values with longer scans, particularly when increasing from 20 to 30 minutes (see Figures 8, 9).

**Concluding Remarks.** These illustrative analyses highlight the value of these data for addressing questions regarding between-condition and between-session reliability. Beyond quantifying reliabilities for connectomic indices, the data available can also be used by investigators to answer questions regarding minimum data requirements (e.g., number of timepoints) and optimal image processing strategies. Finally, it is worth noting that the availability of naturalistic viewing states (Inscapes, movie clips) in the resource will give...
resting state fMRI-focused investigators an opportunity to explore the added value of these states for calculating intrinsic functional connectivity and more (e.g., exploration of inter-subject correlation and inter-subject functional connectivity [23], [62]).

AVAILABLE OF SUPPORTING DATA
The HBN-SSI is available at: http://fcon_1000.projects.nitrc.org/indi/hbn_ssi/. The Configurable Pipeline for the Analysis of Connectomes, which was employed to carry out the image processing for the analyses include in the text can be found at https://fcp-indi.github.io; the configuration file containing the settings for C-PAC can be found at https://www.nitrc.org/frs/downloadlink.php/9275.

LIST OF ABBREVIATIONS
HBN – Healthy Brain Network
SSI – Serial Scanning Initiative
R-fMRI – Resting State Functional Magnetic Resonance Imaging
DKI – Diffusion Kurtosis Imaging
MPRAGE – Magnetization Prepared Rapidly Acquired Gradient Echo
MNI – Montreal Neurological Institute
FSL – FMRIB Software Library
AFNI – Analysis of Functional NeuroImages
ANTS – Advanced Normalization Tools
iFC – Intrinsic Functional Connectivity
ICN – Intrinsic Connectivity Network
CPAC – Configurable Pipeline for Analysis of Connectomes
QAP – Quality Assurance Protocol
SNR - Signal-to-Noise Ratio
CNR - Contrast-to-Noise Ratio
FBER - Foreground-to-Background Energy Ratio
QI1 - Percent artifact voxels
FWHM – Full Width Half Maximum
EFC - Entropy focus criterion
GSR - Ghost-to-Signal Ratio
Mean FD - Mean Frame-wise Displacement
GCOR - Global correlation
ICC – Intra-class Correlation Coefficient
I2C2 – Image Intra-class Correlation Coefficient

ETHICS APPROVAL AND CONSENT TO PARTICIPATE
All experimental procedures were performed with approval of the Chesapeake Institutional Review Board and only after informed consent were obtained.

CONSENT FOR PUBLICATION
All participants consented to have their data shared.

COMPETING INTERESTS
The authors declare that they have no competing interests.

FUNDING
This work was supported by The Healthy Brain Network (http://www.healthybrainnetwork.org) and its supporting initiatives are supported by philanthropic contributions from the following individuals, foundations and organizations: Lee

AUTHOR CONTRIBUTIONS
Conception and Experimental Design:
JE, LP, MPM, RCC, SC, SG, TV

Implementation and Logistics:
DOC, NVP, RCC, SC, TV

Data Collection:
AB, MK, NGV, YO

Data Informatics:
DOC, JP, RCC

Data Analysis:
DOC, LA, MPM, TX

Drafting of the Manuscript:
DOC, MPM, RCC, TX

Critical Review and Editing of the Manuscript:
All Authors contributed equally to the critical review and editing of the manuscript.

ACKNOWLEDGEMENTS

AUTHOR DETAILS
2. Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York
3. Yale University, New Haven, Connecticut
4. City College of New York, New York, New York
5. The Graduate Center of the City University of New York, New York, New York
6. Massachusetts Institute of Technology, Cambridge, Massachusetts
Bibliography


2009.


Figure Legend

Figure 1 – Shown here are sample stimuli from each of the four scan conditions included in the present work. These included: 1) Resting State, (far left), 2) Inscapes (middle left), 3) Movie Clips (e.g., the Matrix; middle right), and 4) Flanker Task (with no-go trials).

Figure 2 - Subset of Quality Assessment Protocol (QAP) spatial anatomical measures for each participant (horizontal axis). Depicted are the following measures: Contrast-to-Noise Ratio (CNR), Signal-to-Noise Ratio (SNR), Entropy focus criterion (EFC). Foreground-to-Background Energy Ratio (FBER), Spatial smoothness (FWHM), Percent artifact voxels (Q11). Each point indicates the measure calculated for an individual scan; for each participant, the data across scan conditions and sessions are depicted using a single color.

Figure 3 - Subset of Quality Assessment Protocol (QAP) spatial functional measures for each participant (horizontal axis). Depicted are the following measures: Ghost to Signal Ratio (GSR), Signal-to-Noise Ratio (SNR), Entropy focus criterion (EFC). Foreground-to-Background Energy Ratio (FBER), spatial smoothness (FWHM). Each point indicates the measure calculated for an individual scan; for each participant, the data across scan conditions and sessions are depicted using a single color.

Figure 4 - Subset of Quality Assessment Protocol (QAP) temporal functional measures for each participant (horizontal axis). Depicted are the following measures: Outliers Detection (Outliers), Global correlation (GCOR), Quality, Mean Frame-wise Displacement, and Standardized DVARS (DVARS). Each point indicates the measure calculated for an individual scan; for each participant, the data across scan conditions and sessions are depicted using a single color.

Figure 5 - Similarity of full-brain connectivity matrices across participants (green), sessions (blue) and scan conditions (yellow), as measured using Pearson correlation coefficients (red). Also depicted in the bottom right are the distributions of correlation coefficients when comparing scans from the same subject (Within Subject), and scans from different subjects (Between Subject). The distribution of correlation values is also shown (bottom right). On the right column are the values for scans from the same subject, and on the left are scans from different subjects. The median, first, and third quartiles are also depicted with horizontal lines.

Figure 6 - Intraclass correlation coefficients (ICC) quantifying between-condition reliabilities (left) and between-session reliabilities at the connection-level. ICC values were obtained using a hierarchical linear mixed model. These connection-level values are grouped on the vertical and horizontal axes based membership of Intrinsic Connectivity Networks (ICN). No overlap indicates that the voxel did not spatially overlap with any ICN.

Figure 7 - Connection-wise ICC values across all subjects, sessions, and scan conditions (top), as well as network-wise calculations of test-retest reliability carried out using the imagewise intraclass correlation coefficient (I2C2), again across all subjects, sessions and scan conditions (bottom).

Figure 8 – Impact of scan duration on test-retest reliability at the connection level. We randomly sampled sessions, and concatenated the time series temporally to create pseudosessions of 10, 20 and 30 minutes of data. For each of the pseudosession durations, we depict intraclass correlation coefficients (ICC) obtained for each scan condition. Note: across durations, the number of pseudosessions was held constant at four.

Figure 9 – Impact of scan duration on test-retest reliability at the network level. We randomly sampled sessions to create pseudosessions of 10, 20 and 30 minutes of data. For each of the pseudosession durations, we depict imagewise intraclass correlation coefficients (I2C2) obtained for each scan condition. Note: across durations, the number of pseudosessions was held constant at four.
Table Legend
Table 1 – HBN-SSI experimental design.
Table 2 – MRI acquisition parameters for scans included in the HBN-SSI.
Table 3 – Questionnaires and physical measures collected.
Table 4 – Test-Retest reliability of quality assurance protocol (QAP) measures, for each scan condition
Inter-Condition | Inter-Session

Intrinsic Connectivity Networks:
- ICN 1 - Visual
- ICN 2 - Somatomotor
- ICN 3 - Dorsal Attention
- ICN 4 - Ventral Attention
- ICN 5 - Limbic
- ICN 6 - FrontoParietal
- ICN 7 - Default
- ICN 8 - No Overlap

Color Scale:
- 0
- 0.2
- 0.4
- 0.6
- 0.8
- 1
Intrinsic Connectivity Networks

- ICN 1 - Visual
- ICN 2 - Somatomotor
- ICN 3 - Dorsal Attention
- ICN 4 - Ventral Attention
- ICN 5 - Limbic
- ICN 6 - FrontoParietal
- ICN 7 - Default
- ICN 8 - No Overlap
<table>
<thead>
<tr>
<th>Session #</th>
<th>Session Type</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1        | Baseline Characterization | • Multiecho MPRAGE  
• Diffusion Kurtosis Imaging  
• Quantitative T1/T2 Mapping  
• Myelin Transfer Ratio  
• FLAIR  
• fMRI: rest (10 min) |
| 2-7, 9-14| Repeat Scanning       | • Multiecho MPRAGE  
• Diffusion Kurtosis Imaging  
• fMRI: rest (10 min)  
• fMRI: Naturalistic Viewing: Inscapes (10 min)  
• fMRI: Naturalistic Viewing: Movie Clips (10 min)  
• fMRI: Flanker Task (10 min) |
<p>| 8        | Full Feature Movie    | • fMRI: Raiders of the Lost Arc (20 min X 6) |</p>
<table>
<thead>
<tr>
<th>Sequence</th>
<th>3D Despot 1</th>
<th>IR-SPGR / 3D Despot 1</th>
<th>3D Despot 2</th>
<th>3D FLASH</th>
<th>ME-MPRAGE / 3D TFL</th>
<th>FLAIR</th>
<th>EPI</th>
<th>EPI</th>
<th>EPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flip Angle</td>
<td>16</td>
<td>5</td>
<td>60</td>
<td>15</td>
<td>7</td>
<td>150</td>
<td>90</td>
<td>90</td>
<td>55</td>
</tr>
<tr>
<td>Inversion Time</td>
<td>NA</td>
<td>400</td>
<td>NA</td>
<td>NA</td>
<td>1000</td>
<td>2500</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Echo Time</td>
<td>2.4</td>
<td>2.4</td>
<td>2.7</td>
<td>11</td>
<td>1.64</td>
<td>89</td>
<td>76.2</td>
<td>93.8</td>
<td>40</td>
</tr>
<tr>
<td>Repetition Time</td>
<td>5.2</td>
<td>5.3</td>
<td>5.4</td>
<td>30</td>
<td>2730</td>
<td>9000</td>
<td>3110</td>
<td>4500</td>
<td>1450</td>
</tr>
<tr>
<td>Bandwidth per Voxel (Readout)</td>
<td>350</td>
<td>350</td>
<td>350</td>
<td>350</td>
<td>651</td>
<td>190</td>
<td>1628</td>
<td>1628</td>
<td>2374</td>
</tr>
<tr>
<td>Parallel Acquisition</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>GP2</td>
<td>GP2</td>
<td>GP2</td>
<td>MB3</td>
<td>MB3</td>
<td>MB3</td>
</tr>
<tr>
<td>Slice Orientation</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Slice Phase Encoding Direction</td>
<td>AP</td>
<td>AP</td>
<td>AP</td>
<td>AP</td>
<td>AP</td>
<td>RL</td>
<td>AP/PA</td>
<td>AP</td>
<td>AP</td>
</tr>
<tr>
<td>Slice Acquisition Order</td>
<td>SA</td>
<td>SA</td>
<td>SA</td>
<td>IA</td>
<td>IA</td>
<td>IA</td>
<td>IA</td>
<td>IA</td>
<td>IA</td>
</tr>
<tr>
<td>Slice Gap</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>50</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
</tr>
<tr>
<td>Field of View</td>
<td>220x220</td>
<td>220x220</td>
<td>220x220</td>
<td>256x256</td>
<td>256x256</td>
<td>201x230</td>
<td>192x192</td>
<td>192x192</td>
<td>192x192</td>
</tr>
<tr>
<td>Reconstructed Image Matrix</td>
<td>128x128x96</td>
<td>128x128x48</td>
<td>128x128x96</td>
<td>256x256x176</td>
<td>256x256x176</td>
<td>448x512x25</td>
<td>96x96x72</td>
<td>96x96x72</td>
<td>78x78x54</td>
</tr>
<tr>
<td>Reconstructed Resolution</td>
<td>1.72x1.72x1.8</td>
<td>1.72x1.72x3.6</td>
<td>1.72x1.72x1.8</td>
<td>1.0x1.0</td>
<td>1.0x1.0</td>
<td>1.0x1.0</td>
<td>0.45x0.45x6.5</td>
<td>2.0x2.0x2.0</td>
<td>2.0x2.0x2.0</td>
</tr>
<tr>
<td>Number of Measurements</td>
<td>8</td>
<td>1</td>
<td>16</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>420</td>
</tr>
<tr>
<td>Acquisition Time</td>
<td>5:00</td>
<td>0:53</td>
<td>8:38</td>
<td>6:41</td>
<td>6:32</td>
<td>2:44</td>
<td>0:16</td>
<td>9:59</td>
<td>10:18</td>
</tr>
<tr>
<td>Fat Suppression</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>No</td>
<td>On</td>
<td>On</td>
<td>On</td>
</tr>
<tr>
<td>Number of Directions</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>64</td>
<td>64</td>
<td>NA</td>
</tr>
<tr>
<td>Number of B Zeros</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>B Value (s)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>0;1000;2000</td>
<td>NA</td>
</tr>
<tr>
<td>Averages</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Questionnaires</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal State Questionnaire (pre-scan, post-scan)</td>
<td>3-item self-report questionnaire assessing hunger and thirst. Participants respond on a visual analogue scale ranging from &quot;I am not hungry/thirsty/full at all&quot; to &quot;I have never been more hungry/thirsty/full&quot;. Responses are rated from 0-100. Participants complete this questionnaire before and after each scan.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York Cognition Questionnaire (NYC-Q) (post-scan)</td>
<td>31-item self-report questionnaire that asks participants about the different thoughts and feelings that they may have had while in the MRI scan. Participants are asked to indicate the extent to which their thinking or experience corresponded to each item on a 9-point scale.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PANAS (post-scan)</td>
<td>The PANAS-S is a self-administered, 20-item Likert scale assessment that measures degree of positive or negative affect. Users are asked to rate 10 adjectives that measure positive feelings such as joy or pleasure, and 10 adjectives that measure negative feelings, such as anxiety or sadness, on a scale of how closely the adjective describes them in the present moment or over the past week. Items are rated on a five-point scale.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical Measures</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitals</td>
<td>Participant vitals (blood pressure, heart rate, blood glucose level, first day of last menstrual cycle) were collected prior to each scan using standard measurement devices in a laboratory environment.</td>
</tr>
<tr>
<td>Voice data samples</td>
<td>Audio samples of participant speech were recorded prior to scanning. Each sample consisted of 10 sentences with 5 different implicit emotions (neutral, happy, sad, angry, fearful), 10 non-words, and 2 minutes of free speech. For each sample different sentences were drawn from the same set of emotions; the non-words also differed in each sample but had similar characteristics (i.e., number of syllables, chunks). Stimuli were presented on a laptop computer screen. Completion of the sample took up to 15 minutes.</td>
</tr>
<tr>
<td>Quotient ADHD System</td>
<td>Quotient is a computer-based task designed to assess three core symptoms of ADHD: hyperactivity, attention, and impulsivity. Participants respond to stimuli presented with random timing and random placement on a screen. Completion of the task takes up to 30 minutes.</td>
</tr>
<tr>
<td>GeneActiv Actimetry Device</td>
<td>Between scanning sessions, participants wore a non-invasive actimetry sensor that recorded heart rate and indices of physical activity and sleep. The device was placed on participants’ non-dominant wrist and data was collected at each scanning session.</td>
</tr>
<tr>
<td>Measure</td>
<td>Rest</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>EFC</td>
<td>0.90</td>
</tr>
<tr>
<td>FBER</td>
<td>0.84</td>
</tr>
<tr>
<td>FWHM</td>
<td>0.58</td>
</tr>
<tr>
<td>GSR</td>
<td>0.56</td>
</tr>
<tr>
<td>SNR</td>
<td>0.92</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.08</td>
</tr>
<tr>
<td>GCOR</td>
<td>0.11</td>
</tr>
<tr>
<td>Quality</td>
<td>0.94</td>
</tr>
<tr>
<td>Mean FD</td>
<td>0.30</td>
</tr>
<tr>
<td>DVARS</td>
<td>0.42</td>
</tr>
</tbody>
</table>