NeuroImage 255 (2022) 119215

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

An open-access accelerated adult equivalent of the ABCD Study neuroimaging dataset (a-ABCD)

Kristina M. Rapuano[a,*], May I. Conley[a], Anthony C. Juliano[b], Gregory M. Conan[g], Maria T. Maza[a,c], Kylie Woodman[a,d], Steven A. Martinez[a,e], Eric Earl[f], Anders Perrone[f,g], Eric Feczko[g,h], Damien A. Fair[g], Richard Watts[a], B.J. Casey[a,*], Monica D. Rosenberg[a,i,*]

[a] Department of Psychology, Yale University
[b] Department of Psychiatry, University of Vermont
[c] Department of Psychology, University of North Carolina, Chapel Hill
[d] Department of Communication, University of California, Santa Barbara
[e] Department of Psychology, Temple University
[f] Department of Psychiatry, Oregon Health and Science University
[g] Masonic Institute for the Developing Brain, University of Minnesota Medical School
[h] Department of Pediatrics, University of Minnesota Medical School
[i] Department of Psychology, University of Chicago, United States

ARTICLE INFO

Keywords:
Development
Working memory
Inhibitory control
Reward processing
fMRI

ABSTRACT

As public access to longitudinal developmental datasets like the Adolescent Brain Cognitive Development Study℠ (ABCD Study®) increases, so too does the need for resources to benchmark time-dependent effects. Scan-to-scan changes observed with repeated imaging may reflect development but may also reflect practice effects, day-to-day variability in psychological states, and/or measurement noise. Resources that allow disentangling these time-dependent effects will be useful in quantifying actual developmental change. We present an accelerated adult equivalent of the ABCD Study dataset (a-ABCD) using an identical imaging protocol to acquire magnetic resonance imaging (MRI) structural, diffusion-weighted, resting-state and task-based data from eight adults scanned five times over five weeks. We report on the task-based imaging data ($n = 7$). In-scanner stop-signal (SST), monetary incentive delay (MID), and emotional *n*-back (EN-back) task behavioral performance did not change across sessions. Post-scan recognition memory for emotional *n*-back stimuli, however, did improve as participants became more familiar with the stimuli. Functional MRI analyses revealed that patterns of task-based activation reflecting inhibitory control in the SST, reward success in the MID task, and working memory in the EN-back task were more similar within individuals across repeated scan sessions than between individuals. Within-subject, activity was more consistent across sessions during the EN-back task than in the SST and MID task, demonstrating differences in fMRI data reliability as a function of task. The a-ABCD dataset provides a unique testbed for characterizing the reliability of brain function, structure, and behavior across imaging modalities in adulthood and benchmarking neurodevelopmental change observed in the open-access ABCD Study.

1. Introduction

The increasing availability of open-access datasets with developmental neuroimaging data provides unprecedented power for characterizing relationships between the developing brain and behavior. To date these datasets include structural and functional magnetic resonance (fMRI), behavioral, and demographic data from more than 20,000 youth around the world (Rosenberg et al., 2018; Simmons et al., 2021) and facilitate reproducibility and replication in developmental cognitive neuroscience.

Some of these studies include longitudinal neuroimaging data collected during cognitive tasks, allowing researchers to characterize how behavioral performance and corresponding fMRI activity change—or remain stable—across development. Although longitudinal designs offer clear benefits, one potential downside is that what appears to be developmental change may in part be driven by practice effects that accompany repeated sampling. For example, participants could hypothetically show improved behavioral performance and distinct patterns of brain activity during a cognitive task when they are fourteen years

* Corresponding authors.
E-mail addresses: kristina.rapuano@yale.edu (K.M. Rapuano), bj.casey@yale.edu (B.J. Casey), mdrosenberg@uchicago.edu (M.D. Rosenberg).

https://doi.org/10.1016/j.neuroimage.2022.119215.
Received 18 December 2021; Received in revised form 14 March 2022; Accepted 13 April 2022
Available online 16 April 2022.

old compared to when they were eight. These differences could reflect neurodevelopment with the transition into adolescence or greater expertise with the task after repeated testing (i.e., practice effects). They could also simply reflect noise in unreliable behavioral and/or brain measures (unless they are statistically reliable across a group of individuals). Put another way, scan-to-scan differences in fMRI task performance and patterns of activation in longitudinal studies may reflect trait-like variability (e.g., developmental change), state-like variability (e.g., differences in task familiarity and cognitive, attentional, and emotional states during scan sessions), an interaction between the two (e.g., changing states or state frequencies across development), and/or noise. Measuring the reliability and meaningful variation of behavior and brain function across repeated study sessions helps disentangle these interactions and interpret longitudinal changes. A growing body of work assesses the reliability of task-based patterns of fMRI activity within-subjects and across sites in youth (Casey et al., 1998; Haller et al., 2018; Kennedy et al., 2021) and adults (Bennett & Miller, 2013; Berman et al., 2010; Buimer et al., 2020; Elliott et al., 2020; Friedman et al., 2008; Gee et al., 2015; Kragel et al., 2021; Li et al., 2020; McGonigle et al., 2000; Raemaekers et al., 2012; for reviews see Herting et al., 2018; Noble et al., 2021). An open question remains, however, about how to "benchmark" session-to-session differences observed in longitudinal developmental neuroimaging datasets—that is, how to disentangle developmental effects from practice effects from repeated testing and other state-related effects and noise. Answering such a question requires a highly sampled within-subject dataset calibrated to a large cohort dataset.

Here we contribute to work on this question with data from a new accelerated adult cohort of a large-scale developmental neuroimaging study: the Adolescent Brain Cognitive Development Study[SM] (ABCD Study®). The ABCD Study (Casey et al., 2018; Luciana et al., 2018; Volkow et al., 2018) is a longitudinal assessment of brain development in a sample of 11,875 9- to 10-year-olds at 21 sites across the United States. ABCD Study participants complete the same 90-minute MRI battery—which includes structural scans, diffusion weighted imaging scans, four 5-minute resting-state runs, and two runs each of the emotional *n*-back (EN-back) task, stop-signal task (SST), and monetary incentive delay (MID) task—every two years for ten years, from 2016-17 through 2026-27. In addition, participants complete iPad questionnaires and tasks during yearly lab visits and intervening phone interviews every six months. ABCD Study data are made openly available to the research community at The National Institute of Mental Health Data Archive (https://nda.nih.gov/abcd) as they are collected.

The current study uses an accelerated adult equivalent of the ABCD (a-ABCD) imaging protocol in eight young adults scanned five times to parallel the five longitudinal waves of ABCD Study imaging data that will be collected over 10 years. Using the task-based data from our accelerated adult cohort, we first characterize the stability of adults' fMRI task performance from one scan session to the next, asking whether performance improves from week to week due to practice effects and/or growing familiarity with the scanning protocol. We next characterize the consistency of fMRI activation maps across scan sessions, asking whether contrast maps reflecting inhibitory control in the SST, reward success in the MID task, and working memory in the EN-back task are similar within individuals over time. These analyses inform the reliability of behavior and brain function in ABCD Study neuroimaging tasks in adulthood. More broadly, the a-ABCD dataset itself may be a useful resource for researchers working with the ABCD Study sample or other longitudinal developmental datasets, providing opportunities to compare adult and developmental datasets with identical scanning protocols (including task and rest fMRI, structural MRI, and diffusion tensor imaging) and disentangle scan-to-scan effects observed on the order of weeks from developmental effects observed on the order of years.

2. Methods

2.1. Participants

2.1.1. Accelerated adult ABCD (a-ABCD) sample

Eight right-handed adults (4 female, mean age = 23 years, SD = 1.3 years, range = 21–25 years, 62.5% White/non-Hispanic; 25% Asian; 12.5% Black/Hispanic) with no history of neurological injury or illness participated[1]. Participants had previously participated in MRI studies or has gained familiarity with MRI through school or work. For six individuals, scan sessions were held 7 days apart. Two participants' fourth and fifth sessions were separated by 8 and 26 days, respectively. Data from the latter participant were excluded from the current analysis due to errors in fMRI task administration (e.g., task runs acquired out of order; task runs launched separately instead of as two runs embedded within one E-Prime execution disrupting task staircasing), resulting in a final sample of seven individuals (3 female, mean age = 23). Data from one participant's second-session monetary incentive delay task were excluded from analysis due to errors in task administration.

All scan sessions began between 11 AM and 2 PM (mean start time = 12:21 PM, SD = 1.25 hours), and all five of each participant's scans started within an hour of each other to minimize time-of-day effects[2]. Participants provided written informed consent in compliance with procedures approved by the Yale University Human Subjects Committee and were paid for their participation.

2.1.2. Comparison youth ABCD Study sample

Behavioral data from year one (baseline) of the ABCD Study were obtained from curated data release 2.0.1 (DOI 10.15154/1504041) as a qualitative comparison sample (n = 11,537; age range = 9-10 years; 48.4% female; exclusion criteria described in Rosenberg et al., 2020). Measures included SST, MID, emotional *n*-back, and post-scan recognition memory performance.

2.2. Study design

Participants completed five ABCD Study-style imaging sessions over the course of approximately five weeks to parallel the ABCD Study protocol that includes five imaging sessions over the course of ten years, with scans every two years from age 9–10 to age 19–20. Replicating the ABCD Study protocol, each testing session began with a pre-scan practice of three neuroimaging tasks: the stop-signal, monetary incentive delay, and emotional *n*-back tasks. Participants then completed a pre-scanning questionnaire indicating how much they were experiencing ten different feelings and emotions (e.g., relaxed, happy, awake, sad). Next,

[1] We included 8 individuals (5 scans per individual = 40 scans total) based on scan time and funding constraints. This sample size is in line with other small *n* "deep phenotyping" studies, including the MyConnectome project (n = 1, 104 sessions; Poldrack et al., 2015) and the Midnight Scan Club dataset (n = 10, 12 sessions each; Gordon et al., 2017), although these datasets include more data per individual. Given this small n, we focus on individual-level rather than group-level effects in all fMRI analyses. For behavioral analyses, we include Bayesian statistics to quantify evidence in favor of the null vs. alternative hypothesis for all group-level tests. Finally, we visualize individual participant data to give readers the opportunity to assess the consistency of behavioral and fMRI results across participants.

[2] We were motivated to keep time-of-scan constant based on work suggesting that sustained attention (Riley et al., 2017) and functional connectivity patterns (Fafrowicz et al., 2019; Orban et al., 2020) vary with time of day in adults. Given the sheer number of scans acquired for the ABCD Study and variability in access to scanners across sites, controlling for time-of-day in the acquisition of the ABCD Study imaging data is not feasible. However, having the time of day logged for ABCD scans provides the scientific community with the opportunity to use these data in their analyses.

during MRI data collection, participants completed structural, diffusion-weighted, resting-state, and three task-based imaging scans using a protocol described in previous work (Casey et al 2018) and detailed at https://abcdstudy.org/scientists/protocols. After the scans were completed, participants performed a recognition memory test for emotional *n*-back task stimuli, completed a post-scanning questionnaire rating how much they were experiencing the same ten feelings and emotions, and completed a post-MID questionnaire. Participants also completed the Wechsler Abbreviated Scale of Intelligence (WASI-II) (Wechsler, 2011) once during the course of the study; these data are not analyzed here. They did not complete other ABCD Study behavioral tasks, interviews, or questionnaires.

2.3. Tasks

During each study session, participants completed three in-scanner tasks: *(1)* the stop-signal task (SST), designed to measure impulsivity and impulse control; *(2)* the monetary incentive delay (MID) task, designed to measure aspects of reward processing; and *(3)* the emotional *n*-back (EN-back) task, designed to measure processes related to working memory and emotion regulation (Casey et al., 2018). After scanning, participants performed a recognition memory task assessing memory for EN-back task stimuli. Task details are described below.

2.3.1. Stop-signal

During each fMRI session, participants performed two approximately 6-min runs of the SST (Casey et al., 2018; Logan, 1994) (437 volumes after discarded acquisitions). Data were collected using the original ABCD Study SST E-Prime scripts before updates addressing issues raised in a commentary on the ABCD Study stop-signal task design (Bissett et al., 2021) were made. Trials began with a left- or right-facing arrow (the go signal). Participants were instructed to press a button indicating the direction of the arrow as quickly and accurately as possible, except when an arrow pointing up (the stop signal) appeared (16.67% of trials). The time between go and stop signal onset—the stop-signal delay—was adjusted in a stepwise manner (i.e., staircased) so that each participant correctly withheld a response to approximately half of stop trials. Specifically, the time between the onset of the go signal and the onset of the stop signal (the stop-signal delay; SSD) was initially set to 50 ms (Bissett et al., 2021; Casey et al., 2018). When a participant failed to stop, task difficulty was decreased by shortening the SSD by 50 ms (to a minimum of 0 ms). When a participant successfully stopped, task difficulty was increased by lengthening the SSD by 50 ms (to a maximum of 900 ms). Here performance is measured with the stop-signal reaction time (SSRT; mean SSD subtracted from mean correct-trial RT).

2.3.2. Monetary incentive delay

During each fMRI session, participants performed two approximately 5.5-min MID task runs (403 volumes after discarded acquisitions) (Casey et al., 2018; Knutson et al., 2000; Yau et al., 2012). On each trial, participants saw a cue indicating whether they could win or lose $0.20 or $5, or whether no money was at stake ($0). After a variable delay of 1500–4000 ms, a target appeared for 150–500 ms. Participants were instructed to respond via button press while the target was on the screen to win or lose the indicated amount. Feedback was provided after each trial. Target timing was staircased such that each participant achieved and maintained approximately 60% accuracy. Initial target duration was based on each participant's RT during the pre-scan practice. Task difficulty was adjusted after every third incentivized trial based on the participant's accuracy on the previous six trials. If accuracy was below 60%, target duration was lengthened and if accuracy was above 60%, target duration was shortened. Here performance is measured as practice trial RT and percent positive feedback (i.e., percent accuracy) on reward, loss, and neutral trials.

2.3.3. Emotional n-back

In each scan session, participants completed two approximately 5-min EN-back task runs (362 volumes per run after discarded acquisitions) (Barch et al., 2013; Casey et al., 2018). Runs included four 0-back blocks (low memory load) and four 2-back blocks (high memory load) with stimuli from one of four categories: happy, fearful, and neutral faces and places. During 0-back blocks, participants were instructed to press a button indicating "match" when they saw an image matching a target picture, and to press a button indicating "no match" otherwise. During 2-back blocks, participants were instructed to press "match" when they saw an image matching the picture from two trials back, and to press "no match" otherwise. Performance is measured by 0-back and 2-back task percent accuracy.

2.3.4. Recognition memory

After each imaging session, participants' memory for stimuli seen during the emotional *n*-back task was probed with an out-of-scanner recognition memory test (Barch et al., 2013; Casey et al., 2018). Participants were asked to rate each of 48 *n*-back and 48 novel stimuli as "old" or "new." Memory performance is assessed with sensitivity (*d'*).

2.4. Neuroimaging data collection

Neuroimaging procedures replicated those of the ABCD Study. Scans were acquired on a Siemens Prisma 3T scanner at the Yale University Faculty of Arts and Sciences Brain Imaging Center (one of the 21 ABCD Study sites). Each scan session included a localizer, high-resolution 3D T1-weighted anatomical scan, two 5-min resting-state fMRI runs, diffusion weighted images, 3D T2-weighted spin echo images, two 5-min resting-state fMRI runs, and six task-based fMRI runs. Mirroring the ABCD Study design, task order and version (which controlled the order in which stimuli were presented) was randomized across participants but held constant within participants (Casey et al., 2018). Functional images were collected using a multiband gradient EPI sequence with the following parameters: TR = 800 ms, TE = 30 ms, flip angle = 52°, 60 slices acquired in the axial plane, voxel size = $2.4 \times 2.4 \times 2.4$ mm^3, multiband slice acceleration factor = 6. T1 images were collected with the following parameters: TR = 2500 ms, TE = 2.88 ms, flip angle = 8°, 176 slices, voxel size = 1 mm^3, parallel imaging factor = 2. T2 images were collected with the following parameters: TR = 3200 ms, TE = 565 ms, variable flip angle, 176 slices, voxel size = 1 mm^3, parallel imaging factor = 2x. Diffusion images were collected with the following parameters: TR = 4100 ms, TE = 88 ms, flip angle = 90°, 81 slices, voxel size = $1.7 \times 1.7 \times 1.7$ mm^3. Detailed imaging acquisition parameters are reported elsewhere (Casey et al., 2018; Hagler et al., 2019).

2.5. Behavioral analyses

Changes in task performance over time were assessed with linear mixed effects models using the *lme4* package in R (Bates et al., 2015). Models included session as a fixed effect and participant as a random intercept. Random slopes also were included if doing so significantly improved model fit. Optimization was performed with the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (Byrd et al., 1995) with the optimx package (Nash & Varadhan, 2011). Significance was assessed with Type III Satterthwaite approximations using the *lmerTest* package (Kuznetsova et al., 2017). Bayes factors, computed with the *bayestestR* package (Makowski et al., 2019), are reported as measures of relative evidence for models including session as a fixed effect and random intercepts for participants (and random slopes when they significantly improved model fit) (H_A) relative to models only including random intercepts for participants (H_0). Bayes factors greater than one are considered evidence in favor of the alternative hypothesis (H_A, the model including session as a predictor). Bayes factors 1–3 are considered anecdotal evidence for H_A, 3–10 substantial evidence, 10–30 strong evidence, 30–

100 very strong evidence, and >100 decisive evidence (Wetzels et al., 2011). Bayes factors less than one are considered evidence in favor of the null hypothesis (H_0, the model excluding session). Bayes factors 1/3–1 are considered anecdotal evidence for H_0, 1/10–1/3 substantial evidence, 1/30–1/10 strong evidence, 1/100–1/30 very strong evidence, and <1/100 decisive evidence (Wetzels et al., 2011).

2.6. Image preprocessing

Neuroimaging preprocessing steps mirrored those used in the publicly available ABCD-BIDS Community Collection (ABCC; Feczko et al., 2021; https://osf.io/psv5m/). Data were first converted to BIDS format using *dcm2bids* (https://github.com/cbedetti/Dcm2Bids), which reorganizes nifti images produced with *dcm2niix* (Xiangrui Li et al., 2016). Raw images were processed using the ABCD-BIDS preprocessing pipeline (for details, see https://github.com/DCAN-Labs/abcd-hcp-pipeline; https://osf.io/89pyd/; Sturgeon et al., 2021), which is based on the Human Connectome Project's minimal preprocessing pipeline (Glasser et al., 2013). The first stage of this pipeline, PreFreeSurfer, performs brain extraction, alignment, and bias field correction on T1w and T2w images. The second stage, FreeSurfer (Dale et al., 1999; Fischl, 2012), segments the resulting T1w images and identifies tissue boundaries which are then registered to a FreeSurfer template. In the PostFreeSurfer stage, brain masks produced by FreeSurfer are used to register T1w images to MNI space via Advanced Normalization Tools (ANTs) symmetric image normalization method (Avants et al., 2008). Surfaces are transformed to standard space using spherical registration and converted to CIFTI format along with the standardized volumes. The fMRIVolume stage corrects for local field inhomogeneities by performing functional image distortion correction using reverse phase-encoded spin echo images. To avoid potential motion confounds, eta squared values are computed for each pair of field map images relative to a participant-level average of all field maps, and the pair with the highest value (i.e., most representative of the average) was selected. Finally, the fMRISurface stage performs spatial smoothing (2-mm full-width half-max).

This processing pipeline was identical to that used for the ABCC (Feczko et al., 2021) and thus did not include procedures to account for the longitudinal nature of the data. In addition to prioritizing consistency across pipelines, this approach also ensures consistency in the template images used across ABCD Study and a-ABCD scans. However, given potential interest in within-participant changes in morphometry over time, the a-ABCD data repository includes raw imaging files that will allow researchers to optimize longitudinal processing procedures depending on their needs.

2.7. Task activation analyses

Because the ABCD-BIDS pipeline minimally processes the task fMRI data, additional preprocessing steps were completed using the abcd-bids-tfmri-pipeline (Juliano et al., 2021), a modified, Python-based version of the TaskfMRIAnalysis stage of the HCP-pipeline. ABCD-BIDS-tfMRI uses FSL utilities and functions (Jenkinson et al., 2012) to prepare the data for higher-level analyses. High-pass filtering was applied using a 200 sec cutoff. First-level GLMs were run using film_gls (Woolrich et al., 2001). This included nuisance regression (motion parameters including six translations, rotations, and their derivatives), as well as the censoring of timepoints that exceeded a framewise displacement of 0.9 mm. Initial timepoints were also censored to account for the MRI signal stabilization. A double-gamma hemodynamic response function was used. Runs for each task were combined for each session separately using a fixed effects model. The emotional *n*-back task GLM included predictors for fixation and happy, fearful, and neutral faces as well as place stimuli in the 0-back and 2-back conditions. The SST model included predictors for correct and incorrect stop and go trials. The MID model included small and large reward and loss cues and no stakes cues, and the corresponding positive and negative feedback.

Representative contrasts for each task condition were selected for visualization and cross-session comparison (SST: correct stop vs. correct go; MID: reward success vs. failure [i.e., positive vs. negative feedback for small and large rewards]; EN-back: 2-back vs. 0-back). Contrasts were selected to match those reported previously in the ABCD Study sample (Casey et al., 2018; Chaarani et al., 2021). Additional contrasts of interest are reported in supplemental materials. Visualizations were performed with Human Connectome Project Workbench and Python. For each participant, binarized contrasts passing a vertex-wise threshold of $p < 0.0001$ were overlaid according to session to visualize within-participant consistency in task activation. In addition, Pearson correlation was computed for every pair of unthresholded *t*-statistic contrast images, and a linear mixed-effects model (including a random effect for session nested within subject) was used to test for significant differences between sessions collected within the same participant versus sessions collected across participants. Correlation coefficients were Fisher *z*-transformed before being submitted to mixed effects models.

3. Results

3.1. Behavioral results

3.1.1. Overall task performance

Adult task performance during the first scan session fell within the ABCD Study baseline (i.e., year 1) cohort performance range (Fig. 1). Median performance during this first session ranged from the 62nd percentile of the corresponding ABCD Study distribution (for MID reward trial accuracy) to the 99th percentile (for 0-back accuracy). Mean performance across all participants and sessions is reported in Table 1.

3.1.2. Changes in performance across sessions

Consistent with the ABCD Study, in-scanner task order—as well as stimulus order within each task—was held constant within participants across sessions. Thus, a potential obstacle for using this task battery to assess developmental changes in cognitive and attentional processes is that participants may show practice effects as tasks become more familiar and trial orders more predictable. Such practice effects could potentially obscure day-to-day variability in cognitive states or be misinterpreted as developmental change.

Contrary to predictions that performance would improve across sessions due to practice effects, fMRI task performance did not change over time (**Fig. 2**). Rather, Bayes factors (which reflect the marginal likelihood of the null hypothesis that performance does not change across sessions and the alternative hypothesis that it does change) provide substantial or stronger evidence for the null hypothesis for SST SSRT, MID task performance, and emotional *n*-back task performance, and anecdotal evidence for the null for SST SSD (**Table 2**). Thus, the SST and MID staircasing algorithms, designed so that participants would correctly withhold response to approximately 50% of SST stop trials and achieve approximately 60% MID task accuracy, resulted in stable SST and MID performance across sessions. Consistent MID practice RTs, which were measured in pre-scan practice tasks and used to set the initial speed of MID trials, suggest that participants did not strategically alter their response times during practice trials in attempts to decrease MID task difficulty and win more money. Furthermore, 0-back and 2-back accuracy did not systematically vary across sessions.

In contrast to the in-scanner task performance measures, post-scan recognition memory for emotional *n*-back stimuli changed across sessions. Bayes factors provide decisive evidence that hit rate increased (as participants became familiar with the images used during the EN-back task) and very strong evidence that false alarm rate increased (as they became familiar with the *lure* images presented in the recognition memory task). Concurrent increases in hit rate and false alarm rate resulted in consistent *d'* scores over time. Because MRI sessions for ABCD Study

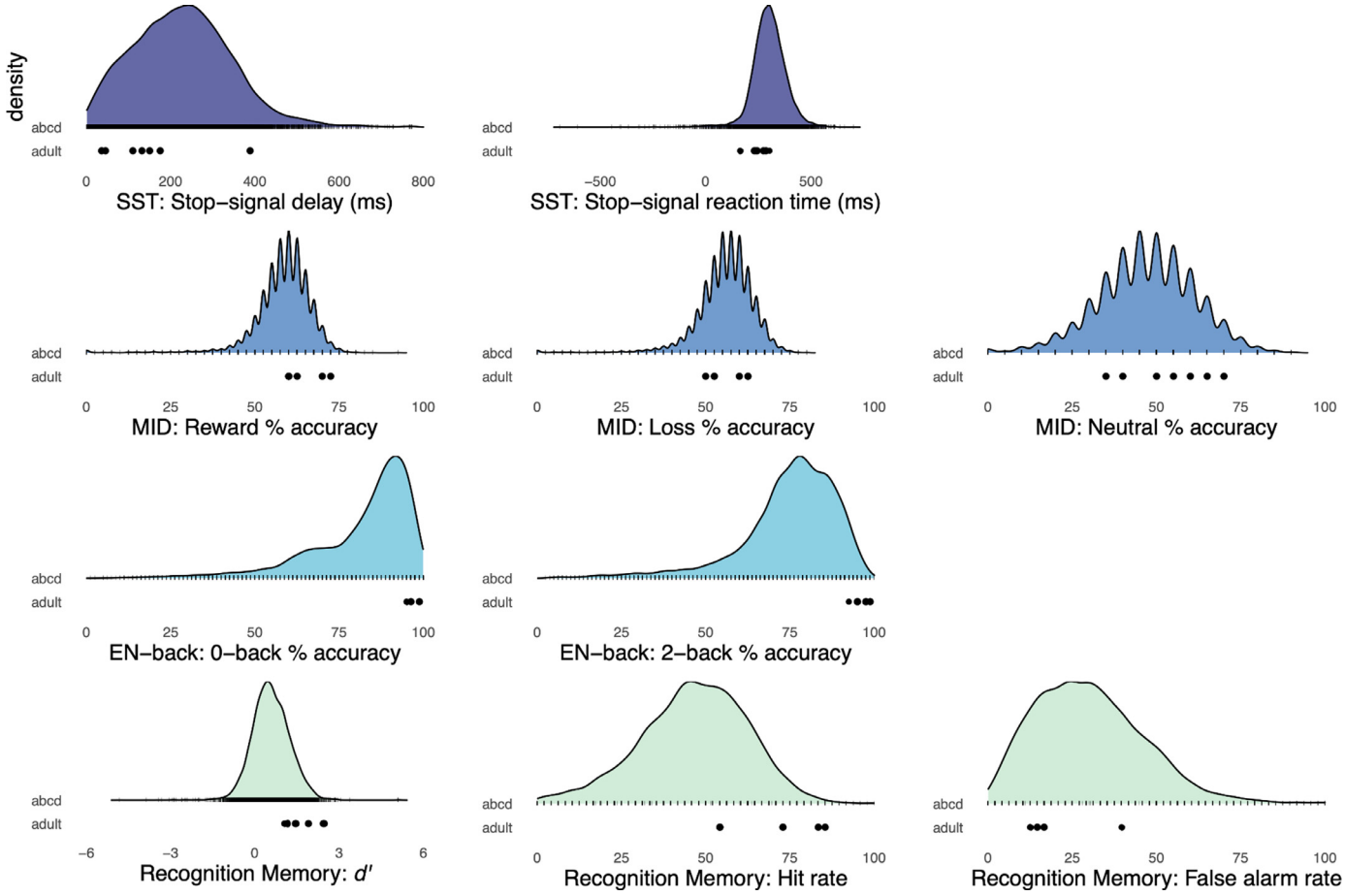

Fig. 1. Adult participant session 1 performance (adult) relative to distributions of the ABCD Study participant baseline (i.e., year 1) performance (abcd). Density plots show performance of 9- to 10-year-olds enrolled in the ABCD Study. Vertical lines below the density plots represent single-child data, and black dots represent single-adult data. Due to missing data, plots include data from the following number of ABCD participants: SST n = 9348, MID n = 9546, EN-back n = 9220, and recognition memory n = 8741.

Table 1
Task performance across adults and scan sessions.

Task	Behavioral measure	min.	max.	median	mean	std. dev.
SST	Stop-signal delay (ms)	25.83	388.33	110.00	110.62	68.77
SST	Stop-signal RT (ms)	162.76	328.57	255.21	250.02	39.07
MID	Practice RT (ms)	210.00	375.00	267.00	270.97	35.89
MID	Reward % acc.	47.50	75.00	60.00	61.25	7.16
MID	Loss % acc.	45.00	70.00	60.00	59.04	6.54
MID	Neutral % acc.	30.00	75.00	55.00	54.71	13.08
EN-back	0-back % acc.	92.50	100	98.75	97.96	2.17
EN-back	2-back % acc.	90.00	100	97.50	96.96	2.73
Recognition memory	d'	1.01	3.30	2.17	2.12	0.66
Recognition memory	Hit rate (%)	39.58	100.00	83.33	78.57	16.33
Recognition memory	False alarm rate (%)	2.08	87.50	12.50	22.68	22.72

participants are separated by two years rather than one week, memory for EN-back stimuli encountered in previous scan sessions may have less pronounced effects on recognition memory, if any.

3.1.3. Self-reported mood and feeling questionnaires

Participants rated how relaxed, happy, scared, awake, upset, angry, excited, tired, sleepy, and sad they felt on a scale from 1 ("very slightly or not at all") to 5 ("extremely") immediately before and after each scan session. We analyzed responses with linear mixed effects models including time (pre-scan vs. post-scan) and session (1–5) as fixed effects and participant as a random effect. There were no significant main effects of time and no significant interactions between time and session for any mood or feeling. There was an effect of session on excitement such that reported excitement decreased across sessions (β = –0.11, SEM = 0.06, $F(1,60)$ = 4.20, p = 0.045); however, this effect did not survive Bonferroni correction for ten comparisons. All other effects of session were nonsignificant (uncorrected $p > 0.05$).

3.2. Imaging results

3.2.1. Activation pattern consistency varies as a function of task

Because the a-ABCD dataset includes approximately 165 minutes of task data per participant (33 minutes/session × 5 sessions) and only seven participants, we focused on individual-level analyses. For each participant and each session, we contrasted fMRI activity associated with correct stop vs. correct go trials during the SST, reward success

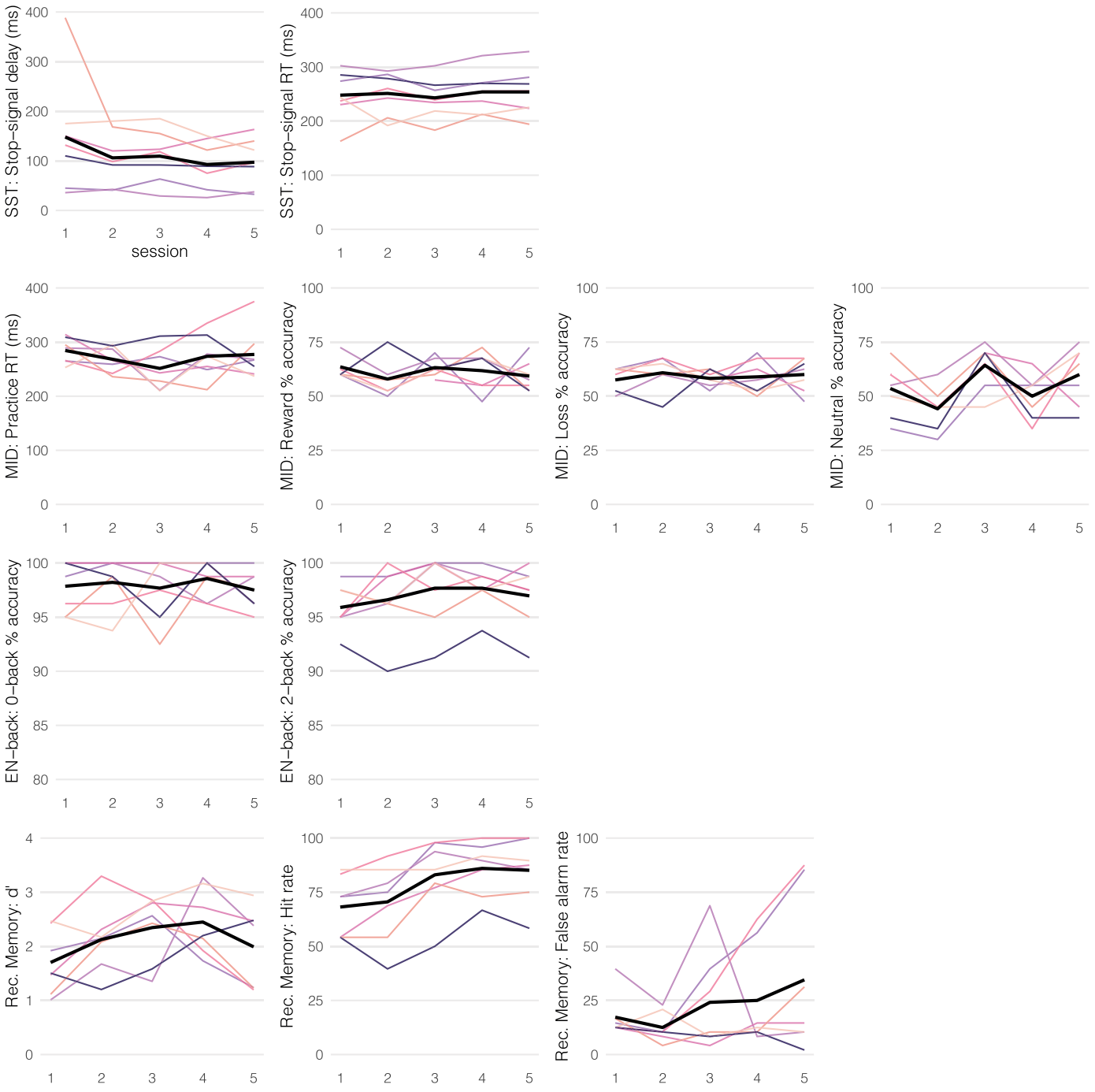

Fig. 2. Performance on the three functional MRI tasks (stop signal, monetary incentive delay, emotional *n*-back) and the post-scan recognition memory task for emotional *n*-back task stimuli. Lines in color represent single-subject data; black lines represent group means.

vs. failure trials during the MID task, and 2-back vs. 0-back blocks in the EN-back task. Representative contrasts were selected to match those reported in previous work (Casey et al., 2018.) For each individual, we visualized the overlap of session-specific contrast maps after applying an uncorrected vertex-wise threshold of $p < 0.0001$ and computed correlations between pairs of unthresholded contrast maps across sessions.

For the SST, the mean correlation between within-subject pairs of unthresholded *t*-statistic contrast images was $r = 0.26$ (subject-wise range [0.19, 0.34]; session-wise range [0.07, .46]; Fig. 3). We observed vertices with significant positive values (correct stop > correct go) in up to five (of five) sessions for each participant, and vertices with significant negative values (correct go > correct stop) in in a maximum of one to three sessions for each participant.

For the MID, the mean correlation between within-subject pairs of unthresholded *t*-statistic contrast images was $r = 0.14$ (subject-wise range [0.09, 0.23]; session-wise range [–0.02, 0.31]; Fig. 4). We observed vertices with significant positive values (reward success > reward fail) in, at most, one to three (of five) sessions, and vertices with significant negative values (reward fail > reward success) in at most two to four sessions for each participant.

For the EN-back task, we observed greater within-subject overlap of activity across sessions than in the SST and MID task. Mean correlation between participants' unthresholded contrast maps was

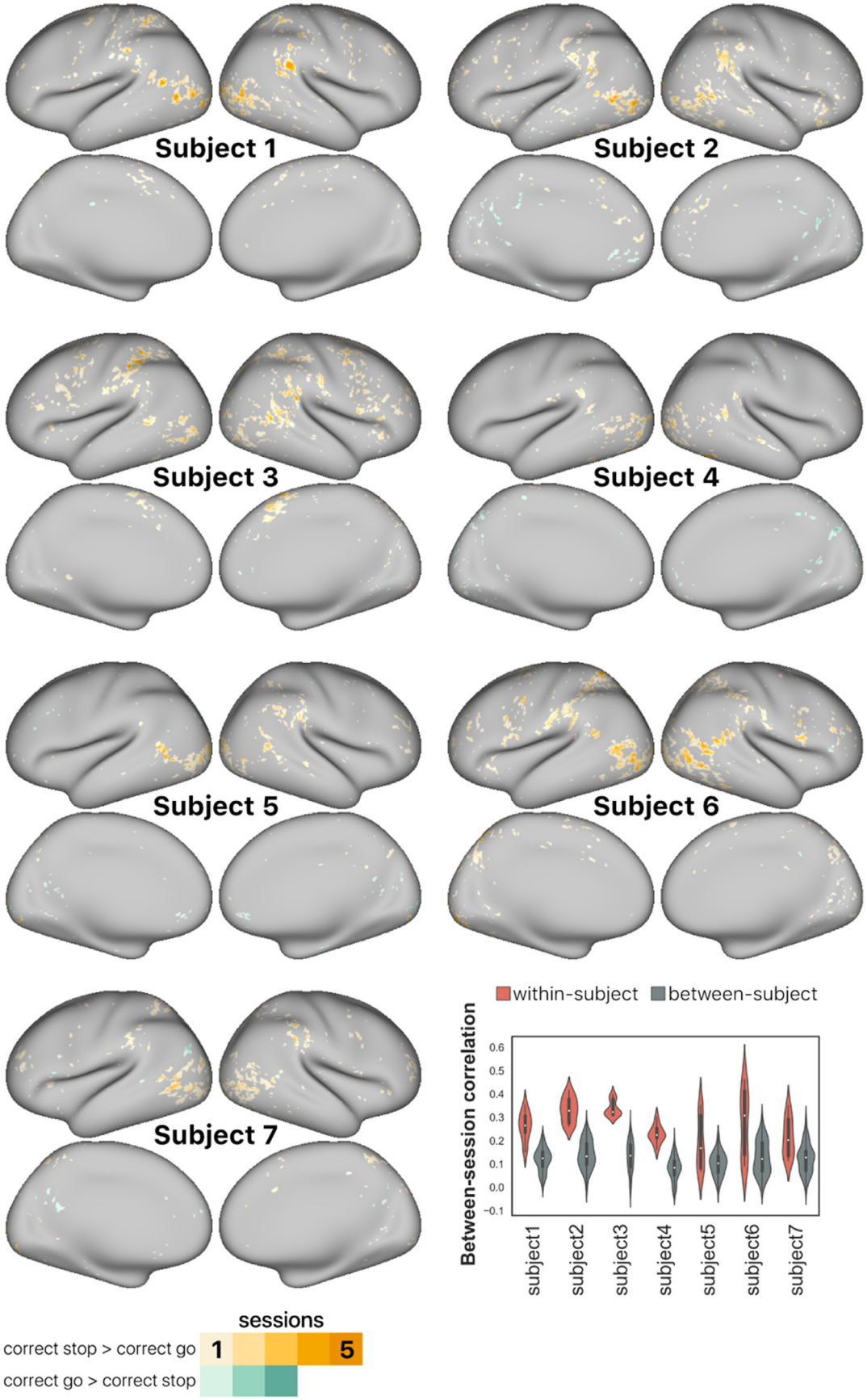

Fig. 3. SST activation maps. Within-participant overlap in SST activation (successful stop versus correct go) across sessions. Darker colors represent activity ($p < 0.0001$) more consistently present across a greater number of sessions (gold = successful stop > correct go; teal = correct go > successful stop). Bottom right panel represents similarity of unthresholded contrasts across sessions for each participant (pink) and between subjects (slate blue). Participants are ordered by their mean within-subject *t*-statistic correlation in the EN-back task.

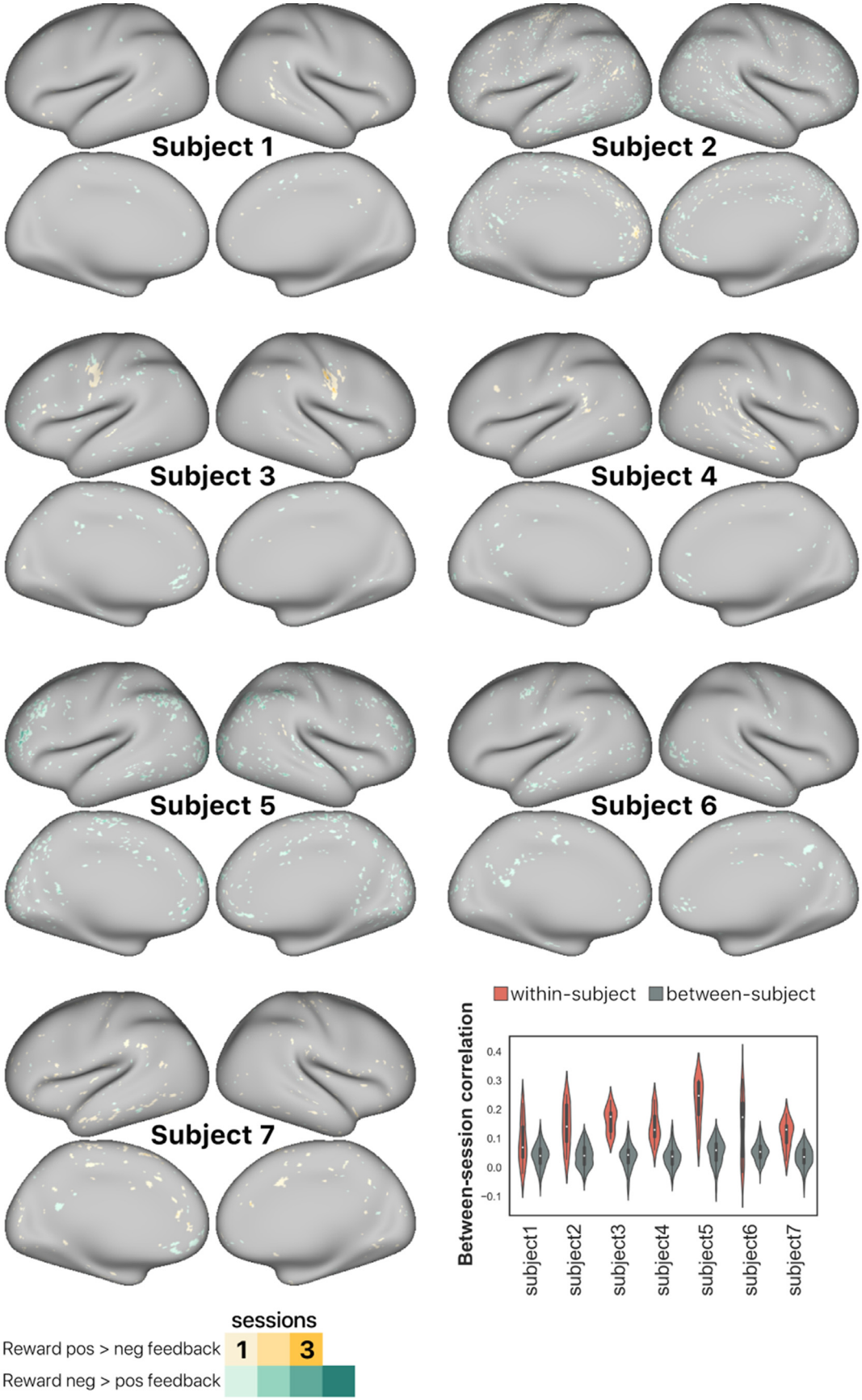

Fig. 4. MID task activation maps. Within-participant overlap in MID task activation (reward success versus fail) across sessions. Darker colors represent activity ($p < 0.0001$) more consistently present across a greater number of sessions (gold = successful reward > fail; teal = fail > successful reward). Bottom right panel represents similarity of unthresholded contrasts across sessions for each participant (pink) and between subjects (slate blue). Participants are ordered according to their mean within-subject *t*-statistic correlation in the EN-back task.

Table 2
Results of mixed effects models testing the effect of session on task performance. Models included session as a fixed effect and participant as a random intercept. Random slopes were included when their addition significantly improved model fit. †Indicates models including random slopes. *Indicates a significant effect of session ($p < 0.05$ corrected for 11 comparisons). Bayes factors reflect evidence for models including session as a fixed effect (H_A) relative to models only including random intercepts for participants (H_0). Bayesian evidence category descriptors are based on Wetzels et al. (2011).

Task	Behavioral measure	*b*	std. err.	df	*F*-stat	*p*	Bayes factor	Bayesian evidence category
SST	Stop-signal delay†	-0.167	0.109	1, 6	2.358	0.176	0.745	Anecdotal evidence for H_0
SST	Stop-signal RT	0.037	0.042	1, 27	.782	0.384	0.026	Very strong evidence for H_0
MID	Practice RT	-0.025	0.116	1, 26.9	.048	0.827	0.050	Strong evidence for H_0
MID	Reward % acc.	-0.073	0.122	1, 32	.365	0.550	0.063	Strong evidence for H_0
MID	Loss % acc.	0.052	0.121	1, 26.2	.188	0.668	0.057	Strong evidence for H_0
MID	Neutral % acc.	0.131	0.112	1, 26.2	.254	0.254	0.095	Strong evidence for H_0
EN-back	0-back % acc.	-0.016	0.105	1, 27	.024	0.877	0.045	Strong evidence for H_0
EN-back	2-back % acc.	0.118	0.067	1, 27	3.050	0.092	0.126	Substantial evidence for H_0
Recognition memory	*d′*	0.136	0.113	1, 27	1.451	0.239	0.099	Strong evidence for H_0
Recognition memory	Hit rate*	0.303	0.049	1, 27	37.645	1.48×10^{-6}	4210.898	Decisive evidence for H_A
Recognition memory	False alarm rate†	0.207	0.177	1, 6	1.370	0.286	33.192	Very strong evidence for H_A

r = 0.51 (subject-wise range [0.31, 0.71]; session-wise range [0.12, 0.76]; Fig. 5). Vertices with significant 2-back > 0-back activity and vertices with significant 0-back > 2-back activity demonstrated overlap in a maximum of all five sessions for each participant.

3.2.2. Anatomical overlap of task activation patterns

Although consistency was generally low for the SST and MID contrast maps, analyses did reveal some consistency in activation within each participant (and across participants) in several brain networks (Fig. 6**A**). Vertices where activity consistently differed (i.e., were significant in more than one session in any participant) between successful stop and correct go SST trials primarily fell within the visual (39% of vertices significant in more than one session in of any participant) and dorsal attention networks (29%; Fig. 6**B**), with higher activity in these regions observed during successful stop trials. Vertices where activity consistently differed between reward success vs. failure MID trials were concentrated within the default mode (54%), ventral attention (25%) and somatomotor networks (17%), with higher activity observed during successful reward trials. Finally, consistency in 2-back vs. 0-back EN-back activation patterns was primarily observed within the frontoparietal (35%) and dorsal attention networks (34%). Note that for the SST and MID contrasts in particular, the total number of vertices used to calculate overlap percentages was low.

3.2.3. Activation patterns are more similar within than between individuals

The within-subject, across-session overlap of activation patterns varied by task and was low in the SST and MID contrasts analyzed here. Given this variability, individual session maps could be (a) noisy observations of an activation pattern that is largely stereotyped across the population or (b) noisy observations of a unique individual-specific activation pattern. It also could be the case that what appears to be session-specific noise reflects session-specific state differences. As an initial approach to disentangling these non-exclusive possibilities, we asked whether each person showed activation patterns that were more similar to themselves than to other individuals. Specifically, for each task, we used linear mixed effects models to compare the similarity of task activation patterns within the same individual across different scan sessions vs. between different individuals. Similarity was measured with the Pearson correlation of unthresholded *t*-statistic contrast images from pairs of sessions. Results revealed that, for all three task contrasts, whole-brain activation patterns were more similar within than between participants (SST correct stop vs. correct go: β [SE] = 0.15 [0.006], $p < 0.0001$; MID reward success vs. failure: β [SE] = 0.10 [0.004]; $p < 0.0001$; EN-back 2-back vs. 0-back: β [SE] = 0.36 [0.008], $p < 0.0001$; Figs. 3-5). Thus, initial analyses suggest that individuals show unique patterns of activity during the SST, MID, and EN-back tasks.

4. Discussion

We collected an accelerated adult equivalent of the ABCD Study dataset, the a-ABCD dataset, in which eight adults completed the ABCD MRI protocol every week for five weeks. We report on task fMRI data from seven adults. This precision fMRI design allowed us to characterize changes in behavior and brain function from one scan session to the next, which may, looking ahead, help differentiate state-like scan-to-scan changes from trait-like developmental change in the large-scale, open-access ABCD Study sample and inform the reliability of fMRI task activation patterns in adulthood. For example, whereas year-to-year changes in EN-back task accuracy that fall within the adult scan-to-scan range may result from state-like changes, year-to-year changes that exceed this range may be attributable to developmental effects.

Despite the power of longitudinal designs for characterizing developmental change, practice, mood, and familiarity effects can introduce confounds for studies in which tasks are repeated across sessions. To characterize potential practice effects in the ABCD Study MRI battery, we assessed how adults' task performance changed across repeated sessions. Contrary to predictions that performance would increase due to growing familiarity with the tasks and/or scanning environment, performance on the in-scanner SST, MID task, and EN-back task did not change over time. This was expected for the SST and MID tasks because difficulty was individualized with staircasing procedures during each scan session. MID practice RTs measured during a practice task before each scan and used to set initial MID trial speed, however, also were consistent across sessions. This finding suggests that participants did not strategically speed up or slow down during pre-scan practice to manipulate upcoming MID task difficulty. Furthermore, performance on the 0-back and 2-back blocks of the EN-back task, which was not individualized, also remained consistent over time. Thus, while perfor-

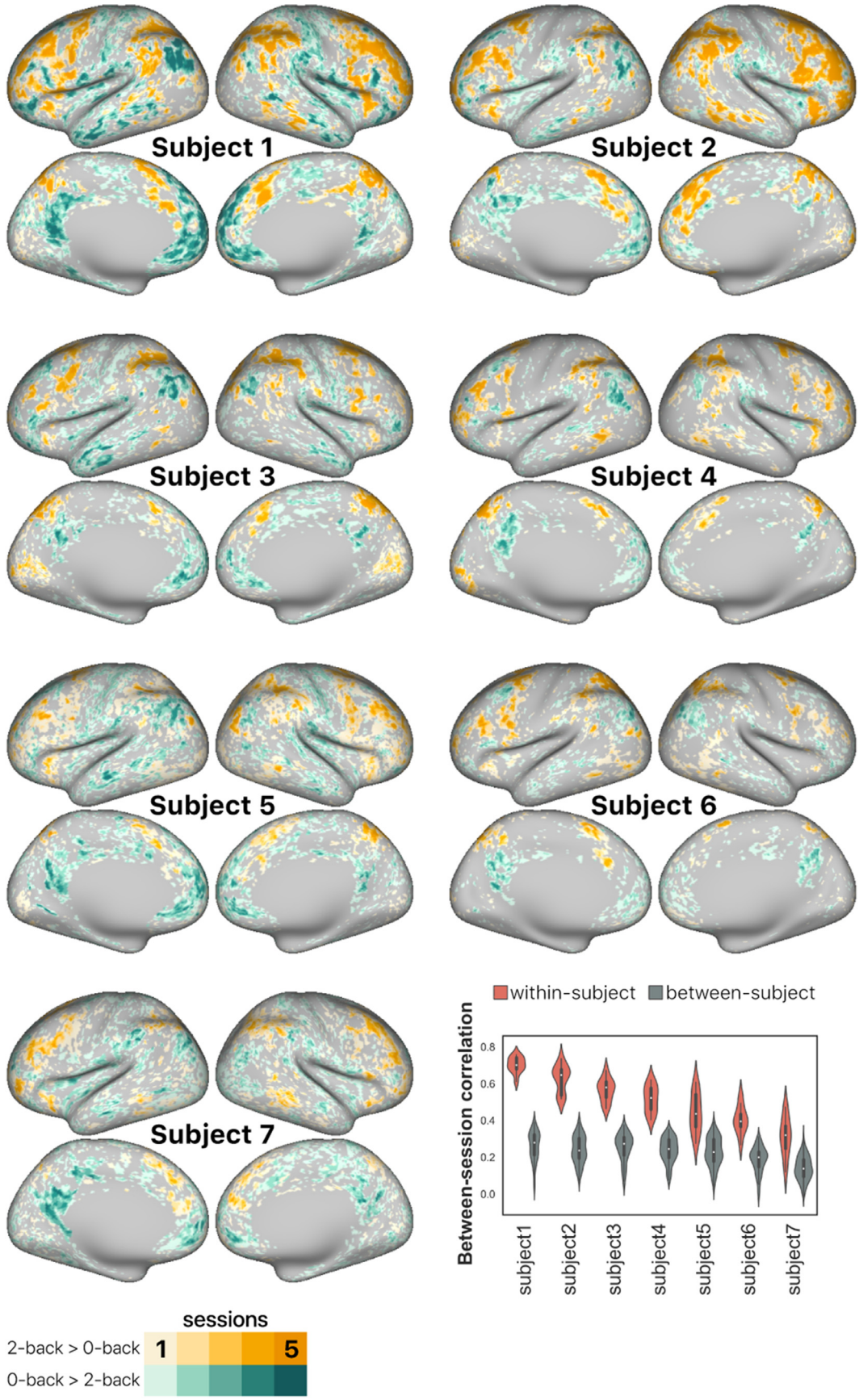

Fig. 5. Within-participant overlap in EN-back task activation (2-back versus 0-back) across sessions. Darker colors represent activity ($p < 0.0001$) more consistently present across a greater number of sessions (gold = 2-back > 0-back; teal = 0-back > 2-back). Bottom right panel represents similarity of unthresholded contrasts across sessions for each participant (pink) and between subjects (slate blue). Participants are ordered according to their mean within-subject *t*-statistic correlation.

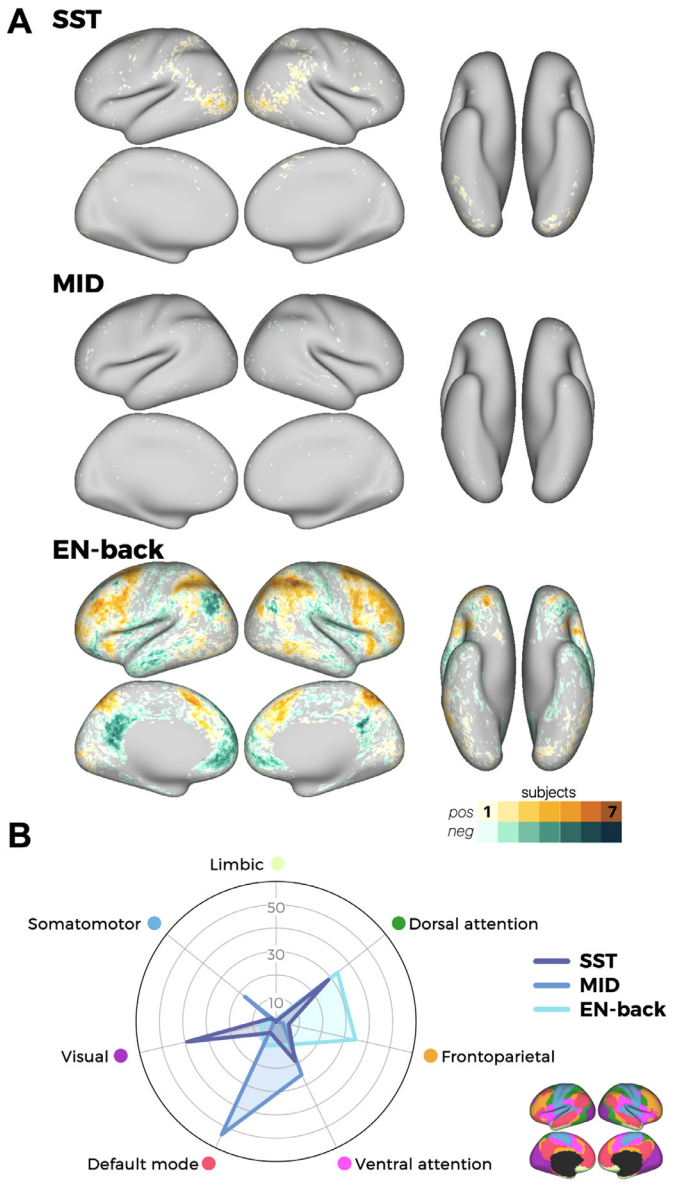

Fig. 6. A) Brain maps reflecting subject-wise overlap of consistent activity across sessions (i.e., vertices that were active during two or more sessions per participant). B) Radar plot illustrating the distributions of activity among functionally-defined networks (Yeo et al., 2011).

mance change from one ABCD Study scan session to the next could reflect both developmental change and task familiarity, we do not see evidence of practice effects in adults in the current design. This is especially noteworthy given that adults performed the tasks only one week apart, whereas ABCD Study youth participants' scan sessions are separated by approximately two years.

In contrast to in-scanner task performance, we did see evidence for changes over time in the out-of-scanner recognition memory task for EN-back stimuli. Over time, participants became more likely to recognize stimuli seen during the EN-back task itself and to incorrectly indicate that new stimuli had been seen during the EN-back task. These two effects—an increase in hit rate to growing familiarity with task stimuli and an increasing false alarm rate due to interference from previous recognition memory tests—counteracted each other, resulting in a stable recognition memory *d'* score over time. Although ABCD Study scan sessions are separated by two years, decreasing the possibility that both "old" images seen during the task and "new" images only seen in the recognition memory test will be recognized from previous testing sessions, researchers investigating changes in recognition memory over time should consider this possibility.

We next characterized the consistency of adults' task activation patterns using representative contrasts of inhibitory control in the SST, reward success in the MID task, and working memory in the EN-back task used in previous work with the ABCD Study sample (Casey et al., 2018). SST and MID task contrast maps were variable across different sessions from the same participant: overlap between significant vertices was low, and the mean correlation between pairs of a participant's session-specific contrast maps was $r = 0.26$ for the SST contrast and $r = 0.14$ for the MID contrast. This pattern of results was generally consistent across the other SST and MID contrasts visualized in the supplemental materials. It also echoes recent work reporting "close to zero" within-session reliability and poor across-session stability for these tasks in ABCD Study youth baseline data and a subset of 2-year-follow-up data (Kennedy et al., 2021). The current results are particularly striking because they show low similarity between SST and MID activation patterns collected weeks—rather than years—apart. EN-back contrast stability was higher, with more overlap between significant vertices observed across sessions and a mean correlation between pairs of a participant's session-specific contrast maps of $r = 0.51$. Thus, the consistency of activation patterns differed by task, potentially enhanced by the EN-back task's block design and disadvantaged by the SST and MID tasks' event-related, individualized designs. An open question to consider in complementary work is the relationship between behavioral and fMRI activation pattern stability in these tasks. That is, why do SST and MID activation patterns vary while performance remains stable due to staircasing? Do activation patterns vary because task timing varies with staircasing, because cognitive states and/or performance strategies vary across scan sessions, or simply due to noise?

What does low a-ABCD task activation pattern stability, especially in the SST and MID tasks, mean for interpretations of ABCD Study data? Based on their observations of poor reliability in ABCD Study task activation patterns, Kennedy and colleagues (2021) argue that these patterns may not be well-suited for individual differences analyses that assume they are stable and trait-like. Our results support their words of caution. They also suggest that, at least with the current ABCD task designs and run lengths, researchers may be more confident in drawing conclusions about individual differences and developmental change from EN-back task contrasts than from SST or MID contrasts.

Looking ahead, future work optimizing task design, imaging protocols, and data collection approaches may help address the relatively low within-subject consistency of SST and MID contrasts. For example, a recent commentary noted design issues with the ABCD Study SST (Bissett et al., 2021) which lead to changes in ABCD Study task scripts (Garavan et al., 2021) after a-ABCD data were collected. In addition, recent work suggests that multiband imaging, such as that used in the ABCD and a-ABCD Studies, decreases the effect size of reward related activity in medial brain regions including the nucleus accumbens and medial prefrontal cortex (Srirangarajan et al., 2021). These advances leave open the possibility that targeted imaging protocols and task designs, and increases in the amount of task data collected per individual (e.g., Elliott et al., 2020; Gordon et al., 2017), may increase task contrast reliability. Task contrast reliability may also increase across development (Kennedy et al., 2021). Although we did not directly compare ABCD and a-ABCD contrast reliability here, future work can explore changes in activation pattern reliability over time.

Maximizing the reliability of task-based measures is central to nearly all research using fMRI. It is especially critical, however, for individualized analytic approaches that use task fMRI to track individual differences in behavior, forecast future outcomes from past developmental trajectories, assess the effects of experiences and environments, and evaluate the efficacy of treatments or interventions. Although initial results in the a-ABCD sample suggest that individuals show relatively *unique* task activation patterns that are more similar within- than be-

tween subjects, future work is needed to both maximize the stability of these patterns and model their sensitivity to varying cognitive, attentional, emotional, and arousal states. Characterizing how activation patterns are influenced by such states can help isolate the trait-like activity patterns most useful for individual-level analyses such as biomarker discovery and personalized prediction.

Future work is also needed to assess changes in performance and stability of task activation patterns over short periods of time in larger participant samples. Although the pattern of behavioral and fMRI results was qualitatively similar in individual participants, it is unlikely that data from the seven individuals analyzed here are representative of all 21–25-year-olds. Large-scale deep phenotyping samples in developmental and adult samples are necessary for evaluating changes observed in large-scale longitudinal samples such as the ABCD Study.

In summary, we anticipate that the a-ABCD dataset will be a useful tool for comparing adult- and adolescent structural, task-based, resting-state, and diffusion-weighted MRI datasets with identical scanning protocols. The dataset is particularly well-suited for disentangling scan-to-scan effects from developmental effects and benchmarking neurodevelopmental change observed in the open-access ABCD Study sample. More broadly, it joins a growing set of tools for characterizing the reliability of brain measures and behavior across childhood, adolescence, and adulthood.

Data and code availability

Behavioral analysis code provided by the ABCD Study Data Analysis, Informatics & Resource Center (DAIRC) are available at https://github.com/ABCD-STUDY/abcd_extract_eprime. Behavioral analysis code specific to this study are available at https://github.com/monicadrosenberg/a-ABCD. Data are available at https://openneuro.org/datasets/ds004097.

Competing Interests

The authors declare no competing interests.

Credit authorship contribution statement

Kristina M. Rapuano: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **May I. Conley:** Data curation, Investigation, Writing – review & editing. **Anthony C. Juliano:** Methodology, Software, Writing – review & editing. **Gregory M. Conan:** Methodology, Software, Writing – review & editing. **Maria T. Maza:** Data curation, Investigation, Writing – review & editing. **Kylie Woodman:** Data curation, Investigation, Writing – review & editing. **Steven A. Martinez:** Data curation, Investigation, Writing – review & editing. **Eric Earl:** Methodology, Software, Writing – review & editing. **Anders Perrone:** Methodology, Software, Writing – review & editing. **Eric Feczko:** Methodology, Software, Writing – review & editing. **Damien A. Fair:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Richard Watts:** Investigation, Methodology, Software, Writing – review & editing. **B.J. Casey:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Monica D. Rosenberg:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing.

ABCD Acknowledgement

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children ages 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147. A full list of supporters is available at abcdstudy.org/nih-collaborators. A listing of participating sites and a complete listing of the study investigators can be found at abcdstudy.org/principal-investigators.html. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

The ABCD data repository grows and changes over time. The ABCD data used in this report came from NIMH Data Archive Digital Object Identifier 10.15154/1504041. DOIs can be found at nda.nih.gov/study.html?id=721.

This research also benefited from the ABCD Workshop on Brain Development and Mental Health, supported by the National Institute of Mental Health of the National Institutes of Health under Award Number R25MH120869 and by UG3-DA045251 from the National Institute of Drug Abuse.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119215.

References

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1). doi:10.1016/j.media.2007.06.004.

Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., 2013. Function in the human connectome: task-fMRI and individual differences in behavior. Neuroimage 80, 169–189. doi:10.1016/j.neuroimage.2013.05.033.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 1 (Issue 1). doi:10.18637/jss.v067.i01, 2015.

Bennett, C.M., Miller, M.B., 2013. fMRI reliability: influences of task and experimental design. Cognit. Affect. Behav. Neurosci. 13 (4), 690–702. doi:10.3758/s13415-013-0195-1.

Berman, M.G., Park, J., Gonzalez, R., Polk, T.A., Gehrke, A., Knaffla, S., Jonides, J., 2010. Evaluating functional localizers: the case of the FFA. Neuroimage 50 (1), 56–71. doi:10.1016/j.neuroimage.2009.12.024.

Bissett, P.G., Hagen, M.P., Jones, H.M., Poldrack, R.A., 2021. Design issues and solutions for stop-signal data from the Adolescent Brain Cognitive Development (ABCD) study. ELife 10. doi:10.7554/eLife.60185.

Buimer, E.E.L., Pas, P., Brouwer, R.M., Froeling, M., Hoogduin, H., Leemans, A., Luijten, P., van Nierop, B.J., Raemaekers, M., Schnack, H.G., Teeuw, J., Vink, M., Visser, F., Hulshoff Pol, H.E., Mandl, R.C.W., 2020. The YOUth cohort study: MRI protocol and test-retest reliability in adults. Dev. Cognit. Neurosci. 45, 100816. doi:10.1016/j.dcn.2020.100816.

Byrd, R., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16 (5), 1190–1208. doi:10.1137/0916069.

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Dale, A.M., 2018. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. Dev. Cognit. Neurosci. doi:10.1016/j.dcn.2018.03.001.

Casey, B.J., Cohen, J.D., O'Craven, K., Davidson, R.J., Irwin, W., Nelson, C.A., Noll, D.C., Hu, X., Lowe, M.J., Rosen, B.R., Truwitt, C.L., Turski, P.A., 1998. Reproducibility of fMRI results across four institutions using a spatial working memory task. Neuroimage 8 (3), 249–261. doi:10.1006/nimg.1998.0360.

Chaarani, B., Hahn, S., Allgaier, N., Adise, S., Owens, M.M., Juliano, A.C., Yuan, D.K., Loso, H., Ivanciu, A., Albaugh, M.D., Dumas, J., Mackey, S., Laurent, J., Ivanova, M., Hagler, D.J., Cornejo, M.D., Hatton, S., Agrawal, A., Aguinaldo, L., Garavan, H.P., 2021. Baseline brain function in the preadolescents of the ABCD Study. Nat. Neurosci. 24 (8). doi:10.1038/s41593-021-00867-9.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9 (2). doi:10.1006/nimg.1998.0395.

Elliott, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., Sison, M.L., Moffitt, T.E., Caspi, A., Hariri, A.R., 2020. What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. Psychol. Sci. 31 (7), 792–806. doi:10.1177/0956797620916786.

Fafrowicz, M., Bohaterewicz, B., Ceglarek, A., Cichocka, M., Lewandowska, K., Sikora-Wachowicz, B., Oginska, H., Beres, A., Olszewska, J., Marek, T., 2019. Beyond the low frequency fluctuations: morning and evening differences in human brain. Front. Hum. Neurosci. 13. doi:10.3389/fnhum.2019.00288.

Feczko, E., Conan, G., Marek, S., Tervo-Clemmens, B., Cordova, M., Doyle, O., Earl, E., Perrone, A., Sturgeon, D., Klein, R., Harman, G., Kilamovich, D., Hermosillo, R., Miranda-Dominguez, O., Adebimpe, A., Bertolero, M., Cieslak, M., Covitz, S., Hendrickson, T., Fair, D. A. (2021). Adolescent brain cognitive development (ABCD) community MRI collection and utilities. *BioRxiv*, 2021.07.09.451638. doi:10.1101/2021.07.09.451638.

Fischl, B., 2012. FreeSurfer. Neuroimage 62 (Issue 2). doi:10.1016/j.neuroimage.2012.01.021.

Friedman, L., Stern, H., Brown, G.G., Mathalon, D.H., Turner, J., Glover, G.H., Gollub, R.L., Lauriello, J., Lim, K.O., Cannon, T., Greve, D.N., Bockholt, H.J., Belger, A., Mueller, B., Doty, M.J., He, J., Wells, W., Smyth, P., Pieper, S., Potkin, S.G., 2008. Test–retest and between-site reliability in a multicenter fMRI study. Hum. Brain Mapp. 29 (8), 958–972. doi:10.1002/hbm.20440.

Garavan, H., Chaarani, B., Hahn, S., Allgaier, N., Juliano, A., Yuan, D. K., Orr, C., Watts, R., Wager, T. D., Ruiz de Leon, O., Hagler, D. J., & Potter, A. (2021). The ABCD stop signal data: response to Bissett et al. *BioRxiv*, 2020.07.27.223057. doi:10.1101/2020.07.27.223057.

Gee, D.G., McEwen, S.C., Forsyth, J.K., Haut, K.M., Bearden, C.E., Addington, J., Goodyear, B., Cadenhead, K.S., Mirzakhanian, H., Cornblatt, B.A., Olvet, D., Mathalon, D.H., McGlashan, T.H., Perkins, D.O., Belger, A., Seidman, L.J., Thermenos, H., Tsuang, M.T., van Erp, T.G.M., Cannon, T.D., 2015. Reliability of an fMRI paradigm for emotional processing in a multisite longitudinal study. Hum. Brain Mapp. 36 (7), 2558–2579. doi:10.1002/hbm.22791.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105–124. doi:10.1016/j.neuroimage.2013.04.127.

Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J.M., Coalson, R.S., Nguyen, A.L., McDermott, K.B., Shimony, J.S., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., Nelson, S.M., Dosenbach, N.U.F., 2017. Precision functional mapping of individual human brains. Neuron 95 (4), 791–807. doi:10.1016/j.neuron.2017.07.011, e7.

Hagler, D.J., Hatton, S., Cornejo, M.D., Makowski, C., Fair, D.A., Dick, A.S., Sutherland, M.T., Casey, B.J., Barch, D.M., Harms, M.P., Watts, R., Bjork, J.M., Garavan, H.P., Hilmer, L., Pung, C.J., Sicat, C.S., Kuperman, J., Bartsch, H., Xue, F., Dale, A.M., 2019. Image processing and analysis methods for the adolescent brain cognitive development study. Neuroimage 202. doi:10.1016/j.neuroimage.2019.116091.

Haller, S.P., Kircanski, K., Stoddard, J., White, L.K., Chen, G., Sharif-Askary, B., Zhang, S., Towbin, K.E., Pine, D.S., Leibenluft, E., Brotman, M.A., 2018. Reliability of neural activation and connectivity during implicit face emotion processing in youth. Dev. Cognit. Neurosci. 31, 67–73. doi:10.1016/j.dcn.2018.03.010.

Herting, M.M., Gautam, P., Chen, Z., Mezher, A., Vetter, N.C., 2018. Test-retest reliability of longitudinal task-based fMRI: Implications for developmental studies. Dev. Cognit. Neurosci. 33, 17–26. doi:10.1016/j.dcn.2017.07.001.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62 (2), 782–790. doi:10.1016/j.neuroimage.2011.09.015.

Juliano, A. C., Conan, G., Moore, L. A., Feczko, E., Fair, D., & Garavan, H. (2021). *ABCD-BIDS task fMRI pipeline*. doi:10.17605/OSF.IO/CSUYB.

Kennedy, J. T., Harms, M. P., Korucuoglu, O., Astafiev, S. V, Barch, D. M., Thompson, W. K., Bjork, J. M., & Anokhin, A. P. (2021). Reliability and stability challenges in ABCD Task fMRI data. *BioRxiv*, 2021.10.08.463750. doi:10.1101/2021.10.08.463750.

Knutson, B., Westdorp, A., Kaiser, E., Hommer, D., 2000. FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12 (1), 20–27. doi:10.1006/nimg.2000.0593.

Kragel, P.A., Han, X., Kraynak, T.E., Gianaros, P.J., Wager, T.D., 2021. Functional MRI can be highly reliable, but it depends on what you measure: a commentary on Elliott et al. (2020). Psychol. Sci. 32 (4), 622–626. doi:10.1177/0956797621989730.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 1 (Issue 13). doi:10.18637/jss.v082.i13, 2017.

Li, Xiangrui, Morgan, P.S., Ashburner, J., Smith, J., Rorden, C, 2016. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J. Neurosci. Methods 264. doi:10.1016/j.jneumeth.2016.03.001.

Li, Xiong, Pan, Y., Fang, Z., Lei, H., Zhang, X., Shi, H., Ma, N., Raine, P., Wetherill, R., Kim, J.J., Wan, Y., Rao, H, 2020. Test-retest reliability of brain responses to risk-taking during the balloon analogue risk task. Neuroimage 209, 116495. doi:10.1016/j.neuroimage.2019.116495.

Logan, G. D., 1994. On the ability to inhibit thought and action: a users' guide to the stop signal paradigm. In: Dagenbach, D., Carr, T.H. (Eds.), Inhibitory Processes in Attention, Memory, and Language. Academic Press, San Diego, CA, pp. 189–239.

Luciana, M., Bjork, J.M., Nagel, B.J., Barch, D.M., Gonzalez, R., Nixon, S.J., Banich, M.T., 2018. Adolescent neurocognitive development and impacts of substance use: overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. Dev. Cognit. Neurosci. 32, 67–79. doi:10.1016/j.dcn.2018.02.006.

Makowski, D., Ben-Shachar, M., Lüdecke, D., 2019. bayestestR: describing effects and their uncertainty, existence and significance within the bayesian framework. J. Open Source Softw. 4 (40). doi:10.21105/joss.01541.

McGonigle, D.J., Howseman, A.M., Athwal, B.S., Friston, K.J., Frackowiak, R.S.J., Holmes, A.P., 2000. Variability in fMRI: an examination of intersession differences. Neuroimage 11 (6), 708–734. doi:10.1006/nimg.2000.0562.

Nash, J.C., Varadhan, R., 2011. Unifying optimization algorithms to aid software system users: optimx for R. J. Stat. Softw. 1 (Issue 9). doi:10.18637/jss.v043.i09, 2011.

Noble, S., Scheinost, D., Constable, R.T., 2021. A guide to the measurement and interpretation of fMRI test-retest reliability. Curr. Opin. Behav. Sci. 40, 27–32. doi:10.1016/j.cobeha.2020.12.012.

Orban, C., Kong, R., Li, J., Chee, M.W.L., Yeo, B.T.T., 2020. Time of day is associated with paradoxical reductions in global signal fluctuation and functional connectivity. PLoS Biol. 18 (2), e3000602. doi:10.1371/journal.pbio.3000602.

Poldrack, R.A., Laumann, T.O., Koyejo, O., Gregory, B., Hover, A., Chen, M.Y., Gorgolewski, K.J., Luci, J., Joo, S.J., Boyd, R.L., Hunicke-Smith, S., Simpson, Z.B., Caven, T., Sochat, V., Shine, J.M., Gordon, E., Snyder, A.Z., Adeyemo, B., Petersen, S.E., Mumford, J.A., 2015. Long-term neural and physiological phenotyping of a single human. Nat. Commun. 6. doi:10.1038/ncomms9885.

Raemaekers, M., du Plessis, S., Ramsey, N.F., Weusten, J.M.H., Vink, M., 2012. Test–retest variability underlying fMRI measurements. Neuroimage 60 (1), 717–727. doi:10.1016/j.neuroimage.2011.11.061.

Riley, E., Esterman, M., Fortenbaugh, F.C., DeGutis, J., 2017. Time-of-day variation in sustained attentional control. Chronobiol. Int. 34 (7). doi:10.1080/07420528.2017.1308951.

Rosenberg, M.D., Casey, B.J., Holmes, A.J., 2018. Prediction complements explanation in understanding the developing brain. Nat. Commun. 9 (1), 589. doi:10.1038/s41467-018-02887-9.

Rosenberg, M.D., Martinez, S.A., Rapuano, K.M., Conley, M.I., Cohen, A.O., Cornejo, M.D., Hagler, D.J., Meredith, W.J., Anderson, K.M., Wager, T.D., Feczko, E., Earl, E., Fair, D.A., Barch, D.M., Watts, R., Casey, B.J., 2020. Behavioral and neural signatures of working memory in childhood. J. Neurosci. 40 (26), 5090LP–505104. doi:10.1523/JNEUROSCI.2841-19.2020.

Simmons, C., Conley, M.I., Gee, D.G., Baskin-Sommers, A., Barch, D.M., Hoffman, E.A., Huber, R.S., Iacono, W.G., Nagel, B.J., Palmer, C.E., Sheth, C.S., Sowell, E.R., Thompson, W.K., Casey, B.J., 2021. Responsible use of open-access developmental data: the adolescent brain cognitive development (ABCD) study. Psychol. Sci. 32 (6), 866–870. doi:10.1177/09567976211003564.

Srirangarajan, T., Mortazavi, L., Bortolini, T., Moll, J., Knutson, B., 2021. Multi-band FMRI compromises detection of mesolimbic reward responses. Neuroimage 244, 118617. doi:10.1016/j.neuroimage.2021.118617.

Sturgeon, D.T., Earl, E., Perrone, A., Snider, K., 2021. DCAN-Labs/abcd-hcp-pipeline: minor update to DCAN BOLD Processing and MRE version (v0.0.4). Zenodo doi:10.5281/zenodo.4571051.

Volkow, N.D., Koob, G.F., Croyle, R.T., Bianchi, D.W., Gordon, J.A., Koroshetz, W.J., Pérez-Stable, E.J., Riley, W.T., Bloch, M.H., Conway, K., Deeds, B.G., Dowling, G.J., Grant, S., Howlett, K.D., Matochik, J.A., Morgan, G.D., Murray, M.M., Noronha, A., Spong, C.Y., Weiss, S.R.B., 2018. The conception of the ABCD study: From substance use to a broad NIH collaboration. Dev. Cognit. Neurosci. 32, 4–7. doi:10.1016/j.dcn.2017.10.002.

Wechsler, D., 2011. *Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II)*. NCS Pearson, San Antonio, TX.

Wetzels, R., Matzke, D., Lee, M.D., Rouder, J.N., Iverson, G.J., Wagenmakers, E.-J., 2011. Statistical evidence in experimental psychology: an empirical comparison using 855 t tests. Perspect. Psychol. Sci. 6 (3), 291–298. doi:10.1177/1745691611406923.

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14 (6). doi:10.1006/nimg.2001.0931.

Yau, W.-Y.W., Zubieta, J.-K., Weiland, B.J., Samudra, P.G., Zucker, R.A., Heitzeg, M.M., 2012. Nucleus accumbens response to incentive stimuli anticipation in children of alcoholics: relationships with precursive behavioral risk and lifetime alcohol use. J. Neurosci. 32 (7), 2544–2551. doi:10.1523/JNEUROSCI.1390-11.2012.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, 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 (3), 1125–1165. doi:10.1152/jn.00338.2011.