Published in final edited form as:

Neuroimage. 2018 May 15; 172: 674–688. doi:10.1016/j.neuroimage.2017.12.044.

Subtyping Cognitive Profiles in Autism Spectrum Disorder Using a Random Forest Algorithm

E Feczko¹, N Balba¹, O Miranda-Dominguez¹, M Cordova¹, SL Karalunas³, L Irwin¹, DV Demeter⁵, AP Hill^{2,6}, BH Langhorst¹, Grieser J Painter¹, J Van Santen^{2,6}, EJ Fombonne^{2,3}, JL Nigg^{1,3,4}, and DA Fair^{1,3,4}

¹Department of Behavioral Neuroscience, Oregon Health & Science University, Portland, OR, 97239, USA

²Department of Pediatrics, Oregon Health & Science University, Portland, OR, 97239, USA

³Department of Psychiatry, Oregon Health & Science University, Portland, OR, 97239, USA

⁴Advanced Imaging Research Center, Oregon Health & Science University, Portland, OR, 97239, USA

⁵The University of Texas at Austin, Department of Psychology, Austin, TX, 78713, USA

⁶Center for Spoken Language Understanding, Institute on Development & Disability, Oregon Health & Science University, Portland, OR, 97239, USA

Abstract

DSM-5 Autism Spectrum Disorder (ASD) comprises a set of neurodevelopmental disorders characterized by deficits in social communication and interaction and repetitive behaviors or restricted interests, and may both affect and be affected by multiple cognitive mechanisms. This study attempts to identify and characterize cognitive subtypes within the ASD population using a random forest (RF) machine learning classification model. We trained our model on measures from seven tasks that reflect multiple levels of information processing. 47 ASD diagnosed and 58 typically developing (TD) children between the ages of 9 and 13 participated in this study. Our RF model was 72.7% accurate, with 80.7% specificity and 63.1% sensitivity. Using the RF model, we measured the proximity of each subject to every other subject, generating a distance matrix between participants. This matrix was then used in a community detection algorithm to identify subgroups within the ASD and TD groups, revealing 3 ASD and 4 TD putative subgroups with unique behavioral profiles. We then examined differences in functional brain systems between diagnostic groups and putative subgroups using resting-state functional connectivity magnetic resonance imaging (rsfMRI). Chi-square tests revealed a significantly greater number of between group differences ($p < .05$) within the cingulo-opercular, visual, and default systems as well as differences in inter-system connections in the somato-motor, dorsal attention, and subcortical systems. Many of these differences were primarily driven by specific subgroups suggesting that our method could potentially parse the variation in brain mechanisms affected by ASD.

Corresponding author: Eric Feczko, Oregon Health and Sciences University, 3181 Sam Jackson Park Rd, Portland, OR, 97221, feczko@ohsu.edu.

Authors Feczko E and Balba N provided equal contributions to the paper and should be treated as co-first authors.

Keywords

supervised learning; random forests; autism; functional connectivity; MRI

Introduction

Issues in diagnosing and treating ASD

Lack of precision medicine in ASD—Autism Spectrum Disorders (ASD) comprise altered social interactions and/or communication, as well as the presence of stereotyped or repetitive behavior ¹. The prevalence of ASD in the global population has been estimated around 1%, but that number has been growing over the past decade ^{2,3}. The variability in symptoms, severity, and adaptive behavior impairment within the ASD population ⁴ complicates the development of effective treatments and improved diagnostic measures. Such variation also suggests the possibility of discrete ASD subphenotypes and is consistent with the evidence that ASD may encompass multiple etiologies ^{1,5}. Therefore, identifying and differentiating subgroups in this population should help refine ASD diagnostic criteria and further the study of precision medicine for individuals with ASD.

Heterogeneity in ASD—The etiology of ASD is complex, and the ASD diagnosis has been related to multiple cognitive, sensory, and motor faculties ⁶. We focused here on the cognitive domain. A thorough review of cognitive mechanisms underlying ASD suggested that non-social cognitive mechanisms, including reward, executive function, attention, visual and auditory processing, may affect the presentation of social behavior regardless of specific impairment or the existence of domain-specific social cognitive mechanisms⁷. We examined seven cognitive domains related to information processing and control that have varying levels of association with ASD: spatial working memory, response inhibition, temporal discounting of reward, attentional vigilance, facial recognition, facial affect processing and vocal affect processing.

Working Memory: Working memory refers here to a limited capacity cognitive system that retains information in an accessible state which supports human thought processes⁸. A vast literature in ASD reveals inconsistent findings as to whether visuospatial working memory may be impaired, suggesting the existence of ASD subgroups, which may drive the observed impairments. Early studies of working memory showed that high ⁹, but not low ¹⁰, functioning children with autism had impairments in verbal and non-verbal working memory. Another found no differences in working memory between children with or without ASD ¹¹. Measures of non-verbal working memory on a non-spatial and non-verbal self-ordered pointing task correlate with visuospatial memory in children with ASD but not children without ASD ¹². In contrast, children without ASD, but not children with ASD, show a relationship between language ability and verbal working memory ¹². Such heterogeneity may reflect differences in how individuals with ASD utilize visuospatial memory to augment non-verbal working memory, whereas individuals without ASD may utilize language to augment verbal working memory ¹³.

More recent studies have supported the hypothesis that children with ASD may use different cognitive mechanisms to support working memory. A large-scale study revealed that children with ASD exhibited lower performance than unaffected children on a spatial span task¹⁴, requiring children to repeat a sequence of fixed spatial locations indicated by a series of changing colors. Interestingly, the ASD participants had significantly lower verbal, but not performance, IQ. This study is consistent with findings from two recent studies on children with ASD^{15,16}, one of which showed that better performance on working memory tasks predicted faster development of play behavior¹⁵. However, another recent study found no differences in a similar spatial span task¹⁷. Taken together, all of these findings suggest working memory differences between children with and without ASD are inconsistent, and may be affected by sample differences that comprise different ASD subgroups.

Response Inhibition: Response inhibition refers here to the ability to inhibit a prepotent response, a lower level component of executive function¹⁸. Over 40 studies have examined whether response inhibition is different between individuals with and without ASD¹⁹. While a number of these studies are underpowered, several use large sample sizes and previously validated psychophysical tests. The results from these studies are quite variable, despite large sample sizes and similar task designs. For example, Guerts and colleagues used a stop task to compare stop signal reaction times between TD and ASD children and found a large effect of diagnosis²⁰, while a more recent study employing the same task found only a small effect of ASD when examining commission errors²¹. Although sampling variation may explain divergent results, an interesting possibility is that heterogeneity in ASD helps explain the inconsistency across the literature¹⁹.

Temporal Discounting of Reward: Temporal discounting refers here to the weakening of the subjective value of a reward due to a delay²². A few studies^{23–25} reveal that those with ASD have altered performance on delayed reward discounting tasks. On average, people naturally prefer immediate to delayed rewards of similar values. Different types of rewards may be discounted differently, and may reflect varying preferences for rewards associated with goal-oriented behavior. For example, individuals with ASD discount monetary and social rewards similarly, whereas typically developing (TD) individuals discount social rewards more than monetary rewards²⁴. ASD individuals may also discount monetary rewards more steeply with respect to time than TD individuals²⁵.

Attentional Vigilance: Attentional vigilance refers to the ability to maintain an alert state in the absence of an alerting stimulus. It is often measured using continuous performance tasks (CPTs). ASD performance on CPTs show mixed results. An early study found no difference between children with and without ASD on CPT performance. However, the task used long displays and the parameters of the task were not shifted throughout²⁶. A more recent study using the same version of the task also failed to find differences between children with and without an ASD. However, they did find differences in EEG signals that are important for sustained and selective attention²⁷, suggesting that individuals with ASD may use an alternative, perhaps compensatory, strategy to perform similarly on CPTs. Consistent with this hypothesis, individuals with ASD show impaired performance on CPTs where the ratio of distractors to targets²⁸ or inter-stimulus interval²⁹ varies over the task duration. On the

other hand, increasing attentional demands by crowding the visual display does not seem to affect performance in participants with ASD ³⁰.

Processing of facial features, vocal affect, and facial emotion: Previous work has repeatedly suggested that individuals with ASDs may have trouble processing the arrangements of facial features, which may impair facial identity recognition and the ability to link speech to facial expressions. Individuals with ASD show impairments in searching for the eye region on a face ³¹. Unlike TD individuals, individuals with ASD are not faster at recognizing a part of the face when it is placed in the context of a whole face ³², and performance on facial identity recognition is not maintained when the orientation of a face is altered ³³. Impairments in face processing may affect other domains; individuals with an ASD have difficulty integrating visual facial and auditory speech information ³⁴ and do not use visual information from the mouth to guide speech perception ³⁵.

However, results on facial emotion recognition are more mixed ³⁶. Earlier studies found wide variation in facial emotion recognition performance in adults with an ASD ^{37,38}. More recent studies have shown that facial recognition can be improved in ASD, but that this improvement may not generalize when recognizing emotions from faces ³⁹. ASD participants trained to recognize basic emotions like ‘happy’ or ‘sad’ for a particular set of identities did not improve recognition on faces from novel identities. Furthermore, ASD participants did not improve at recognizing emotion when the eyes were presented in the context of a whole face, suggesting that such training did not enable individuals with ASD to process the eyes holistically ³⁹.

In summary, multiple information processing streams may be affected in individuals with ASD, but the types of impairment may be heterogeneous within the ASD population, with different individuals showing varying patterns of difficulty. Critically, it is difficult to disentangle from these studies whether individuals with an ASD diagnosis comprise distinct subgroups, as shown by working memory and response inhibition findings. Therefore, it is critical to test whether ASD is heterogeneous categorically and/or multi-dimensionally. The identification of distinct ASD subgroups may enable better mapping of the cognitive domains affected by and/or responsible for ASD.

Lack of clear biomarkers in ASD

Due to the wide variation in behavioral measures related to ASD, many studies have sought brain-based biological markers to identify a common etiology across individuals with ASD. Markers that are measurable via MRI are highly desirable, because they may represent potential targets for diagnostic tools and or treatments. Unfortunately, the results of these studies are varied due to differences in both study design and sample composition.

Structural brain biomarkers indicating heterogeneity—Reviews of structural MRI findings in ASD have found a wide range of putative biomarkers across independent studies ^{40–42}. Whole brain-volume ⁴³ developmental trajectories may differ between individuals with and without ASD. Regionally, the temporal-parietal junction ⁴⁴, anterior insula ^{44,45}, posterior cingulate ^{46,47}, lateral and medial prefrontal ⁴⁶, corpus-callosum ⁴⁸, intra-parietal sulcus ^{45,49}, and occipital cortex ⁴⁷, have all been shown to be different between samples

with and without ASD. This has led a number of reviewers to suggest that the heterogeneity within the disorder may account for the divergent findings^{40,41}. Indeed, an interesting study by Christine Nordahl in 2007 examined differences between individuals diagnosed with high-functioning autism, Asperger's, and low functioning autism. Compared to TD individuals, these three samples showed varying cortical folding signatures, indicating that the mechanisms underlying the diagnosis for these samples may differ⁴⁵.

Functional brain biomarkers indicating heterogeneity—Studies of functional brain biomarkers for ASD have largely centered on studies of resting state functional connectivity MRI (rsfMRI) for two reasons. First, the hemodynamic response in ASD children has been shown to be largely similar to the hemodynamic response in TD children⁵⁰, suggesting that differences in functional MRI reflect differences in neural activity. Second, the absence of a task enables one to examine differences across multiple brain regions and/or networks, similar to structural MRI.

Unfortunately, findings from rsfMRI have also varied considerably from study to study. Studies have found altered connectivity within the dorsal attention network⁵¹; default mode-network (DEF; ⁵²); whole-brain^{53,54} and subcortical-cortical⁵⁵ underconnectivity; whole-brain⁵⁶ and cortical-subcortical⁵⁷ hyperconnectivity; and altered connectivity within a discrete set of regions dubbed the “social brain”⁵⁸. Some studies^{59,60} found no differences in functional connectivity. All of these studies differ not only in MRI processing strategies, but also in the diagnostic inclusion/exclusion criteria. More recent studies^{51,58,59} also examined differences in processing strategy, but continued to show discrepant results. Taken together, the findings strongly suggest that ASD heterogeneity may limit the replicability of findings.

Machine Learning approaches in classifying ASD

Machine learning algorithms provide data-driven methods that can characterize ASD heterogeneity by identifying data-driven subgroups of individuals with ASD. However, most studies using machine-learning algorithms focused only on the identification of individuals with ASD, despite recent studies demonstrating moderate success using such algorithms. A large number of studies have tested whether imaging biomarkers can classify whether an individual has or does not have ASD. Early studies had small sample sizes under 100 individuals and showed high classification rates ranging from 80 to 97 percent accurate^{61–64}. Larger scale studies greater than 100 individuals typically showed modest accuracy in range of 60 to 80 percent^{65–67}. The discrepancies may indicate poor control of motion in some cases or over-fit models in others⁶⁸. Alternatively, the discrepancies might be the result of ASD heterogeneity. Along these latter lines, one of the best classifications of ASD was performed using Random Forests (RF; ⁶⁷). RFs are random ensembles of independently grown decision trees, where each decision tree votes as a weak classifier, and classification into the same group can occur through different pathways. ASD classification was improved when behavioral features were incorporated into models, suggesting that ASD may be stratified by differences in brain function and behavior⁶⁵. Interestingly, random forests can also enable the identification of subgroups⁶⁹, however, to our knowledge no machine learning approach has attempted to do so for individuals diagnosed with ASD.

Novel use of Random Forest (RF) in identifying subgroups within sample

Here we implement a novel approach for using RFs to identify more homogenous ASD subgroups. RFs is a random ensemble classification approach that iteratively grows decision trees to classify data. The RF model produces a proximity matrix that indicates the similarity between participants. This proximity matrix illustrates how often a pair of subjects were grouped into the same terminal node of each decision tree within the RF and is similar to a correlation matrix. Conceptually, we can recast the proximity matrix as a graph, and a community detection algorithm⁷⁰ can be used to identify putative subgroups. Several recent studies have used community detection to characterize subpopulations⁷¹. However, one limitation from the approach as it is currently being used is that the community detection approach does not tie the sub-grouping to the outcome measurement of interest. In other words, prior studies have not evaluated whether the similarity measured between participants, which drives the community detection, is associated with the clinical diagnosis. Thus, an approach that ties the defined sub-populations to the clinical diagnosis is better equipped to identify clinically relevant subgroups. We posit that the combination of random forest classification and community detection can assist with this goal.

In the current report we classify children with and without ASD using several information processing and control measures. To attempt to validate the group assignments identified from the cognitive measures, we then compared the strength of rsfcMRI connections, within or between neural systems, across the identified subgroups. Such a link would provide external evidence that these subgroups differ in functional brain organization as it pertains to an ASD diagnosis.

Methods

Participants/Demographics

Participants—The study sample consisted of 105 children between the ages of 9 and 13. Age demographics are shown in Table 1, PDS in Table S1, and all other demographics are shown in Table 2. The ASD group was recruited by community outreach and referrals from a nearby autism treatment center and included 47 children (11 females) with a mean age of 12.15 years ($SD = 2.12$) across all tests. All ASD children had their diagnosis confirmed (using DSM-IV criteria) by a diagnostic team that included two licensed psychologists and a child psychiatrist, and were assessed with a research reliable Autism Diagnostic Observation Schedule Second Edition (ADOS; mean ASD = 12.36, $SD = 3.371$), Autism Diagnostic Interview-Revised interview (ADI-R) and by the Social Responsiveness Scale Second Edition (SRS; TD mean = 17.8, $SD = 10.45$; ASD mean = 92.32, $SD = 27.02$) surveys filled out by parents of the children. The TD group included 58 children (31 females) with a mean age of 10.29 years ($SD 2.16$) for all tests. A Fisher's exact test indicated that gender was significantly different between the two groups ($p = 0.025$). It should be noted that the gender difference between our groups is consistent with the fact that males are at increased risk for autism in the general population. Parental pubertal developmental stage (PDS) report was used to assess pubertal stage. The PDS information was acquired once for all participants, but was untied to the tasks or MRI visits, which limits our ability to infer from it. For each MRI and task visit, we calculated the difference between the date of PDS acquisition and the

date the task/MRI was acquired. For each task, any participant that had a PDS within 6 months of the task/MRI visit was included. As a result, the reported subject numbers for the PDS, as linked to the task and MRI, vary. However, we did have a single PDS measure acquired for all participants. Median PDS values were calculated from the observable measures on the PDS (e.g. hair growth or skin changes), measures that did not involve observation (e.g. whether the parent will discuss puberty with his/her child) were excluded. Unsurprisingly, differences in PDS were strikingly similar to the differences observed in age (see: Table S1). Exclusion criteria for both groups included the presence of seizure disorder, cerebral palsy, pediatric stroke, history of chemotherapy, sensorimotor handicaps, closed head injury, thyroid disorder, schizophrenia, bipolar disorder, current major depressive episode, fetal alcohol syndrome, severe vision impairments, Rett's syndrome, and an IQ below 70. Participants in the TD group were also excluded if diagnosed with attention-deficit hyperactivity disorder. Subjects taking prescribed stimulant medications completed medication washout prior to testing and scanning. Children performed tasks and completed MRI visits following a minimum of five half-life washouts, which ranged from 24 to 48 hours given the preparation. Participants on non-stimulant psychotropic medication (e.g. anxiolytics or anti-depressants) were excluded from this study.

Data collection procedures—ASD participants came in for a screening visit to determine if they qualified for the study. During this initial visit, informed written consent or assent was obtained from all participants and their parents, consistent with the Oregon Health & Science University institutional review board. Additionally, children completed the ADOS and the Wechsler Intelligence Scale for Children IV (WISC-IV; ⁷²) block design subtest while parents completed the SRS, ADI-R, and Developmental and Medical History surveys. Participants who qualified for the study came back for a second visit where they completed our Delay Discounting, Spatial Span, CPT, and Stop tasks. All participants also experienced a “mock scanner” to acclimate to the scanner environment and to train themselves to lie still during the procedure. Participants then came in for a third visit where they were scanned. At the fourth visit, participants completed our Face Identity Recognition, Facial Affect Matching, and Vocal Affect Recognition tasks.

Participants in the TD group were recruited from a partner study with similar protocol. During the initial screening visit, participants underwent a diagnostic evaluation based on the Kiddie-Schedule for Affective Disorders and Schizophrenia (KSADS) interview, as well as parent and teacher standardized ratings, which were reviewed by their research diagnostic team. TD participants completed their study visits and tasks in a similar timeline and were recruited for our study during their MRI visit. TD participants were then screened and enrolled in an additional visit in which they completed the Face Identity Recognition, Facial Affect Matching, and Vocal Affect Recognition tasks.

Most of the participants consented to a longitudinal study where they returned on an annual basis to be reassessed on these same tasks and were re-scanned. For this study, we used data from each participant's earliest time point for each completed task and MRI scan. Per task and scan, a t-test was conducted to test whether the cross-sectional ages were significantly different for that test. In all cases, ASD participants were significantly older than TD participants (all $p < 0.05$). We controlled for non-verbal intelligence, as measured by the

WISC block design, by ensuring that block design scores were not significantly different between the groups ($p = 0.285$). We also calculated and tested the difference in visit age for the ASD (mean years = 1.51, s.d. (years) = 1.36) and typical (mean years = 1.14, s.d. (years) = 1.17) samples selected. We found no significant group effects on average visit difference ($t(103) = 1.49$, $p = 0.14$).

Tasks

Measures derived from seven tasks were used as input features for the random forest. These seven tasks cover multiple levels of information processing, which may affect or be affected by the presence of an ASD diagnosis. Per measure, an independent samples, two-tailed, t -test was conducted to evaluate whether ASD and TD participants differed significantly. Table 3 lists each feature along with the t -statistic and p -value associated with the test. Because the random forest approach is robust against the presence of non-predictive features⁷³, our initial feature selection was inclusive. Despite this liberal inclusion, these non-predictive features did not contribute meaningfully to the classification model and thus did not affect results materially (supplementary materials).

Delay Discounting

The Delay Discounting task measures an individual's impulsivity by asking them to evaluate a reward's subjective value following a delay. The task design employed here has been described in detail previously^{74,75}. In short, this computerized task consisted of 91 questions and requested participants to choose between two hypothetical amounts of money, one smaller amount that would be available immediately, and one larger amount that would be available after a fluctuating delay (between 0 to 180 days). No actual money was obtained. We used 9 variables from this task in our RF model: the indifference score at 5 time points (7, 30, 90, or 180 days), the calculated area under the curve (AUC) based on these indifference scores, the proportion of variance explained between the scores and their timepoints, their k value (a measure of overall rate of discounting), and the natural log-transformation of these k values. Three validity criteria were applied⁷⁶: 1) an indifference point for a specific delay could not be greater than the preceding-delay indifference point by more than 20% (\$2); 2) the final (180 day) indifference point was required to be less than the first (0 day) indifference point, indicating evidence of variation in subjective value of rewards across delays; and 3) the 0-day indifference point was required to be at least 9.25. Lower values for the 0-day indifference point indicate that the child chose multiple times to have a smaller reward now over a larger reward now, suggesting misunderstanding or poor task engagement. Data that did not meet validity criteria were treated as missing in analyses.

Spatial Span

The Spatial Span task measures an individual's visuospatial working memory capacity. Our participants received a spatial span subtest identical to the computerized Cambridge Neuropsychological Test Battery (CANTAB; ⁷⁷). Briefly, this computerized task presents a series of 10 white boxes randomly placed on the screen, a subset of which would change color in a fixed order. Participants were instructed to watch for boxes that changed color and to keep track of their sequence. In the spatial forward task, participants were instructed to

click on the boxes in the same sequential order in which they were presented. In the spatial backward task, participants were instructed to click on the boxes in the reverse order in which they were presented. The tasks were counterbalanced, and every subject had the opportunity to practice before administration. At the beginning of both tasks, the numbers of squares that changed started at three and increased to nine, with two trials at each sequence length (a total of 24 trials for both tasks). The task discontinued when a child failed both trials at a sequence length. We used 8 measures from this task in our RF model: reaction time, accuracy, number completed, and span number correct for both the forward and backward tasks.

Stop Task

A tracking version of the Logan stop task was administered to all participants^{78,79}. The Stop Task is a dual go-stop task. The go portion of the task measures reaction time and variability of reaction time on a simple choice detection task; the stop portion measures speed at which the individual can interrupt a prepotent response (how much warning is needed). For this computerized task participants fixated on a small cross in the center of computer screen, which appeared for 500ms on each trial. For the “go trials” (75% of total trials), either a rainbow “X” or an “O” would appear on the screen for 1000ms. Participants then had 2000ms to indicate whether they saw an “X” or an “O” using a key press, after which the next trial would automatically start. The “stop trials” (25% of total trials) were identical except that an auditory tone was played briefly after the presentation of the visual cue. The timing of the tone was varied stochastically to maintain approximately 50% success at stopping. Participants were instructed to not respond with the key press if they heard the tone. Each participant performed 20 practice trials to ensure they understood the task, before completing eight 32 trial blocks of the task. We used 5 measures from this task in our RF forest model: accuracy of the X/O choice on “go-trials”, probability of successful stopping on the “stop-trials”, stop signal reaction time (computed as the difference between go RT and timing of the stop delay warning signal), mean reaction time on go-trials, and the standard deviation of reaction times during “go-trials”.

Continuous Performance Task

The Continuous Performance task was an identical-pairs version of the common CPT, which measures vigilance. For this computerized task participants viewed a series of four digit numbers (250ms per cue) and were instructed to press a button whenever they saw a pair of identical numbers back-to-back. The task consisted of three types of trials: 1) trials where the paired numbers were made of distinct digits called “stim trials”, 2) trials where paired numbers only differed by one digit called “catch trials” and 3) trials where the pair of numbers were identical (target trials). The task included a total of 300 stimuli and required about 10 minutes to complete. There were 20% target trials, 20% catch trials, and 60% “stim” or non-target trials. We used 6 measures from this task in our RF model: dprime (a measure of discriminability⁸⁰) per discrimination type (essentially, “hard” and “easy” discriminations), bias score for each discrimination type, and the natural log of bias per discrimination type.

Face Identity Recognition Task

The Face Identity Recognition Task was designed by the Center for Spoken Language Understanding (CSLU) at OHSU to measure facial processing skills. In this computerized identification task, for each of the 25 trials (inter-trial interval = 2s), participants were presented with a “target face” on the left side of the screen, a colored photograph of a human face presented in standardized poses with neutral facial expressions. At the same time participants were shown an additional four facial photographs on the right side of the screen (all photographs were selected from the Glasgow Unfamiliar Faces Database ⁸¹, see Fig 1B), one of which matched the target face. Participants were asked to select the target face out of the lineup by touching the screen with stylus pen. Reaction times were calculated from the moment the trial began to the participant’s response; however, participants were not told they were being timed or instructed to complete the task as quickly as possible. Each participant was allowed five practice trials to ensure they understood the task. We used 2 measures from this task in our RF model which included the number of correct responses and the median reaction time for all trials.

Facial Affect Matching Task

The Facial Affect Matching Task and was designed by the CSLU at OHSU to measure affect discrimination skills using facial expressions. In this computerized task, for each of the 25 trials (inter-trial interval = 2s), participants were presented with a “target emotion”, a colored photograph of a human face expressing one of six possible emotions (happiness, sadness, surprise, disgust, fear or anger), on the left side of the screen. At the same time participants were shown an additional four facial photographs on the right side of the screen (all photographs were selected from the NimStim set of facial expressions ⁸², see Fig 1A), one of which matched the target emotion. Participants were asked to select the target emotion out of the lineup by touching the screen with stylus pen. Reaction times were calculated from the moment the trial began to the participant’s response; however, participants were not told they were being timed or instructed to complete the task as quickly as possible. Each participant was allowed five practice trials to ensure they understood the task. We used 2 measures from this task in our RF model which included the number of correct responses and the median reaction time for all trials.

Vocal Affect Recognition

The Affect Matching Task was designed by the CSLU g at OHSU to measure affect discrimination skills using auditory cues. In this computerized task, for each of the 24 trials (inter-trial interval = 2s), participants were presented with an audio recording of an actor reading neutral phrases (e.g., “we leave tomorrow”) but expressing one of four possible emotions (happiness, sadness, fear or anger) during the reading. Participants were asked to identify what type of emotion the actor was expressing by selecting one of four black and white drawings of facial expressions, each depicting one of the 4 basic emotions (see Fig 1C). Reaction times were calculated from the moment the trial began to the participant’s response; however, participants were not told they were being timed or instructed to complete the task as quickly as possible. Each participant was allowed four practice trials to

ensure they understood the task. We used 2 measures from this task in our RF model which included the number of correct responses and the median reaction time for all trials.

MRI scans

Data acquisition

Participants were scanned in a 3.0 T Siemens Magnetom Tim Trio scanner (Siemens Medical Solutions, Erlangen, Germany) with a 12 channel head coil at the Advanced Imaging Research center at Oregon Health and Science University. One T1 weighted structural image (TR = 2300ms, TE = 3.58ms, orientation = sagittal, FOV = 256×256 matrix, voxel resolution = 1mm×1mm×1.1mm slice thickness), and one T2-weighted structural image (TR = 3200ms, TE = 30ms, flip angle = 90° FOV = 240mm, slice thickness = 1mm, in-plane resolution = 1 × 1mm) was acquired for each participant. Functional imaging was performed using blood oxygenated level-dependent (BOLD) contrast sensitive gradient echo-planar sequence (TR = 2500ms, TE = 30ms, flip angle = 90°, in-plane resolution 3.8×3.8mm, slice thickness = 3.8mm, 36 slices). For fMRI data acquisition, there were three 5-minute rest scans where participants were asked to relax, lie perfectly still and fixate on a black cross in the center of a white display.

General preprocessing

All functional images went through identical Human Connectome Project preprocessing pipelines as described previously⁸³ in order to reduce artifacts. These pipelines included 1) PreFreeSurfer, which corrects for MR gradient and bias field distortions, performs T1w and T2w image alignment, and registers structural volume to MNI space; 2) FreeSurfer⁸⁴, which segments volumes into predefined cortical and subcortical regions, reconstructs white and pial surfaces, and aligns images to a standard surface template (FreeSurfer's fsaverage); 3) PostFreeSurfer, which converts data to NIFTI and GIFTI formats, down sampled from a 164k to a 32k vertices surface space, applies surface registration to a Conte69 template, and generates a final brain mask. 4) fMRIVolume, which removes spatial distortions, performs motion correction, aligns fMRI data to the subject's structural data, normalizes data to a global mean, and masks the data using the final brain mask, and 5) fMRISurface which maps the volume time series to a standard CIFTI grayordinate space.

Functional connectivity processing

All resting state functional connectivity MRI data received additional preprocessing that have been widely used in the imaging literature⁸⁵ to account for signals from non-neuronal processes. These steps included: 1) removal of a central spike caused by MR signal offset, 2) slice timing correction 3) correction for head movement between and across runs, 4) intensity normalization to a whole brain mode value of 1000, 5) temporal band-pass filtering ($.009\text{Hz} < f < .08\text{ Hz}$), 6) regression of nuisance variables: 36 motion related parameters, and three averaged signal timecourses from the grayordinates, white matter, and cerebrospinal fluid (CSF). Additionally, because previous research has indicated that minor head movement can result in changes in MRI signal, we performed motion-targeted "scrubbing" on all rs-fcMRI data⁸⁵. These steps included censoring any volumes with frame

displacement (FD) > .2mm, and the elimination of any run with less than a total of two and a half minutes of data.

Correlation matrix generation

All timecourses and correlations were derived from a set of 333 Regions of Interest (ROIs) produced from a published data-driven parcellation scheme (Figure 4)⁸⁶, and a set of 19 subcortical areas parcellated by FreeSurfer during preprocessing. The resulting parcellations set comprised 352 ROIs. Correlations between ROIs were calculated using Pearson product-moment coefficient between each pair of ROIs over the extracted time series following preprocessing and motion censoring. We created a correlation matrix for each participant and then created group correlation matrices by averaging individual matrices across groups and subgroups.

Data Analysis

Exploratory Data Analysis—Prior to construction of the RF model, we measured the quantity of missing data. Machine-learning model performance can be greatly affected by missing data. Therefore, we excluded any measures and participants that were missing more than 15 percent of data. The remaining missing data is imputed separately for the training and test datasets using the random forest algorithm below, where the missing data's column is the outcome measure and the remaining variables are used as predictors. Prior to our exploratory data analysis we had a total of 143 subjects (73 ASD, 70 TD) with partially completed data, after eliminating subjects with more than 15 percent missing data we finalized our subject list down to 105 (47 ASD, 58 TD). In the final dataset, less than 3 percent of all possible data was missing. An inspection of the missing data was unable to find any patterns that distinguish the missing ASD data from the remaining cases.

Random Forest classification

General algorithm: The RF algorithm constructs a series of decision trees. Per tree, a bootstrapped dataset is generated from a subset of the training data and a subset of features are randomly used to predict group classification or outcome measure in the case of imputation. The Gini impurity is used as the cost function to determine the optimal tree for classification and the mean square error is used as the cost function to determine the optimal tree for regression. Finally, a testing dataset comprising participants that were excluded from the training dataset is used to evaluate classification model performance. We implemented this algorithm via in-house custom-built MATLAB programs that used the MATLAB TreeBagger class. 1000 trees were used for the classification model and 20 trees were used for the surrogate imputation. Missing data was imputed separately for training and testing datasets. For classification, 1000 iterations of the RF algorithm were run to assess the performance of the RF models. Per iteration, 60 percent of participants formed the training dataset and the remaining 40 percent formed the testing dataset.

Optimization and validation: Distributions of overall, ASD, and control accuracy were constructed from the 1000 iterations and compared against a distribution of 1000 null-models. Per null-model, the group assignments are randomly permuted and the RF procedure above is performed on the permuted data. If the RF classification models are

significantly better than the null models, then we interpret the RF models as valid for predicting a given outcome measure. An independent samples t-test was used to evaluate the significance of the RF model performance against the null model performance based on the models' accuracy, specificity, and sensitivity rates.

Community detection: Since each tree has different terminal branches, the RF algorithm may identify different paths for participants with the same diagnosis. Therefore, validated models can be further analyzed to identify putative subgroups that reflect the same diagnosis but perhaps different etiologies. Briefly, the RF algorithm produces a proximity matrix, where the rows and columns reflect the participants and each cell represents the proportion of times, across all trees and forests, a given pair of participants ended in the same terminal branch. For the classification model, the Infomap algorithm (Rosvall, 2007) was used to identify putative subgroups from the proximity matrix for participants with an ASD and from the proximity matrix for control participants. Because we have no basis for determining what constitutes an edge, an iterative procedure was used⁸⁷, where we identified a consensus set of community assignments across all possible thresholds.

Radar plot visualization: Task measures were then examined via radar plots to identify features that distinguish putative subgroups. Since plotting all measures may obscure differences between the groups, visualized task measures were chosen via statistical testing. For the ASD and the TD samples separately, one-way ANOVAs, with subgroup as the factor and each subgroup a level, were conducted for each task measure. Significant ($p < 0.05$) task measures were chosen for visualization. Individual task measures were converted to percentiles and visualized by task.

Functional connectivity cluster analysis: We used a chi-square approach to identify potential differences between subgroups within or between functional systems, as opposed to individual functional connections⁸⁸. Briefly, three sets of mass univariate tests were conducted for all Fisher-Z transformed functional connections: a set of one-way ANOVA using ASD subgroup as the factor, a set of one-way ANOVAs using control subgroup as the factor, and a set of t-tests between ASD and control groups. Per set, a matrix of coefficients are extracted and binarized to an uncorrected $p < 0.05$ threshold. This binary matrix is then divided into modules based on the published community structure⁸⁹ which reflects groups of within system (e.g. connections within the default mode system) and between system (e.g. connections between the default mode system and the visual system) functional connections. The subcortical parcellation was defined as its own system for this analysis because of prior research suggesting differences between cortical and subcortical connectivity⁵⁵. A ratio of expected significant to non-significant functional connections (i.e. the expected ratio) is calculated by dividing the total number of significant connections by the total number of all connections. Per module, the number of expected significant and non-significant functional connections is determined by multiplying the expected ratio by the total number of functional connections within the module. A chi-squared statistic is then calculated using the observed and expected ratio of significant connections. Permutation tests were conducted for all functional connections across the 352 ROIs to calculate the p value per module, and

evaluate whether the observed clustering is greater than what would be observed by random chance.

Results

Random Forest Classification results

Random forest successfully classified individuals as having ASD or not—RF model accuracy is shown in Figure 2A. Applying the RF algorithm on behavioral data from 7 different tasks (34 variables) achieved an overall classification accuracy of 73% ($M = .727$, $SD = .087$) and an independent sample t-test revealed that the RF model was significantly more accurate than the permutation accuracy measure of 51% [$M = 50.9$, $SD = .103$; $t(1998) = 51.325$, $p < .001$]. The RF model had a sensitivity of 63% ($M = .631$, $SD = .153$) when classifying ASD subjects, the ability to correctly identify true positives, and an independent sample t-test revealed that the model's sensitivity was significantly higher compared to the permutation sensitivity of 44%. [$M = .441$, $SD = .166$; $t(1998) = 26.643$, $p < .001$]. The RF model also had a specificity of 81% ($M = .807$, $SD = .153$) when classifying control participants, the ability to correctly identify true negatives, and an independent sample t-test revealed that this was significantly more accurate compared to the permutation specificity of 56%. [$M = .564$, $SD = .153$; $t(1998) = 40.501$, $p < .001$]. Taken together, these findings show that the RF model identified patterns in the cognitive data that stratified individuals with an ASD diagnosis from individuals without. (Note: Due to confound age and gender factors, a secondary RF analysis was performed on the behavioral data, controlling for both factors. Despite the large confounds, the RF analysis accurately classified ASD from control participants greater than chance. This analysis is discussed in supplemental materials).

Proximity matrices from random forest model suggest subgroups in ASD and Control samples—We next applied community detection to the proximity matrices generated through the random forest modeling. The community detection algorithm identified three putative ASD subgroups and four putative control subgroups (Figure 2B). For children with an ASD diagnosis, the largest subgroup comprised 25 individuals, while the other two subgroups numbered 13 and 9 children respectively. For children without an ASD diagnosis, the largest subgroup comprised 39 individuals; three other subgroups were evenly split with five, five, and three children respectively. Six controls were not identified as part of any community, which were placed into a fifth “unspecified” subgroup. To characterize these subgroups, we first examined whether accuracy of classification varied between subgroups, and then examined variation in the task measures between the subgroups.

ASD subgroups differed in terms of classification accuracy: We next compared the classification accuracy of individuals within each ASD subgroup to see if specific subgroups may have differentially affected RF model performance (Figure 2C). It also allowed us to validate that these subgroups were indeed systematically different from one another based on the cognitive data used in the RF model.

Because we constructed multiple RFs, each subject was included in the test dataset a large number of times, therefore we can calculate the rate of accurate classification per subject. A

one-way between subjects ANOVA was conducted to compare the rate of classification accuracy between the 3 ASD subgroups identified by community detection. There was no significant difference between the groups [$F(2, 44) = 1.859, p = .168$]. An independent sample t-test was conducted to see if subgroup classification accuracy significantly differed from chance (.5) using a Bonferroni adjusted alpha level of .0167 per test (.05/3). Subgroup 1 was significantly better at classification than chance [$M = .726, SD = .367; t(24) = 3.0732, p = .005$] but subgroups 2 [$M = .607, SD = .383; t(12) = 1.01, p = .334$] and subgroup 3 [$M = .443, SD = .431; t(8) = -.399, p = .701$] were not.

These results suggest that there may be differences in our subgroups that are important for distinguishing ASD from TD. This difference is subtle, because effects of subgroup on accuracy are small and could largely be driven by the small sample size in subgroups 2 and 3. However, variation in classification accuracy may reflect differences in cognitive profiles. Subjects in subgroup 3 had a classification accuracy of only 44%, which may indicate that these individuals had cognitive scores more similar to our control group than our ASD group, while subgroup 1 had a classification accuracy of nearly 73% suggesting that their cognitive scores may be far different from both our control group, and ASD subgroup 3.

Control subgroups differed in terms of classification accuracy: We also compared the classification accuracy of individuals within each control subgroup to again see if specific subgroups were differentially affecting our RF model's performance (Figure 2D).

A one-way between subjects ANOVA was conducted to compare classification accuracy for each of the 4 control subgroups plus the controls that were lumped into a fifth subgroup, identified by community detection. There was a significant effect of subgroups on classification accuracy [$F(4, 53) = 24.018, p < .001$]. Post-hoc comparisons using an independent-sample t-test indicated that the classification accuracy for subgroup 5 ($M = .120, SD = .086$) was significantly worse (using a Bonferroni adjusted alpha level of .006 per test) than subgroup 1 [$M = .922, SD = .137; t(43) = -13.871, p < .001$], subgroup 2 [$M = .804, SD = .422; t(9) = -3.910, p = .004$], and subgroup 4 [$M = .995, SD = .0089; t(7) = -16.903, p < .001$], but not subgroup 3 [$M = .636, SD = .362; t(9) = -3.411, p = .008$]. Additionally, an independent sample t-test was conducted to see if subgroup classification accuracy significantly differed from chance (.5) using the Bonferroni adjusted alpha level of .006 per test. Participants in subgroups 1 [$t(38) = 19.276, p < .001$] and 4 [$t(2) = 96.00, p < .001$] were classified as controls significantly more than chance, while participants in subgroup 5 [$t(5) = -10.773, p < .001$] were classified as controls significantly less than chance.

Community Detection identified these subgroups in ASD and Control samples who differed in behavioral tasks and classification accuracy: To test whether ASD subgroups may reflect quantitative variation in autism symptom severity, we examined whether identified ASD subgroups varied by Social Responsiveness Scale (SRS). A one-way ANOVA revealed no significant differences between the subgroups on SRS (Figure 2E; $F(2, 44) = 0.006, p = 0.994$), suggesting that ASD subgroups had similar autism severity but varied in other ways. Because normal variation in cognitive profiles may affect the manifestation of a developmental disorder⁷¹, we then examined the variation in task performance for ASD (Figure 3; left) and control (Figure 3; right) subgroups. For control

subgroups, the fourth subgroup was not examined due to the small sample size and the fifth subgroup was not examined because it represented “unspecified” subjects. A series of subgroupXtask measure repeated measures ANOVA were performed to assess whether we should examine task performance between specific subgroups. The ASD subgroups ($F(66,1056) = 7.65, p = 7.5 \times 10^{-54}$), control subgroups ($F(66,1452) = 2.19, p = 2.4 \times 10^{-7}$), and accurately identified subgroups ($F(33,1716) = 10.64, p = 3.3 \times 10^{-49}$) showed significant differences across task, indicating that identified subgroups varied by task measure. Post-hoc one-way ANOVAs identified 11 significant different features for control subgroups ($F(2, 46) > 3.29, p < 0.0462$) and 16 significant different features for ASD subgroups ($F(2, 44) > 3.45, p < 0.0405$). For both ASD and control subgroups, similar relative cognitive profiles were observed. The largest subgroup in both cohorts performed best on stop and continuous performance tasks. The second largest subgroup in both cohorts had the smallest spatial span, and the highest accuracy and longest reaction times for the facial and affect processing tasks. The third subgroup in both cohorts was characterized by highest spatial span, but lowest accuracy and shortest reaction time for the face processing tasks. Participants who show a combination of low accuracy and short reaction time may be showing a speed accuracy trade-off⁹⁰, where individual participants are making quicker responses at a cost of more accurate responses. For the most part, delayed discounting did not differentiate the subgroups, which is unsurprising, because evidence is mixed whether delayed discounting varies by ASD or ASD subgroups. A prior study suggests that ASD and control subgroups discount monetary rewards similarly²⁴; the relationship between discounting and time varies by ASD subgroup, which is consistent with findings from a separate study where some ASD participants may discount monetary rewards more steeply than controls²⁵. The similar cognitive profiles observed between controls and ASD subgroups suggests that normal variation in cognitive profiles may impact how ASD manifests in individuals.

Functional Connectivity Results

Functional connectivity differences between ASDs and Controls—To test our hypothesis that our ASD and controls groups differed in terms of resting-state functional connections between, and within, different functional systems, we used the chi-squared approach described earlier. The Gordon parcellation plus 19 subcortical regions were used to define the modules (Figure 4). We conducted the analysis on the 26 ASD subjects and 42 control subjects with satisfactory fMRI data (Figure 5A). The chi-squared analysis revealed significant clustering effects between the cingulo-opercular system and the default mode system ($\chi^2 = 48.86, p = .0002$), the somato-motor hand system and the default mode system ($\chi^2 = 12.81, p = .0016$), the visual system and default mode system ($\chi^2 = 11.74, p = .001$), and between the subcortical system and the dorsal attention system ($\chi^2 = 35.05, p = .0024$). It also revealed significant clustering effects within the cingulo-opercular ($\chi^2 = 259.36, p = .0002$), the default mode system ($\chi^2 = 11.66, p = .0002$), and the visual system ($\chi^2 = 35.05, p = .0002$). These findings are consistent with prior reports of rsfMRI differences between TD and ASD samples (see: Discussion).

Subgroup differences within ASD and control samples—Because ASD subgroups differed in classification accuracy with respect to chance (Figure 2C), we also tested whether variance between each of the ASD subgroups and the large control subgroup differed in

terms of resting-state functional connections between, and within, different function systems, using the chi-squared analysis. Unfortunately, due to the MRI ‘scrubbing’ procedure, we did not have sufficient data in the other control subgroups to include them in this analysis. We conducted a one-way ANOVA with four groups on 57 subjects: the 31 subjects from Control subgroup one, 12 subjects from ASD subgroup 1, 8 subjects from ASD subgroup two, and 6 subjects from ASD subgroup three who had satisfactory fMRI data. We again used a permutation test to determine each system’s expected ratio and compared this to the observed ratio using the chi-squared analysis (Figure 5B). We used the estimated marginal means from the ANOVA to visualize which subgroups drove significant clustering (Figure 5C). This test revealed significant increases in connectivity for ASD subgroup 1, relative to all other subgroups, between the cingulo-parietal system and the auditory system ($\chi^2 = 12.06$, $p = .0014$). Significant increases in ASD subgroup 2 and 3 between the cingulo-opercular system and the default system ($\chi^2 = 24.01$, $p = .0002$), and between the dorsal attention system and the somato-motor hand system ($\chi^2 = 15.37$, $p = .0006$). Significant increases in ASD subgroup 1 and 2 connectivity between the salience system and the visual system ($\chi^2 = 11.36$, $p = .0016$). Significant increases in control connectivity were observed within the default system ($\chi^2 = 22.36$, $p = .0010$) and between the dorsal attention system and the subcortical system ($\chi^2 = 11.85$, $p = .002$). Connectivity between the default system and the somato-motor hand system ($\chi^2 = 28.85$, $p = .0002$) showed mixed results, with ASD subgroups deviating from controls. The estimated marginal means for these tests are summarized in Table 3.

These differences overlapped substantially with the differences observed between ASD and controls (Figure 5A), suggesting that normal variation in mechanisms that are also affected by ASD may cause variation in how ASD may manifest^{1,91}. These findings should be interpreted cautiously, however, because these data are not predictive of diagnosis.

Discussion

Accuracy of the Random Forest model

Link our results to prior findings using machine learning ASD classification—

Using a RF model, ASD and control participants were accurately classified 73 percent of the time using a comprehensive battery of cognitive tasks often identified as affected by an ASD diagnosis. Despite differences in age between samples, it is unlikely that the accurate classification was driven by age for two primary reasons. First, task measures important for classification did not show strong correlations with age (see: supplemental materials for discussion); when corrected for multiple comparisons, no relationships between gender and task performance are observed. Second, we performed a second RF model controlling for age and gender across all features, which continued to perform above chance (see: supplemental materials for discussion).

Higher performance has been reported for behavior when constructing a model using visual face scanning (88.1%;⁹²) or goal-oriented reach (96.7%;⁹³) measures. However, high classification accuracy may be a function of validation strategies or sample size. Liu et al used a leave-one-outcross validation (LOOCV) strategy, which improves classification accuracy within a test dataset, but may reduce the generalizability of the model to other

datasets. Crippa et al. also used a LOOCV validation strategy, and were also limited in sample size. Machine learning approaches using imaging data have shown that validation accuracy decreases as the sample size increases, suggesting that these small sample sizes may be overfitting the data ^{65,68}.

Recent classification studies incorporating brain measures have shown comparable results to our initial classification and further suggest that heterogeneity of clinically relevant ASD subgroups may limit high classification accuracy. Duchesnay et al. found that PET imaging could be used to predict ASD with 88% accuracy in a sample of 26 participants ⁶¹. Murdaugh et al. used the intra-DEF connectivity to predict ASD with 96% accuracy in a sample of 27 participants ⁶². Wang et al., using whole-brain functional connectivity, correctly predicted ASD with 83% accuracy in a sample of 58 participants ⁶³. Jamal et al. used EEG activity during task switching to predict ASD with 95% accuracy in a sample of 24 participants ⁶⁴. Using large data consortiums like the Autism Brain Imaging Data Exchange (ABIDE), recent classification studies have developed and tested models using datasets with over 100 participants. Collectively, these large-sample studies demonstrate performance accuracy from 59% to 70% when testing untrained data ^{65–68,94}. Our data highlights the importance of considering heterogeneity for such tests.

Extension of prior Machine Learning studies

Individual classification results and their relation to subgroups—Our RF approach extends prior studies by identifying putative subgroups from a validated ASD classification model. Specifically, we identified three ASD and four control putative subgroups, with a fifth group of isolated subjects. To further characterize these subgroups, we examined whether subgroups were stratified via classification accuracy. Because of our extremely stringent inclusion criteria, we are extremely confident that all ASD subjects indeed have an ASD, therefore ASD subgroups that contain misclassified individuals may represent clinically important subgroups that our initial RF model failed to capture. Control subgroups that contain misclassified individuals may represent subgroups that our initial RF model confused for ASD individuals. We found that the largest subgroup for ASD and the largest and smallest subgroup for controls were significantly more accurate than chance. Other ASD and control subgroups were not, and the distinction in classification accuracy may reflect the heterogeneity within the disorder. In an earlier study, ASD participants were sub-grouped on the basis of symptom severity, verbal IQ, and age, which caused classification rates to increase by as much as 10% ⁶⁵. On the other hand, the fact that control subgroups also showed misclassification suggests that variation in such skills may represent the existence of broad cognitive subgroups that are independent of diagnosis, whose variation may impact the presentation of ASD symptoms ¹. Prior work by Fair et al. has shown similar heterogeneity in both TD and ADHD children; as with Katuwal, taking into account this heterogeneity improved diagnostic accuracy ⁷¹.

ASD subgroups are not associated with variance in symptom severity—It is controversial whether clinical subgroups even exist in ASD. Recently, it has been suggested that ASD represents the tail end of a continuous distribution of social abilities. Categorically distinct subtypes are either artificial constructs ⁹⁵ or unknown ¹. Categorically distinct

subtypes may be difficult to discover due to the heterogeneity present within the typical population⁷¹ as well as the heterogeneity in genetic causes of ASD¹. According to Constantino et al., such genetic subtypes may interact with the environment of the individual, leading to varying manifestations of ASD. Findings that the trajectories of adaptive functioning and autism symptom severity are distinct from one another^{14,91} further suggests a dissociation between adaptive functioning and symptom burden.

Therefore, our subgroups may reflect the variation in autism symptom severity or in cognitive mechanisms that may impact ASD profiles, independent of severity. To test this hypothesis, we examined whether our ASD subgroups varied by autism symptom severity, as measured by the SRS⁹⁶ and the ADOS (Supplemental Analysis 3). We found that our subgroups did not differ on the SRS or the ADOS, suggesting that autism symptom severity was similar across the three subgroups, despite differences in classification accuracy. Because we are confident in the ASD diagnosis, we suspect that the variation between these three subgroups reflects typical variation in cognitive mechanisms, which may be independent of autism symptom severity but influence ASD presentation^{14,91}. Identification of such subgroups may be critical for the development of personalized treatment approaches in future studies and has the potential for improving ASD diagnosis and long term outcomes¹. Future studies could better characterize putatively identified subgroups by examining how subgroups may differ on measures of adaptive functioning, or examining whether the subgroups may be characterized by a set of measured ASD symptoms. Critically, future studies should also seek to assess the stability of identified subgroups using longitudinal data.

Describe identified subgroups

To further characterize the identified subgroups, we examine how the subgroups differed on the tasks incorporated into the model. With such an analysis, we can compare our results to prior research that has identified subgroups in independent datasets using similar tasks⁷¹. Replication of similar subgroups would suggest these subgroups may be meaningful. However, because the data from these tasks were used to construct the model, an independent set of measures is necessary to establish the validity of the identified subgroups. Therefore, we also examined differences in functional brain organization in a subset of participants, to see whether differences in functional brain organization between the subgroups reflects the effect of an ASD diagnosis on functional brain organization.

Differences in behavior and how that compares to previous literature—Due to fragmentation and limited sample size, we examined variation in task performance between the three ASD subgroups and between the largest three control subgroups only. Similar to prior research, subgroup differences were largely similar, independent of clinical diagnosis. Per sample, the largest subgroups performed best on CPT and stop tasks, and worst on face processing tasks. The second largest subgroups had the smallest spatial span and were slower but more accurate on the face processing tasks. The third largest subgroups had the largest spatial span and were faster, but less accurate, on the face processing tasks. The distinctions between these subgroups are consistent with prior research, which characterized heterogeneity in typical and ADHD samples and found multiple subgroups characterized by

either a small spatial span, slow RT, and high information processing, or high spatial span, fast RT and low information processing⁷¹. Taken together, these findings suggest that clinical heterogeneity may emerge from normal variation in cognitive profiles, and are consistent with a recent study showing that clinical heterogeneity within ASD may be driven by normative development⁹⁷. Our study here extends the prior findings to ASD and establishes a predictive model, which provides some clinical validity to the identified subgroups.

Our finding that the differences between subgroups were similar in both ASD and TD samples may appear inconsistent with prior studies that show an effect of ASD on the relationship between cognitive measures and task performance^{12,27,37,38}. However, differences in diagnostic criteria may explain some of the apparent contradiction here. Our study used a team of experts to confirm ASD diagnosis per individual, whereas these prior studies often used only a DSM diagnosis plus one or two instruments that assess autism symptom severity (e.g. the ADOS and/or ADIR). The inconsistency in findings may be interpreted as further evidence of heterogeneity within ASD. Differences in cognitive profiles across individuals with ASD could explain the variation in attention, working memory, and face processing. In addition, prior work suggests that cognitive subtypes within ASD may be similar to cognitive subtypes found in typical populations⁹⁸.

Differences in fMRI data and validation of subgroups how that compares to previous literature—To provide further validation of the subgroups, we examined whether significant differences in the functional organization of the brain between subgroups overlapped with significant effects of ASD on functional brain organization. Since this data was never used in the RF model, variation that overlaps with differences between ASD and typical children may reflect clinically or etiologically important distinctions between subgroups. Because we did not observe differences in symptom severity between subgroups, the findings above are more likely to reflect typical variation in neural mechanisms underlying cognitive performance, as opposed to manifestations of ASD symptoms.

Differences between children with and without ASD are consistent with prior studies but also show some novel findings. Children with an ASD have shown altered visual system responses to stacks of oriented lines⁹⁹, and at rest they've exhibited altered DEF functional connectivity⁵², but not altered cingulo-opercular connectivity⁵⁹. Between system differences have been less studied in ASD, however, sub-cortical cortical connectivity has been shown to be altered^{55,57} as well as the dorsal attention network organization⁵¹, which is consistent with altered connections between subcortical and dorsal attention networks. However, differences between the DEF and visual, somatomotor, and cingulo-opercular systems have not been documented. The differences found between somatomotor and DEF may be consistent with findings of altered motor system function in ASD¹⁰⁰, while differences between DEF and cingulo-opercular systems may be consistent with altered rich-club organization⁵¹.

We would like to emphasize that the ANOVA chi-squared analysis may be underpowered⁸⁸ and, though enticing, is not definitive. Nevertheless, the subgroup chi-squared ANOVAs hint that the identified subgroups may reflect differences in both mechanisms relevant to an ASD

diagnosis, and mechanisms that reflect variation across the subgroups. Four of the seven connectivity modules significantly affected by an ASD diagnosis showed variation in the ANOVA analysis: connectivity within the DEF; connectivity between the DEF and cingulo-opercular systems, between the DEF and somatomotor systems, and between the dorsal attention and subcortical systems. We also found significant variation in the ANOVA chi-squared analysis from the ASD and typical comparisons. Like with behavioral measures in children with and without ADHD, it is possible that variation within the ASD subgroups identified here may actually be “nested” within the normal variation found in brain networks across typical children ^{1,71,91}.

The correspondence between the subgroups and the connectivity profiles are intriguing, and hint that the first ASD subgroup may have altered visual processing mechanisms, the third ASD subgroup may have altered attention mechanisms, and the second ASD subgroup may have both. Speculatively, the first ASD subgroup shows the best ASD performance on both stop and CPT tasks, just as individuals in the first control subgroup performs better than the other control subgroups. Inter-system connectivity between the default mode and task control and attention systems (i.e. CIP and DOA) are control-like in the first ASD subgroup, as well as connectivity between attention and motor systems. As discussed extensively in the introduction, such variation is consistent with the literature and may reflect typical heterogeneity variability related to the presentation of ASD. The third ASD subgroup shows the worst performance on facial and affect tasks of the three ASD subgroups; the first control group performs worse on the same tasks compared to the other control subgroups (Figure 3; right). Such tasks would involve visual processing, and the chi-squared comparison reveals that the third ASD and first control subgroups have similar visual system connectivity. Variation in facial task performance may be implicated in some children with autism ³⁹, but not others ³⁷. It will be interesting to see whether future studies identify similar variation in system-level connectivity between ASD subgroups, and whether these groups are stable over time. In addition, future studies with larger sample sizes may be able to uncover additional or more refined sub-populations within the disorder.

Effects of demographics on RF model performance and subgroup affiliation

Due to the age and gender differences between our ASD and TD samples, we wanted to test whether the typical variation affecting ASD subgroups may reflect differences in demographic variables. We conducted six supplemental analyses (see: Supplemental Materials) to address this question. The analyses detailed extensively in Supplemental Materials are alluded to here. Specifically, we evaluated the effect of age and gender on the behavioral measures (Supplemental Analysis 1), performed the RF classification on behavioral measures when controlling for age and gender (Supplemental Analysis 2), examined the effect of ASD subgroup on ADOS symptom scores (Supplemental Analysis 3), tested whether subgroup affiliation affected age, IQ, or gender (Supplemental Analyses 4 and 5), and measured how much each behavioral measure improved RF classification (Supplemental Analysis 6).

The results from the supplemental RF were concerning, and hinted that controlling for age and gender may have, in fact, biased the analysis in unintended ways. There is some

literature¹⁰¹ that suggests such biases may occur when the differences in groups might differ by the controlling variables, but the features important for classification (i.e. in this case the behavioral measures) are not associated with those variables (i.e., here, age or gender). When we compared the association between age/gender and behavioral measures (Supplemental Analysis 1; Figure S1) to the behavioral measure importance (Supplemental Analysis 6; Figure S8), we found that only a few variables were associated with age or gender; the most important behavioral features showed no association with either demographic variable. After conducting the RF analysis we found several sub-groups that differed primarily by age and gender. Such findings were minimal in the main analysis. The findings provide important context for the primary findings, and highlight the importance of first examining the relationship between nuisance variables and input features. If no associations between input features and regressors are found, but regressors are associated with the outcome variable, then such regression may bias subsequent models in unintended ways. Similar concerns have been found when using parametric tests like analysis of covariance (ANCOVA) in psychiatric research¹⁰¹. Nonetheless, several considerations arise from these supplementary analyses.

When controlling for age and gender, the supplemental RF (Supplemental Analysis 2) showed a reduction in classification accuracy from 73 to 62 percent (Figure S2A). Nevertheless, the RF model remained significantly above chance for both ASD (64 percent) and TD (60 percent) individuals. Notably, the drop in model performance was driven entirely by the TD group, where performance dropped over 20 percent. Inspection of the subgroups shows that the second TD subgroup was more similar to the ASD subgroups; accuracy for the second TD subgroup was almost zero, suggesting that they were all being classified with ASD (Figure S2D), and therefore ASD classification may be driven by typical heterogeneity. Such an interpretation is consistent with the findings from the original RF. Additionally, We found little evidence that ASD subgroups varied by either SRS (Figure S2E and 2E) or ADOS (Figure S4; Supplemental Analysis 3) measures, which further indicates that ASD subgroups vary by typical heterogeneity and not autism symptom severity. However, it is unclear whether typical heterogeneity reflects demographic variables like age and gender or more cognitive variables like IQ or general task performance (Figure S3). Therefore further analyses investigated what aspects of typical heterogeneity affected subgroup affiliation.

The original RF model may be driven by cognitive profile, while the supplemental RF model may be driven more by demographics

Surprisingly, as noted above, the supplemental RF analysis identified subgroups that varied more by age and gender than the original RF (Supplemental Analysis 4). When we examined variation between control subgroups, we found that the supplemental RF subgroups (Figure S5D and Figure S6D) were split by gender, while the original subgroups showed no demographic differences (Figure S6B and S6D). The ASD supplemental subgroups varied by age (Figure S5C), gender, and IQ (Figure S6C), while the ASD original subgroups varied by age (Figure S5A) and IQ (Figure S6A). The variation in age and IQ differed between the supplemental and original analysis. The most accurately classified ASD subgroup in the original analysis was closest in terms of age and IQ to the control

subgroups, while the least accurately classified ASD subgroup in the supplemental analysis was most similar to the control subgroups. Because such demographic differences between accurately classified subgroups may explain the RF classification, we were interested in whether accurately classified ASD subgroups differed from accurately classified control subgroups. Since gender did not vary in the original analysis between ASD subgroups and between control subgroups, we focused on age and IQ variables. We found that age and IQ varied more in the supplemental than in the original analysis, however, IQ was numerically lower in the original ASD subgroup when compared to the control subgroups. Taken together, the findings suggest that the original RF was driven by variation in typical cognitive profiles, whereas the supplemental RF may be affected by variation in gender and age.

Limitations and Future Directions

The current study developed a novel approach for identifying and characterizing putative clinical subgroups and applied the approach to the question of ASD. Putative ASD subgroups identified from a diagnostically validated RF model are more likely to reflect properties relevant to clinical diagnosis than simple examination of the similarity between subjects. Therefore, this approach extends prior work that seeks to cluster and identify clinically important subgroups⁷¹. However, we want to emphasize that the identified subgroups in this sample are putative. To maintain a sensible scope and limit overfitting, we limited the model features to cognitive measures covering executive function and face processing, which do not cover every implicated domain of ASD⁶.

To provide more support for the validity of these subgroups, we compared the subgroups using rs-fcMRI data. While we had sufficient power to examine the effect of ASD on resting state data⁸⁸, we were underpowered when examining differences between subgroups. Therefore, while the rs-fcMRI analysis provides some evidence that speaks to the validity of the subgroups, the analysis does not fully validate them. In addition, our quality control procedure did not identify enough samples in the 2nd, 3rd, and 4th control subgroups to include them in the rs-fcMRI comparison. Such a comparison would help ascertain how the heterogeneity within and between control and ASD subgroups on behavior relates to rs-fcMRI patterns.

Prior identification of subgroups have relied on calculating the distance between each individual directly¹⁰². Unfortunately, such a method has two primary limitations. First, the diagnosis (or question) of interest is not tied directly to the distance measure. By tying the distance measure directly to diagnosis we can be more confident that the identified subgroups are relevant to the clinical question. Second, the proximity method employed by the RF approach implicitly handles missing data. Using the same dataset, we could not calculate a correlation matrix because we do not have all the data, and excluding participants by missing data would bias the subgroup identification. Nevertheless, we hope future studies will compare our approach with other detection techniques in order to better evaluate the performance.

Validate results and subgroups on external data set—Because we are continuing to acquire data on these and new participants, we will be able to use external data to test whether behaviorally identified subgroups can be predicted from functional connectivity, or structural morphometry data. Furthermore, the approach outlined here can be used in future independent studies to identify putative clinical subgroups acquired at different sites. Ultimately, such independent studies can test whether previously identified subgroups are replicable, and whether such subgroups can be predicted in independent data.

It is important to note that the interpretations of our analyses here are limited by the age and gender gap between the ASD and TD cohorts. Interestingly, we saw no differences of age or IQ across the accurately identified subgroups, nor were any subgroups split by gender. In fact, controlling for age and gender by regressing out these variables, as the supplemental RF model identified subgroups that were strongly affected by age, gender, and IQ. Taken together, these points suggest that the original RF is capturing variation in typical cognitive profile, as opposed to age or gender. Nonetheless, this limitation should not outweigh the novel impact of the approach presented here; future studies can use this RF approach on independent samples and verify whether identified ASD subgroups display similar characteristics to the subgroups identified in the present sample.

With the aid of large-scale data consortiums (e.g. ABIDE¹⁰³), future studies will and should be able to identify putative subgroups and validate them further in larger samples. New tools will be needed to reduce the large dimensionality of the imaging data, and potential site differences in data quality; however, such resources, which provide imaging data on over 2000 participants across multiple sites, should provide a natural extension to the current work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors of this study would like acknowledge all the members of Dr. Damien Fair's, Dr. Joel Nigg's, and Dr. Bonnie Nagel's labs for their hard work on this study. Additionally, we thank Dr. Shannon Mcweeny and Dr. Garet Lahvis for their advice and for critically reading our manuscript. We are also extremely grateful to all the families who took the time to participate in this study. Eric Feczko was supported by the National Library of Medicine Postdoctoral Fellowship. This research was supported by DeStefano Family Foundation and the National Institutes of Mental Health (R01 MH096773, R00MH091238, R01 MH096773-03S1, R01 MH 096773-05, R01 MH086654, R01 MH086654).

References

1. Constantino JN, Charman T. Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression. *Lancet. Neurol.* 2016; 15:279–291. [PubMed: 26497771]
2. Fombonne E. The prevalence of autism. *JAMA.* 2003; 289:87–89. [PubMed: 12503982]
3. Mcpheeters ML, et al. A Systematic Review of Medical Treatments for Children With Autism Spectrum Disorders abstract. *Pediatrics.* 2011; doi: 10.1542/peds.2011-0427
4. Hill EL. Executive dysfunction in autism. *Trends Cogn. Sci.* 2004; 8:26–32. [PubMed: 14697400]

5. Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* 2011; 1380:42–77. [PubMed: 21129364]
6. Hughes JR. A review of recent reports on autism: 1000 studies published in 2007. *Epilepsy Behav.* 2008; 13:425–437. [PubMed: 18627794]
7. Stone VE, Gerrans P. What's domain-specific about theory of mind? *Soc Neurosci.* 2006; 1:309–319. [PubMed: 18633796]
8. Baddeley A. Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 2003; 4:829–839. [PubMed: 14523382]
9. Bennetto L, Pennington BF, Rogers SJ. Intact and impaired memory functions in autism. *Child Dev.* 1996; 67:1816–1835. [PubMed: 8890510]
10. Russell J, Jarrold C, Henry L. Working memory in children with autism, with moderate learning difficulties. *J Child Psychol Psychiatry.* 1996; 37:673–686. [PubMed: 8894948]
11. Ozonoff S, Strayer DL. Further Evidence of Intact Working Memory in Autism. *J. Autism Dev. Disord.* 2001; 31:257–263. [PubMed: 11518480]
12. Joseph RM, Steele SD, Meyer E, Tager-Flusberg H. Self-ordered pointing in children with autism: Failure to use verbal mediation in the service of working memory? *Neuropsychologia.* 2005; 43:1400–1411. [PubMed: 15989932]
13. Joseph RM, McGrath LM, Tager-Flusberg H. Executive dysfunction and its relation to language ability in verbal school-age children with autism. *Dev Neuropsychol.* 2005; 27:361–378. [PubMed: 15843102]
14. Chen S-F, et al. Deficits in executive functions among youths with autism spectrum disorders: an age-stratified analysis. *Psychol. Med.* 2016; :1–14. DOI: 10.1017/S0033291715002238
15. Faja S, et al. Executive function predicts the development of play skills for verbal preschoolers with autism spectrum disorders. *Autism Res.* 2016; :1–11. DOI: 10.1002/aur.1608
16. Bowler DM, Poirier M, Martin JS, Gaigg SB. Nonverbal short-term serial order memory in autism spectrum disorder. *J. Abnorm. Psychol.* 2016; 125:886–893. [PubMed: 27732024]
17. Macizo P, Soriano MF, Paredes N. Phonological and Visuospatial Working Memory in Autism Spectrum Disorders. *J. Autism Dev. Disord.* 2016; 46:2956–2967. [PubMed: 27314268]
18. Diamond A. Executive functions. *Annu. Rev. Psychol.* 2013; 64:135–168. [PubMed: 23020641]
19. Geurts HM, van den Bergh SFWM, Ruzzano L. Prepotent response inhibition and interference control in autism spectrum disorders: Two Meta-Analyses. *Autism Res.* 2014; 7:407–420. [PubMed: 24596300]
20. Geurts HM, Verte S, Oosterlaan J, Roeyers H, Sergeant JA. How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *J. Child Psychol. Psychiatry.* 2004; 45:836–854. [PubMed: 15056314]
21. Adams NC, Jarrold C. Inhibition in autism: Children with autism have difficulty inhibiting irrelevant distractors but not prepotent responses. *J. Autism Dev. Disord.* 2012; 42:1052–1063. [PubMed: 21830170]
22. Critchfield TS, Kollins SH. Temporal discounting: basic research and the analysis of socially important behavior. *Journal of Applied Behavior Analysis.* 2001; 34:101–122. [PubMed: 11317983]
23. Demurie E, Roeyers H, Baeyens D, Sonuga-Barke E. Temporal discounting of monetary rewards in children and adolescents with ADHD and autism spectrum disorders. *Dev. Sci.* 2012; 15:791–800. [PubMed: 23106733]
24. Demurie E, Roeyers H, Baeyens D, Sonuga-Barke E. Domain-general and domain-specific aspects of temporal discounting in children with ADHD and autism spectrum disorders (ASD): A proof of concept study. *Res. Dev. Disabil.* 2013; 34:1870–1880. [PubMed: 23578902]
25. Chantiluke K, et al. Disorder-specific functional abnormalities during temporal discounting in youth with Attention Deficit Hyperactivity Disorder (ADHD), Autism and comorbid ADHD and Autism. *Psychiatry Res. - Neuroimaging.* 2014; 223:113–120.
26. Pascualvaca DM, Fantie BD, Papageorgiou M, Mirsky AF. Attentional capacities in children with autism: Is there a general deficit in shifting focus? *J. Autism Dev. Disord.* 1998; 28:467–478. [PubMed: 9932233]

27. Tye C, et al. Attention and inhibition in children with ASD, ADHD and co-morbid ASD + ADHD: an event-related potential study. *Eur. Child Adolesc. Psychiatry*. 2014; 7:e1210–5.
28. Corbett BA, Constantine LJ. Autism and attention deficit hyperactivity disorder: assessing attention and response control with the integrated visual and auditory continuous performance test. *Child Neuropsychol*. 2006; 12:335–348. [PubMed: 16911977]
29. Lundervold AJ, et al. Attention Deficits in Children With Combined Autism and ADHD: A CPT Study. *J. Atten. Disord*. 2016; 20:599–609. [PubMed: 22940852]
30. Yasuda Y, et al. Cognitive inflexibility in Japanese adolescents and adults with autism spectrum disorders. *World J. psychiatry*. 2014; 4:42–48. [PubMed: 25019056]
31. Pruett JR, et al. Impaired Eye Region Search Accuracy in Children with Autistic Spectrum Disorders. *PLoS One*. 2013; 8
32. Nakahachi T, et al. Disturbed holistic processing in autism spectrum disorders verified by two cognitive tasks requiring perception of complex visual stimuli. *Psychiatry Res*. 2008; 159:330–338. [PubMed: 18417223]
33. Morin K, et al. Atypical face perception in autism: A point of view? *Autism Res*. 2015; 8:497–506. [PubMed: 25683613]
34. Stevenson RA, et al. Arrested Development of Audiovisual Speech Perception in Autism Spectrum Disorders. *J. Autism Dev. Disord*. 2014; 44:1470–1477. [PubMed: 24218241]
35. Bebko JM, Schroeder JH, Weiss JA. The McGurk effect in children with autism and asperger syndrome. *Autism Res*. 2014; 7:50–59. [PubMed: 24136870]
36. Annaz D, Karmiloff-Smith A, Johnson MH, Thomas MSC. A cross-syndrome study of the development of holistic face recognition in children with autism, Down syndrome, and Williams syndrome. *J. Exp. Child Psychol*. 2009; 102:456–486. [PubMed: 19193384]
37. Barton JJS, et al. Are patients with social developmental disorders prosopagnosic? Perceptual heterogeneity in the Asperger and socio-emotional processing disorders. *Brain*. 2004; 127:1706–1716. [PubMed: 15215211]
38. Hefter RL, Manoach DS, Barton JJS. Perception of facial expression and facial identity in subjects with social developmental disorders. *Neurology*. 2005; 65:1620–1625. [PubMed: 16301491]
39. Tanaka JW, et al. The perception and identification of facial emotions in individuals with Autism Spectrum Disorders using the Let's Face It! Emotion Skills Battery. *J. Child Psychol. Psychiatry*. 2012; 53:1259–1267. [PubMed: 22780332]
40. Chen R, Jiao Y, Herskovits EH. Structural MRI in autism spectrum disorder. *Pediatr Res*. 2011; 69:63R–8R.
41. Amaral DG, Schumann CM, Nordahl CW. Neuroanatomy of autism. *Trends Neurosci*. 2008
42. Brambilla P, et al. Brain anatomy and development in autism: Review of structural MRI studies. *Brain Res. Bull*. 2003; 61:557–569. [PubMed: 14519452]
43. Lange N, et al. Longitudinal Volumetric Brain Changes in Autism Spectrum Disorder Ages 6–35 Years. *Autism Res*. 2015; 8:82–93. [PubMed: 25381736]
44. Dierker DL, et al. Analysis of cortical shape in children with simplex autism. *Cereb. Cortex*. 2015; 25:1042–1051. [PubMed: 24165833]
45. Nordahl CW, et al. Cortical folding abnormalities in autism revealed by surface-based morphometry. *J Neurosci*. 2007; 27:11725–11735. [PubMed: 17959814]
46. Valk SL, Di Martino A, Milham MP, Bernhardt BC. Multicenter mapping of structural network alterations in autism. *Hum. Brain Mapp*. 2015; 36:2364–2373. [PubMed: 25727858]
47. Wallace GL, et al. Longitudinal cortical development during adolescence and young adulthood in autism spectrum disorder: Increased cortical thinning but comparable surface area changes. *J. Am. Acad. Child Adolesc. Psychiatry*. 2015; 54:464–469. [PubMed: 26004661]
48. Kucharsky Hiess R, et al. Corpus Callosum Area and Brain Volume in Autism Spectrum Disorder: Quantitative Analysis of Structural MRI from the ABIDE Database. *J. Autism Dev. Disord*. 2015; 45:3107–3114. [PubMed: 26043845]
49. Shokouhi M, Williams JH, Waiter GD, Condon B. Changes in the sulcal size associated with autism spectrum disorder revealed by sulcal morphometry. *Autism Res*. 2012; 5:245–252. [PubMed: 22674695]

50. Feczko E, et al. The hemodynamic response in children with Simplex Autism. *Dev. Cogn. Neurosci.* 2012; 2:396–408. [PubMed: 22795455]
51. Ray S, et al. Structural and functional connectivity of the human brain in autism spectrum disorders and attention-deficit/hyperactivity disorder: A rich club-organization study. *Hum. Brain Mapp.* 2014; doi: 10.1002/hbm.22603
52. Monk CS, et al. Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *Neuroimage.* 2009; 47:764–772. [PubMed: 19409498]
53. Cherkassky VL, Kana RK, Keller TA, Just MA. Functional connectivity in a baseline resting-state network in autism. *Neuroreport.* 2006; 17:1687–1690. [PubMed: 17047454]
54. Anderson JS, et al. Decreased interhemispheric functional connectivity in autism. *Cereb. Cortex.* 2011; 21:1134–1146. [PubMed: 20943668]
55. Di Martino A, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry.* 2014; 19:659–67. [PubMed: 23774715]
56. Supekar K, et al. Brain Hyperconnectivity in Children with Autism and its Links to Social Deficits. *Cell Rep.* 2013; 5:738–747. [PubMed: 24210821]
57. Cerliani L, et al. Increased Functional Connectivity Between Subcortical and Cortical Resting-State Networks in Autism Spectrum Disorder. *JAMA Psychiatry.* 2015; 72:1–11.
58. Gotts SJ, et al. The perils of global signal regression for group comparisons: A case study of Autism Spectrum Disorders. *Front Hum Neurosci.* 2013; 7
59. Redcay E, et al. Intrinsic functional network organization in high-functioning adolescents with autism spectrum disorder. *Front. Hum. Neurosci.* 2013; 7:573. [PubMed: 24062673]
60. Tyszka JM, Kennedy DP, Paul LK, Adolphs R. Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cereb. Cortex.* 2014; 24:1894–1905. [PubMed: 23425893]
61. Duchesnay E, et al. Feature selection classification of imbalanced datasets. Application to PET images of children with autistic spectrum disorders. *Neuroimage.* 2011; 57:1003–1014. [PubMed: 21600290]
62. Murdaugh DL, et al. Differential Deactivation during Mentalizing and Classification of Autism Based on Default Mode Network Connectivity. *PLoS One.* 2012; 7
63. Wang H, Chen C, Fushing H. Extracting Multiscale Pattern Information of fMRI Based Functional Brain Connectivity with Application on Classification of Autism Spectrum Disorders. *PLoS One.* 2012; 7:1–14.
64. Jamal W, et al. Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates. *J. Neural Eng.* 2014; 11:046019. [PubMed: 24981017]
65. Katuwal GJ, Baum SA, Cahill ND, Michael AM. Divide and Conquer?: Sub-Grouping of ASD Improves ASD Detection Based on Brain Morphometry. *PLoS One.* 2016; 11:1–24.
66. Abraham A, et al. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *Neuroimage.* 2016; doi: 10.1016/j.neuroimage.2016.10.045
67. Chen CP, et al. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *NeuroImage. Clin.* 2015; 8:238–45. [PubMed: 26106547]
68. Sabuncu MR, Konukoglu E. Clinical Prediction from Structural Brain MRI Scans?: A Large-Scale Empirical Study. *Neuroinformatics.* 2014; doi: 10.1007/s12021-014-9238-1
69. Breiman L, Cutler A. Breiman and Cutler's random forests for classification and regression. Packag. 'randomForest'. 2012; 29doi: 10.5244/C.22.54
70. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:1118–1123. %U<http://www.ncbi.nlm.nih.gov/PMC/pmc/lookup/quickdisp?req=doi:10.1073/pnas.0706375105> [PubMed: 18216267]
71. Fair, Da, et al. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. *Front. Syst. Neurosci.* 2012; 6:80. [PubMed: 23382713]

72. Wechsler, D. Wechsler Intelligence Scale for Children—4th Edition (WISC-IV®). Harcourt Assessment; 2003.
73. Breiman LEO. Random Forests. *Mach. Learn.* 2001; 45:5–32.
74. Mitchell SH. Measures of impulsivity in cigarette smokers and non-smokers. *Psychopharmacol.* 1999; 146:455–464.
75. Wilson VB, Mitchell SH, Musser ED, Schmitt CF, Nigg JT. Delay discounting of reward in ADHD: application in young children. *J. Child Psychol. Psychiatry.* 2011; 52:256–264. [PubMed: 21083561]
76. Johnson MW, Bickel WK. An algorithm for identifying nonsystematic delay-discounting data. *Exp. Clin. Psychopharmacol.* 2008; 16:264. [PubMed: 18540786]
77. Robbins TW, et al. Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia.* 1994; 5:266–281. [PubMed: 7951684]
78. Logan GD. in *Inhibitory processes in attention, memory, and language.* 1994:189–239.
79. Nigg JT. The ADHD response-inhibition deficit as measured by the Stop Task: replication with DSM-IV combined type, extension, and qualification. *J Abnorm Child Psychol.* 1999; 27:393–402. [PubMed: 10582840]
80. Green, DM., Swets, JA. *Signal Detection Theory and Psychophysics.* Wiley; 1966.
81. Burton, aM, White, D., McNeill, A. The Glasgow Face Matching Test. *Behav. Res. Methods.* 2010; 42:286–291. [PubMed: 20160307]
82. Tottenham N, et al. The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 2009; 168:242–249. [PubMed: 19564050]
83. Glasser MF, et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage.* 2013; 80:105–24. [PubMed: 23668970]
84. Fischl B. FreeSurfer. *Neuroimage.* 2012; 62:774–781. [PubMed: 22248573]
85. Power JD, et al. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage.* 2014; 84:320–341. [PubMed: 23994314]
86. Gordon EM, et al. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb. Cortex.* 2014; 26:288–303. [PubMed: 25316338]
87. Karalunas SL, et al. Subtyping Attention-Deficit/Hyperactivity Disorder Using Temperament Dimensions?: Toward Biologically Based Nosologic Criteria. *JAMA psychiatry.* 2014; 71:1015–24. [PubMed: 25006969]
88. Eggebrecht AT, et al. Joint Attention and Brain Functional Connectivity in Infants and Toddlers. *Cereb. Cortex.* 2017; :1–12. DOI: 10.1093/cercor/bhw403
89. Gordon EM, et al. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb. Cortex.* 2014; doi: 10.1093/cercor/bhu239
90. MacKay DG. The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychol. Rev.* 1982; 89:483–506.
91. Szatmari P, et al. Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA psychiatry.* 2015; 72:276–83. [PubMed: 25629657]
92. Liu W, Li M, Yi L. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Res.* 2016; 9:888–898. [PubMed: 27037971]
93. Crippa A, et al. Use of Machine Learning to Identify Children with Autism and Their Motor Abnormalities. *J. Autism Dev. Disord.* 2015; 45:2146–2156. [PubMed: 25652603]
94. Katuwal GJ, Cahill ND, Baum SA, Michael AM. The Predictive Power of Structural MRI in Autism Diagnosis. in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 2015; :4270–4273. DOI: 10.1109/EMBC.2015.7319338
95. Volkmar FR, McPartland JC. Moving beyond a categorical diagnosis of autism. *Lancet. Neurol.* 2016; 15:237–238. [PubMed: 26497770]
96. Constantino, J., Gruber, C. *Social responsiveness scale.* Western Psychological Services; 2005.

97. van der Meer MJM, et al. Homogeneous Combinations of ASD--ADHD Traits and Their Cognitive and Behavioral Correlates in a Population-Based Sample. *J. Atten. Disord.* 2014; 1087054714533194.
98. Rommelse NNJ, van der Meer MJM, Hartman CA, Buitelaar JK. Cognitive Profiling Useful for Unraveling Cross-Disorder Mechanisms Support for a Step-Function Endophenotype Model. *Clin. Psychol. Sci.* 2016; 4:957–970.
99. Vandenbroucke MW, Scholte HS, van Engeland H, Lamme VA, Kemner C. A neural substrate for atypical low-level visual processing in autism spectrum disorder. *Brain.* 2008; 131:1013–1024. [PubMed: 18192288]
100. Nebel MB, et al. Disruption of functional organization within the primary motor cortex in children with autism. *Hum. Brain Mapp.* 2014; 35:567–580. [PubMed: 23118015]
101. Miller GA, Chapman JP. Misunderstanding analysis of covariance. *J. Abnorm. Psychol.* 2001; 110:40–48. [PubMed: 11261398]
102. Fair DA, et al. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. *Front. Syst. Neurosci.* 2012; 6:80. [PubMed: 23382713]
103. Di Martino A, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry.* 2014; 19:659–67. [PubMed: 23774715]

Figure 1. Depiction of stimuli used in face and affect processing experiments. (A) Example from visual facial affect recognition task. (B) Example from facial identity recognition task. (C) Example from auditory facial affect recognition task.

Figure 2.

(A) Plot of accuracy for observed (blue) vs. permuted (red) RF models. Wide bars refer to the 25th/75th percentiles and thinner bars refer to the 2.5th/97.5th percentiles. (B) Sorted proximity matrix, where each row and column represents a participant and each cell represents the number of times two participants ended in the same terminal node across all the RF models. (C) Plot of RF classification accuracy for ASD subgroups, error bars represent 1 standard error of the mean (SE). Dashed blue line represents 50% mean accuracy. (D) Plot of RF classification accuracy for control subgroups. Error bars represent 1 SE. Dashed blue line represents 50% mean accuracy. (E) Plot of SRS for ASD subgroups. The color code for each subgroup is maintained throughout all subfigures.

Figure 3.

Radar plots represent the 50th percentile for performance per group. All data are normalized within each radar plot from 0 to 100 percent. Per sample, one-way ANOVAs were conducted on raw data to reduce the number of points plotted based on differences between subgroups. The colors for each subgroup are the same as in Figure 2.

Figure 4. Visualization of systems of the brain used in the chi-squared analysis. Aud = Auditory. CIO = Cingulo-opercular. CIP = Cingulo-parietal. Def = Default-mode. DoA = Dorsal attention. FrP = Frontal-parietal. ReT = Retrosplenial. SMh = Somato-motor hand. Smm = somato-motor mouth. VeA = ventral attention. Vis = Visual. SBcT = subcortical. Non = none.

Figure 5.

(A) Plot of t-statistics for significant clustering observed between ASD and controls. (B) Plot of F ratios for significant clustering observed in the ANOVAs by subgroup. Colors surrounding significant clusters reflect the functional systems involved in the module (e.g. within or between system connectivity). (C) Visualization of estimated marginal means via subgroup by network interactions. Error bars represent 2 times the standard error of the mean. Colors surrounding the boxes reflect the functional systems involved and are consistent with the colors in B.

Table 1

Age table for ASD and TD samples per test. TD = Typically Developing; ASD = Autism Spectrum Disorder; M = mean; SD = Standard Deviation. Independent-sample t-tests revealed that subgroups were significantly differed in terms of age on the Facial and Affect Processing Tasks, Spatial Span, Delay Discounting, CPT, Stop Task, and MRI scans. Note that the demographics for the Facial and Affect Processing Tasks applies to the Face Identity Recognition, Facial Affect Matching, and the Vocal Affect Recognition tasks.

Age			
Average for All Tests: TD = 58, ASD=47			
TDM(SD)	ASD M(SD)	Tstat	p-value
10.29 (1.48)	12.15 (2.16)	5.237	0.006
Facial&Affect Processing: TD = 55. ASD = 46			
11.26 (1.51)	12.51 (2.14)	3.426	<.001
Delay Discounting : TD = 58, ASD =47			
10.12 (1.74)	12.41 (2.50)	5.509	<.001
Stop Task: TD = 58, ASD = 46			
9.91 (1.62)	12.00(2.29)	5.46	<.001
Spatial Soan Task: TD = 58 ASD =47			
10.08 (1.69)	11.89 (2.27)	4.69	<.001
CPT: TD = 58, ASD =47			
9.91 (1.62)	12.07 (2.24)	5.64	<.001
MRI Data: TD = 42, ASD = 26			
10.73 (1.74)	12.73 (2.03)	5.08	<.001

Table 2

Demographics table for ASD and TD samples per test. WISC BD = Wechsler's Intelligence Scale for Children IV: Block design raw score.

IQ scores	TD	ASD
WISC BD(M)	38	40.81
WISC BD(SD)	12.78	13.96
T score		1.074
p-value		0.279
Gender		
Males(N)	27	38
Males(%)	46.56	80.85
Females(N)	31	9
Females(%)	54.45	19.15
Pearson's Chi-square		12.951
p-value		<.001
Ethnicity		
Non-Hispanic(N)	49	39
Non-Hispanic(%)	84.48	82.38
Hispanic (N)	9	8
Hispanic (%)	15.52	17.02
Pearson's Chi-square		0.07
p-value		0.966
Race		
White(N)	51	39
White(%)	87.93	82.98
Black/African American (N)	1	2
Black/African American (%)	1.72	4.26
Asian (N)	4	1
Asian (%)	6.7	2.13
Native Hawaiian/Pacific Islander (N)	0	1
Native Hawaiian/Pacific Islander (%)	0	2.13
Pearson's Chi-square		4.295
p-value		0.368

Table 3

Table of task measures used in RF analysis. Independent samples t-tests were conducted between all available ASD and TD data. RT = reaction time.

Behavior Data						
Task	Variable	TD M(SD)	ASD M(SD)	T Score	p-value	df
Delay Discounting	7-day indifference score	8.16(1.99)	7.65 (2.9)	1.06	0.29	103
Delay Discounting	90-day indifference score	4.81 (3.03)	4.92(3.31)	-0.16	0.88	103
Delay Discounting	Natural log of k	-4.48(1.83)	-4.74 (2.2)	0.64	0.53	94
Delay Discounting	30-day indifference score	6.32(2.91)	5.89 (3.64)	0.66	0.51	103
Delay Discounting	K-value	0.0521 (0.185)	0.057 (0.14)	0.14	0.89	94
Delay Discounting	180-indifference score	3.77(3.1)	4.11 (3.82)	0.50	0.62	103
Delay Discounting	Initial indifference score	9.97(0.417)	9.89 (0.464)	0.99	0.33	103
Delay Discounting	AUG	0.527 (0.26)	0.543(0.318)	-0.27	0.79	95
Delay Discounting	Timepoint and Score R2 value	0.682 (0.352)	0.666 (0.39)	1.58	0.12	102
Stop Task	Probability of Stopping on Stop Trials	51 (3.75)	51.4(5.28)	-0.53	0.60	102
Stop Task	Stop Signal RT (ms)	253 (72.4)	303(129)	-2.49	0.01	102
Stop Task	Mean RT on Go Trials (ms)	703(130)	816(205)	-3.40	0.00	103
Stop Task	Go Trials Accuracy	95.2(3.48)	95 (4.63)	0.35	0.73	103
Stop Task	SD Go Trials RT (ms)	199(59.2)	242 (93.5)	-2.84	0.01	103
CPT	Bias score Stim Trials	-0.301 (0.133)	-0.317(0.157)	0.52	0.60	100
CPT	DprimeStim Trials	2.69 (0.82)	2.67 (0.866)	0.10	0.92	100
CPT	Dprime Catch trials	1.21 (0.684)	1.41 (0.828)	-1.36	0.18	100
CPT	Natural Log of Bias Score Stim Trials	1.6(0.793)	1.72(1.05)	-0.67	0.50	100
CPT	Natural log of Bias Score Catch Trials	-0.12(1.08)	0.0662(1.42)	-0.74	0.46	100
CPT	Bias Score Catch Trials	0.057 (0.73)	-0.0191 (0.525)	0.58	0.56	100
Spatial Span	SS Backwards Number Completed	8.59 (2.31)	8.19(2.99)	0.75	0.42	103
Spatial Span	SS Backward RT (ms)	1290(438)	1370 (701)	-0.70	0.48	103
Spatial Span	SS Backward Response Consistency	0.619(0.105)	0.559 (0.143)	2.46	0.02	103
Spatial Span	SS Backward Span Number Correct	5.41 (1.86)	4.87 (2.51)	1.26	0.21	103
Spatial Span	SS Forward Number Completed	9.34 (2.26)	8.79 (2.95)	1.08	0.28	103
Spatial Span	SS Forward RT (ms)	1170(348)	1170 (443)	0.10	0.92	103

Behavior Data					
Task	Variable	TD M(SD)	ASD M(SD)	T Score	p-value
Spatial Span	SS Forward Response Consistency	0.627 (0.0935)	0.622 (0.123)	0.24	0.81
Spatial Span	SS Forward Span Number Correct	5.93(1.93)	5.66 (2.43)	0.63	0.53
Facial Affect	Total Correct	18.4(2.27)	17.1 (3.24)	2.39	0.02
Facial Affect	Median RT (s)	5.06(1.74)	5.45(1.92)	-1.07	0.24
Facial Recognition	Total Correct	22.6(1.91)	19.6(4.06)	4.77	0.00
Facial Recognition	Median RT (s)	6.05 (2.37)	6.22 (3.05)	-0.31	0.76
Vocal Affect	Total Correct	16.7(2.17)	15.8(3.02)	1.63	0.11
Vocal Affect	Median RT (s)	1.96(0.658)	1.68(0.72)	2.04	0.04

Table 4

A list of the estimated marginal means from the chi-square ANOVA test. Values in parentheses reflect the standard error of the marginal means.

Table of marginal means from chi-square test				
Module	control SG1	ASD SG1	ASD SG2	ASD SG3
AUD-CIP	−0.04 (0.009)	0.12 (0.01)	−0.02 (0.01)	−0.04 (0.004)
CIO-DEF	−0.09 (0.002)	−0.09 (0.005)	−0.02 (0.005)	−0.02 (0.004)
DEF-DEF	0.17 (0.002)	0.09 (0.005)	0.13 (0.005)	0.1 (0.005)
DEF-SMH	−0.02 (0.002)	−0.037 (0.005)	−0.06 (0.005)	−0.002 (0.003)
DOA-SMH	−0.005 (0.002)	−0.004 (0.006)	0.04 (0.006)	0.03 (0.004)
DOA-SBCT	−0.02 (0.002)	−0.06 (0.008)	−0.07 (0.007)	−0.05 (0.005)
SAL-VIS	−0.04 (0.004)	0.05 (0.01)	0.03 (0.01)	−0.03 (0.008)