

Automatic classification of children with autism spectrum disorder by using a computerized visual-orienting task

Qiao He,^{1†} Qiandong Wang,^{2†} Yaxue Wu,^{3,4} Li Yi,^{3,4} and Kunlin Wei^{3,4} 

¹Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China, ²Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (Beijing Normal University), Faculty of Psychology, Beijing Normal University, Beijing, China, ³School of Psychological and Cognitive Sciences, Peking University, Beijing, China, ⁴Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

Abstract: Early screening and diagnosis of autism spectrum disorder (ASD) primarily rely on behavioral observations by qualified clinicians whose decision process can benefit from the combination of machine learning algorithms and sensor data. We designed a computerized visual-orienting task with gaze-related or non-gaze-related directional cues, which triggered participants' gaze-following behavior. Based on their eye-movement data registered by an eye tracker, we applied the machine learning algorithms to classify high-functioning children with ASD (HFA), low-functioning children with ASD (LFA), and typically developing children (TD). We found that TD children had higher success rates in obtaining rewards than HFA children, and HFA children had higher rates than LFA children. Based on raw eye-tracking data, our machine learning algorithm could classify the three groups with an accuracy of 81.1% and relatively high sensitivity and specificity. Classification became worse if only data from the gaze or nongaze conditions were used, suggesting that “less-social” directional cues also carry useful information for distinguishing these groups. Our findings not only provide insights about visual-orienting deficits among children with ASD but also demonstrate the promise of combining classical behavioral paradigms with machine learning algorithms for aiding the screening for individuals with ASD.

Keywords: autism diagnosis; autism screening; autism spectrum disorder; automatic classification; machine learning; visual orienting

Correspondence Dr Kunlin Wei and Li Yi, School of Psychological and Cognitive Sciences & Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China. Email: wei.kunlin@pku.edu.cn and yilipku@pku.edu.cn

Received 15 February 2020. Accepted 9 March 2021.

[†]These authors contributed equally as first author.

Autism spectrum disorder (ASD) is a pervasive neurodevelopmental disorder with impairments in social interactions and communications, and the presence of repetitive or stereotypical behaviors (American Psychiatric Association, 2013). While ASD has become a worldwide public health issue, its diagnosis currently relies on experienced clinicians who combine the clinical judgment of symptoms with semistructured, play-based behavioral observations (e.g., Autism Diagnostic Observation Scale [ADOS]; Lord et al., 2000), and standardized interviews or

questionnaires (e.g., Autism Diagnostic Interview-Revised [ADI-R]; Lord, Rutter, & Le Couteur, 1994). However, the diagnostic procedures of ASD based on these scales are laborious and time-consuming, and heavily dependent on the examiners' expertise.

To provide more objective methods, the use of machine learning algorithms in aiding ASD diagnosis or screening has been gaining popularity in recent years. Machine learning algorithms are capable of making decisions based on complex data. Specifically, these algorithms have the

potential to classify people with ASD and other populations automatically. For this purpose, researchers have evaluated various potential data formats (termed *features* in machine learning) that can be roughly summarized as two categories: the behavioral measures (Bone et al., 2016; Crippa et al., 2015; Duda, Ma, Haber, & Wall, 2016; Wall, Kosmicki, Deluca, Harstad, & Fusaro, 2012; Wedyan & Al-Jumaily, 2016), and the brain-imaging data (Anderson et al., 2011; Deshpande, Libero, Sreenivasan, Deshpande, & Kana, 2013; Ecker et al., 2010; Ghiassian, Greiner, Jin, & Brown, 2016; Iidaka, 2015; Ingallhalikar, Parker, Bloy, Roberts, & Verma, 2011; Jamal et al., 2014; Jiao et al., 2010; Murdaugh et al., 2012; Segovia et al., 2014; Stahl, Pickles, Elsabbagh, Johnson, & Team, 2012; Uddin et al., 2011; Wee, Wang, Shi, Yap, & Shen, 2014; Zhou, Yu, & Duong, 2014). Though having yielded promising results, previous studies have shared some methodological concerns that limit their potential use in clinical settings. First, the behavioral features used in previous studies often relied on manually labeled features such as the retrospective coding of children's behaviors in videotapes (Baranek, 1999; Clifford & Dissanayake, 2008; Osterling, Dawson, & Munson, 2002; Ozonoff et al., 2010; Teitelbaum, Teitelbaum, Nye, Fryman, & Maurer, 1998). For instance, an early study had achieved an excellent accuracy of 93.75% when distinguishing the ASD group, the developmental delay group, and the typically developing (TD) group (e.g., Baranek, 1999). However, they used retrospective video-based analyses of sensorimotor and social behaviors of infants. Like traditional clinical diagnosis, these features were essentially based on subjective ratings, which reduced the appeal of using machine learning algorithms. Second, most studies had performed the binary classification to distinguish children with ASD from TD children (Anzulewicz, Sobota, & Delafeld-Butt, 2016; Crippa et al., 2015; Kosmicki, Sochat, Duda, & Wall, 2015; Wall et al., 2012), or children with ASD from children with attention deficit hyperactivity disorder (ADHD; Duda et al., 2016). The accuracies for the binary classification are higher than multiclass classification, owing to its higher base rate. However, this has raised criticisms that the binary classification fails to take into account the heterogeneity in the ASD population (Bone et al., 2016; Wall et al., 2012). For example, different individuals with ASD differ remarkably in cognitive level or intelligence. Individuals with ASD with a cognitive level below the normal range (i.e., low-functioning individuals

with ASD, LFA) may exhibit behaviors that are confounded with those of individuals with developmental delays (Osterling et al., 2002) and even with those of normal people with autistic traits, such as stereotyped behaviors, aggression, and noncompliance (J. E. Dawson, Matson, & Cherry, 1998). Third, automatic classification with brain-imaging data is generally limited to adults or older children (over 7 years) due to the challenge of collecting brain-imaging data from young children, especially those with ASD (Anderson et al., 2011; Deshpande et al., 2013; Ecker et al., 2010; Ingallhalikar et al., 2011; Jiao et al., 2010; Murdaugh et al., 2012). Therefore, the application of machine learning to young children or infants, who are arguably in more need of automatic classification, is challenging at present (but see Hazlett et al., 2017).

To address these limitations, the present study aimed to combine the machine learning algorithms with objective behavioral measures to automatically classify children with LFA, high-functioning ASD (HFA, i.e., children with ASD who have a cognitive level in the normal range), and TD children. To achieve this goal, we gamified the well-known visual-orienting tasks and measured children's eye movements when performing the task. Visual orienting, or adjusting visual attention toward a target based on directional cues (especially the direction of others' gaze), is widely identified as one of the characteristic deficits for individuals with ASD, and thus has the potential to provide useful features for ASD classification (Congiu, Fadda, Doneddu, & Striano, 2016; G. Dawson, Meltzoff, Osterling, Cleary, & Brown, 1998; Goldberg et al., 2008; Lewy & Dawson, 1992; Senju, Tojo, Dairoku, & Hasegawa, 2004). Individuals with ASD have impairments in visual-orienting tasks, especially in those with social directional cues such as human gazes; for example, they have difficulty in sharing attention with others by following their gazes (Clifford & Dissanayake, 2008; G. Dawson et al., 2004; Lewy & Dawson, 1992). Even if children with ASD show gaze-following behaviors, their subsequent behaviors seem qualitatively different from those of TD children. For example, TD children spent more time looking at the attended object than the unattended one whereas children with ASD spent a similar amount of looking time on the attended and the unattended objects (Falck-Ytter, Thorup, & Bölte, 2014). The prevailing explanation is that individuals with ASD have difficulty inferring others' intentions from gazes, which leads to deficits in

visual-orienting behaviors such as spontaneous gaze following or attention sharing (Bedford et al., 2012). This explanation appears to be consistent with brain-imaging data of individuals with ASD who have shown reduced activation in the superior temporal sulcus, a brain region associated with interpreting others' intentions (Pelphrey, Morris, & McCarthy, 2005).

It is noteworthy that individuals with ASD also have shown impairments in visual-orienting tasks with nongaze cues (G. Dawson et al., 1998; G. Dawson et al., 2004; Stauder, Bosch, & Nuij, 2011). Children with ASD reacted significantly slower than TD controls even when the direction was cued by an arrow symbol, suggesting that the impairment of visual orienting extended to cues that arguably contain less social information (Stauder et al., 2011). Furthermore, ASD-related deficits shown in visual-orienting tasks were associated with the cognitive level and age (Osterling & Dawson, 1994; Povinelli & Povinelli, 1996; Webster & Potter, 2008). For example, researchers have reported that the performance of children with ASD in discriminating others' gazing targets tended to increase with cognitive levels (Webster & Potter, 2008). Thus, these previous findings have suggested that the visual orienting based on cues with varying degrees of social components might provide potential features not only for distinguishing individuals with ASD from TD individuals but also for distinguishing ASD individuals with different levels of cognitive abilities.

In the present study, we adopted a computerized visual-orienting task to measure the eye movements among children with HFA, children with LFA, and TD children. We used four types of directional cues: arrow, finger, the gaze of a cartoon animal's face, and human gaze. Arguably, these cues are associated with increasing levels of social information. Participants were required to follow the cued direction to look at a target picture and sustain their fixations on the target to obtain rewarding feedback. Their eye-movement data were continuously recorded and used as features without human ratings or prior assumptions. Given that these participant groups have shown the distinct performance of visual orienting (e.g., Congiu et al., 2016; G. Dawson et al., 2004; Goldberg et al., 2008; Lewy & Dawson, 1992; Senju et al., 2004; Stauder et al., 2011), we hypothesized that their eye movements alone would enable excellent automatic classification of groups.

Methods

Participants

Twenty-six children with HFA ($M_{\text{age}} = 5.08$ years, $SD_{\text{age}} = 0.83$ years; 23 boys and 3 girls), 24 age-matched children with LFA ($M_{\text{age}} = 4.98$ years, $SD_{\text{age}} = 1.09$ years; 19 boys and 5 girls), and 24 age-matched TD children ($M_{\text{age}} = 5.24$ years, $SD_{\text{age}} = 0.81$ years; 18 boys and 6 girls) were recruited. All children with ASD were recruited in a preschool specialized for children with ASD, and have all been previously diagnosed by professional pediatricians in the licensed hospitals according to the diagnostic criteria for ASD in the *Diagnostic and Statistical Manual of Mental Disorders-5th Edition* (DSM-5, American Psychiatric Association, 2013). We were unable to administer the Autism Diagnostic Observation Schedule or the Autism Diagnostic Interview-Revised (Lord et al., 2000; Lord et al., 1994) to confirm the ASD diagnosis in our study since the Chinese translated versions of these two scales have not been officially released and validated. All children with ASD scored above the threshold on the Chinese version of the Social Communication Questionnaire (SCQ; Berument, Rutter, Lord, Pickles, & Bailey, 1999). We divided the children with ASD into two subgroups according to their nonverbal and verbal mental intelligence as measured by Combined Raven's Test (CRT-C2; Raven, Raven, & Court, 1998) and the Peabody Picture Vocabulary Test-Revised (PPVT-R; Sang & Miao, 1990; Vance & Stone, 1990), respectively. Twenty-four children with ASD who failed to complete the CRT-C2 or scored lower than the average score of the norm in the PPVT-R test were assigned to the LFA group. The remaining 26 children with ASD were assigned to the HFA group. The HFA group and the TD group were matched by both the nonverbal mental intelligence, $t(48) = -.49$, $p = .624$, and the verbal mental intelligence, $t(46) = -.95$, $p = .356$, and both groups scored higher than the LFA group in PPVT, $ps < .01$. Both the HFA and the LFA groups scored higher on the SCQ than the TD group, $t(47) = 5.068$, $p < .001$; $t(47) = 8.28$, $p < .001$, respectively, and the two ASD groups scored similarly on the SCQ, $t(48) = -1.70$, $p = .096$. The present research had been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and approved by the Committee for Protecting Human and Animal Subjects at the School of Psychological and Cognitive Sciences at Peking

University, China. We obtained written parental consents before the formal experiment. Demographic data as well as cognitive and SCQ scores of the participants are summarized in Table 1.

Materials

In each trial, a directional cue was presented at the center of a screen, and four object pictures were presented in four directions around the cue (Figure 1A). The directional cue had four different types: the gaze of a human face, the gaze of a cartoon face, a pointing finger, and an arrow (Figure 1B). Each cue type included 40 different pictures (width: 200 pixels, height: 280 pixels, resolution: 72 pixels per inch), 10 for each of the four possible directions (left, right, up, and down; Figure 1B). Thus, we had a total of 160 cue pictures, each of which was randomly assigned to one of the 160 trials. For each trial, we also displayed four object pictures around the cue picture; one was the target, and the other three were the confounds. Thus, a total of 640 object pictures were used. These pictures depicted familiar objects for children, including furniture, musical instruments, sports equipment, stationery, fruits, and water containers. In each trial, the four object pictures were randomly chosen from the 640 different pictures without replacement and randomly presented as the target and the confounds. The distance between the center of the cue and the center of each peripheral object was 250 pixels.

We used a Tobii X120 eye tracker (120 Hz sample rate) to record the eye movements. Stimuli were presented using Psychtoolbox (<http://psychtoolbox.org>) on the Matlab platform (Math Works Inc., Natick, MA). The display monitor was 25.5 cm in height and 34 cm in width with 1024 × 768 pixel resolution.

Procedure

Children sat approximately 60 cm away from the screen. A five-point calibration (four at the corners and one at the center of the screen) procedure was provided, and the calibration was considered as successful only when the average visual angles were smaller than 1° for all five calibration points. Five practice trials with similar settings, but with different pictures, were first administered to the children to familiarize them with the task. The 160 trials of the formal test, with 10 trials for each cue direction (up, down, left, and right) in each cue condition (human gaze, cartoon gaze, finger, and arrow), were randomly presented.

The experimenter started each trial by pressing an enter key when the children paid attention to the screen. A trial began with a red cross blinking at the center of the black screen for 1000 ms. A center cue picture and four surrounding object pictures were then displayed simultaneously. The children were instructed to look at the target and sustain their visual attention on it (Figure 1A). The instruction was: "Look at the person/little animal, what is she/it looking at? Will you look at that, too?" If the children managed to fixate on the target for 1000 ms, the trial would end, and a yellow smiling face was shown on the screen. The experimenter would praise the children for each successful trial by saying "Great job!". If the children failed to fixate on the target for 1000 ms within 4000 ms after the trial began, this trial was marked as a failed trial, and no feedback was provided. The experimenter would then encourage the child by saying "We need to look at what he is looking at to get a smiling face." Note that the experimenter did not know a priori whether a child with ASD belonged to the LFA group or the HFA group.

It should be noted that the choice of the timing parameter 1000 ms was based on the task goal; that is, participants

Table 1
Demographics of the participants

Groups	TD	HFA	LFA	Difference (<i>t</i> test)		
				HFA vs. TD	LFA vs. TD	HFA vs. LFA
Male (Female)	18 (6)	23 (3)	19 (5)	N/A	N/A	N/A
Age (years)	5.24 (0.81)	5.08 (0.83)	4.98 (1.09)	−0.683	−0.94	0.369
Raw Raven Score	22.17 (9.03)	20.77 (10.83)	11.00 (2.00)	−0.49	−5.91*	4.34
Raw PPVT Score	80.38 (30.16)	77.44 (22.39)	22.88 (12.30)	−0.9515	−10.59***	10.55***
SCQ Score	8.50 (5.13)	12.23 (6.92)	17.32 (5.23)	−3.73	−8.82***	−5.09**

Note. SDs are shown in parentheses. TD = typically developing children; HFA = high-functioning children with autism spectrum disorder; LFA = low-functioning children with autism spectrum disorder; N/A = not applicable; Raw Raven = Combined Raven's Test; PPVT = Peabody Picture Vocabulary Test-Revised; SCQ = Chinese version of the Social Communication Questionnaire.

* $p < .05$. ** $p < .01$. *** $p < .001$.

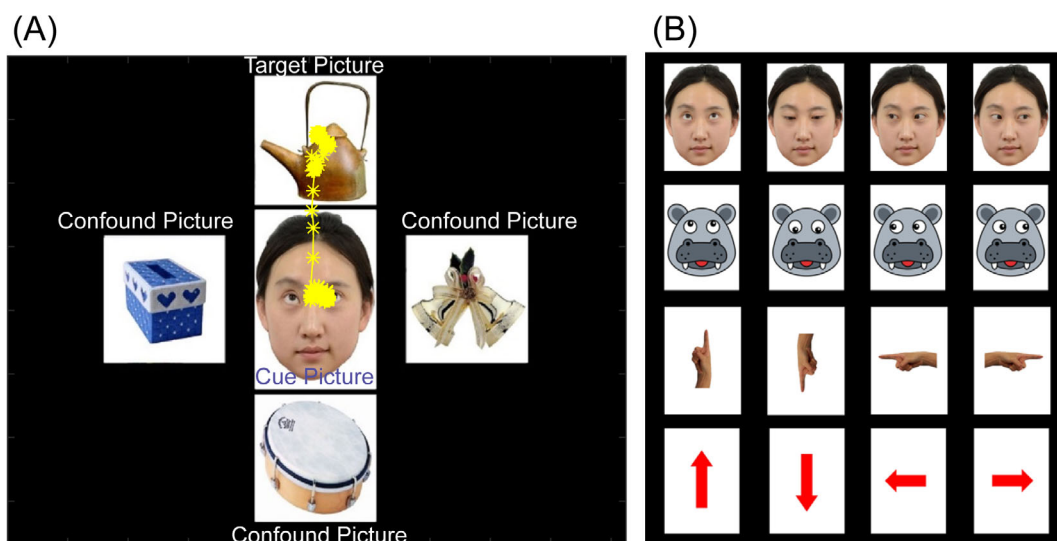


Figure 1. Visual stimuli used in the experiment. (A) The layout of stimuli and eye gaze data from a sample trial. The participant first fixated at the center cue to obtain the directional information, then moved his or her gaze to the corresponding target and stayed on it for 1000 ms to complete the trial. Yellow stars denote gaze samples, which are connected by lines representing saccadic paths. (B) Examples of directional cues. From top to bottom: human face, cartoon face, finger, and arrow; from left to right: up, down, left, and right directional cues. Photographs were personally approved for publication.

should fixate on the target to indicate their choice. Our pilot experiment showed that children with ASD might occasionally choose one target, but then change their mind and fixate on another (often the right) target. Thus, we decided to have a long fixation time (1000 ms) to give them an opportunity to change their mind. The 4000 ms was defined by task heuristics, as our pilot experiment indicated that some children with ASD would lose their attention during a trial. This is common for children of this age. In this case, we would need to terminate the trial and move on to the next trial. In addition, for visual orienting, most studies used 3000 ms to 6000 ms per trial (e.g., Falck-Ytter et al., 2014; Wang et al., 2020). Thus, the threshold of 4000 ms to define a failed trial is determined by piloting and task heuristics.

Data analysis

We defined four areas of interest (AOIs): the cue picture (cue region), the target picture (target region), the three confounding pictures (confound region), and all other areas (other region). The sample positions of the left and the right eyes were averaged for subsequent analyses. A fixation was defined as the continuous gazes within a radius of 50 pixels for 100 ms or longer. We excluded trials with failed data recording, either due to large head movements or failure to follow the instruction. The excluded trials

constituted 1.16%, 3.54%, and 0.81% of the total trials for the HFA, the LFA, and the TD groups.

To measure the children's general performance, we computed the success rate, which was defined as the percentage of successful trials within each condition. We also calculated the latency of initial responses (hereafter abbreviated as *response latency*) as the delay from the onset of the picture presentation until the gaze moved out of the cue picture in the successful trials. The fixation time on the cue picture, the target picture, and the confounding picture(s) were calculated and then converted as percentages of their corresponding trial durations.

Statistical tests for these descriptive measures were performed to compare between participant groups and cue conditions. Two-way mixed-design analyses of variance (ANOVAs), with cue conditions as the within-group factor and participant group as the between-group factor, were conducted on the aforementioned measures. We further compared the task performance (success rate) between the first and the last 10 trials for each cue type, testing whether there was any learning effect throughout the experiment. The Bonferroni correction was used for post hoc pairwise comparisons. All tests were two-sided with α set at .05. All the statistical tests were performed using SPSS Version 20.0 (IBM Corp., NY, USA).

For the automatic classification of the three groups, we used all eye-movement data without prior assumptions or

subjective selections of features. We computed the statistics of the spatial distributions of gaze samples (e.g., M , SD , skewness, kurtosis). Table 2 lists all the statistics that serve as features. The features were computed for x and y coordinates separately and for four AOIs (i.e., cue, target, confound, and other regions) separately. As the target picture and the confounds appeared at random locations on each trial, we used local coordinates instead of global coordinates of gaze samples that fell in the target region and the confound region. Samples in the cue region and the other region used their original global coordinates. As each trial had three confounds regions, we overlaid them as one region. In addition to the statistics of sample distributions, we also counted the frequency of samples in these regions. As a result, we obtained a total of 68 features, all of which were used without selection in automatic classification.

We used the k -nearest neighbors (KNN) algorithm to train and test our classification models (Dudani, 1976). We chose KNN for its simplicity and for the low dimensionality of our data (i.e., x - y coordinates of gaze data). Furthermore, the KNN performed well with a relatively small sample size and overlapped features that are characteristics of our data set. Piloting on the data from one previous study (Liu, Li, & Yi, 2016), the number of neighbors for KNN was set at 2 and 5 for two-group and three-group classifications, respectively. Subject-wise leave-one-out cross-validation method was used to evaluate the model performance. This method separates the original data into a training data set (all subjects except one) and a testing data set (one subject not in the training data set). The training data set was used to optimize the model whereas the test data set was used to test the model's prediction. We computed the overall accuracy as well as specificity and sensitivity of model predictions.

Table 2
Features computed Based on Spatial Distributions of Gaze Samples

Feature Name	No. of Features
Mean (of x , y in each region)	8
Median	8
Standard Deviation	8
Skewness	8
Kurtosis	8
Interquartile Range (IQR)	8
Minimum Coordinate	8
Maximum Coordinate	8
Number of Sample Points in each Region	4

Results

Behavioral results

To give the first impression of eye-movement patterns during our visual-orientation task, we presented exemplary trials from three representative participants—one from each of the three experimental groups (Figure 2). The TD children performed the task well: They fixated at the center cue picture for some time, then immediately shifted their gaze to the target and stayed on the target picture for 1000 ms to receive their final reward (Figure 2A). The children with HFA performed worse than the TD children; sometimes they shifted their gaze to the target, but did not remain there for long. Instead, they shifted their gaze to a confound (Figure 2B). In this case, they might fail the trial as they did not fixate on the target for 1 s. Children with LFA performed the worst among the three groups, and their scan paths appeared to be more random and less purposeful than the other two groups. In our LFA example, the child failed to orient his gaze to the correct target and instead explored multiple confounding pictures (Figure 2C).

Success rate

The success rate of trials reflects children's general behavioral performance. Two-way mixed-design ANOVA on success rate yielded a significant main effect on group, $F(2, 71) = 119.830$, $p < .001$, $\eta^2 = .771$, and a significant main effect on cue type, $F(3, 69) = 9.965$, $p < .001$, $\eta^2 = .302$. The cue group interaction effect was not significant, $F(6, 213) = 1.544$, $p = .165$, $\eta^2 = .042$. Post hoc tests showed a ranking order among three participant groups (Figure 3): The TD group had the highest success rate (74.7%), and the HFA group (29.1%) performed better than the LFA group (5.7%). Post hoc tests for cue type showed that subjects had a significantly higher success rate for both nonface cues (41.6% and 39.8% for finger and arrow, respectively) than both face cues (33.8% and 30.7% for human and cartoon, respectively), $ps < .01$. To test whether there was a learning effect throughout the experiment, we compared the number of successful trials in the first and the last 10 trials for each cue condition. The success rate did not differ in any group for any condition, $ps > .05$. Thus, children did not improve their visual-orienting performance over the trials.

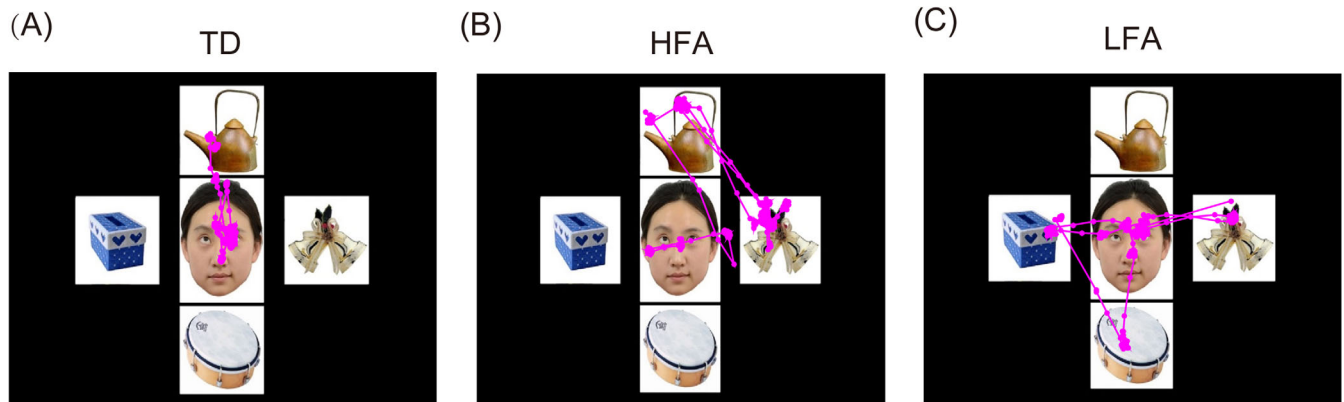


Figure 2. Scan paths overlaid on the visual stimuli presented in the visual-orienting task. (A) Three typical trials from the typically developing children (TD), (B) high-functioning children with autism spectrum disorder (HFA), and (C) low-functioning children with autism spectrum disorder (LFA) groups are shown separately.

Response latency

The response latency is a reaction time-like measure, capturing the delay from the onset of stimulus presentation to the time that a participant's gaze first exits the cue region (Figure 4). Similar to the results of the success rate, the cue type group interaction was not significant, $F(6, 213) = .921$, $p = .481$, $\eta^2 = .025$. Significant main effects were found between groups, $F(2, 71) = 12.301$, $p < .001$, $\eta^2 = .257$, and cue types, $F(3, 69) = 15.461$, $p < .001$, $\eta^2 = .402$. Pairwise comparisons using Bonferroni correction showed that the LFA group ($M = 1277$ ms, $SD = 308$ ms) had significantly longer response latency than both the HFA ($M = 1117$ ms, $SD = 276$ ms), $p < .05$,

and the TD ($M = 986$ ms, $SD = 185$ ms) groups, $p < .001$. There was no difference between the HFA and the TD groups, $p = .079$. Post hoc test on cue type showed that subjects had significantly shorter response latency for non-face cues (1074 ms and 1015 ms for finger and arrow cues, respectively) than for face cues (1223 ms and 1195 ms for human and cartoon cues, respectively), $ps < .001$.

Duration in cue, target, foil, and other noninformation regions

We computed the total time that a participant's gaze fell in the four AOI regions (Figure 5). For the center cue region, the interaction effect between cue condition and group was

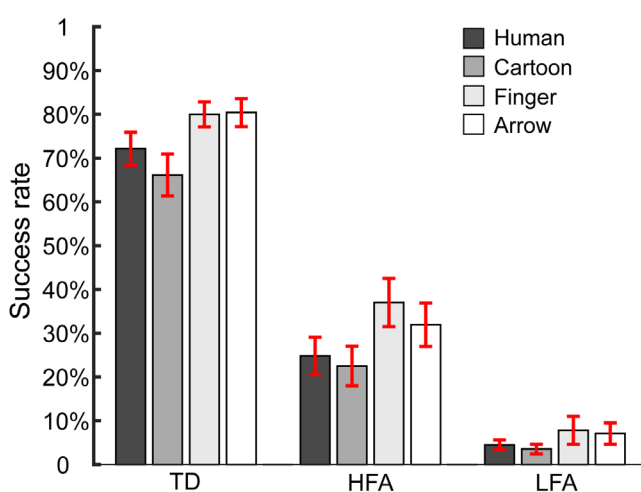


Figure 3. Success rates are shown for each participant group and each cue type separately. The error bars denote standard error of the mean across participants. TD = typically developing children; HFA = high-functioning children with autism spectrum disorder; LFA = low-functioning children with autism spectrum disorder.

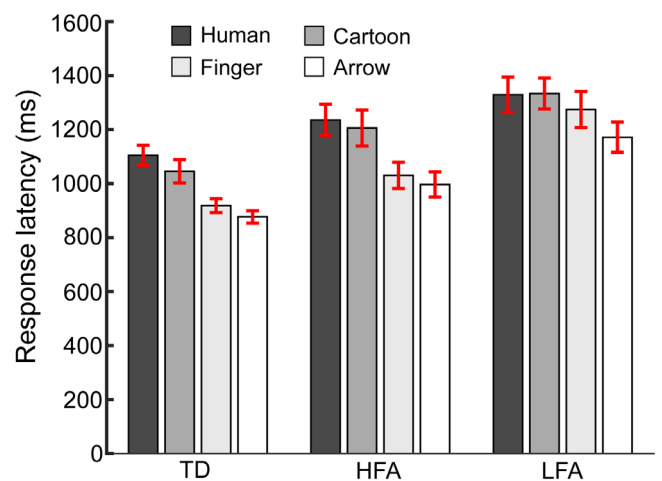


Figure 4. Response latency of the three groups in the four cue conditions. It measures the time it takes to release attention from the cue picture. The error bars denote standard error of the mean across participants. TD = typically developing children; HFA = high-functioning children with autism spectrum disorder; LFA = low-functioning children with autism spectrum disorder.

not significant, $F(4.278, 151.866) = 1.420$, $p = .227$, $\eta^2 = .038$. We found significant main effects of group, $F(2, 71) = 5.071$, $p < .01$, $\eta^2 = .125$, and cue condition, $F(3, 69) = 36.641$, $p < .001$, $\eta^2 = .614$. Post hoc tests on group difference found that the HFA group had a significantly longer duration than the LFA group in this region, $p < .01$. Post hoc tests on cue conditions revealed that the duration in the human face and cartoon face conditions were significantly longer than in the finger and arrow conditions, $ps < .001$. Thus, participants from all groups appeared to spend more time analyzing the gaze cue pictures than the nongaze cue pictures. Especially, the LFA group appeared to make their effort to decode the cued direction as they also spent more time on the more difficult cue pictures.

For the target region, the interaction effect between cue condition and group was significant, $F(4.748, 168.545) = 3.094$, $p < .01$, $\eta^2 = .080$. We also found significant main effects of group, $F(2, 71) = 118.822$, $p < .001$, $\eta^2 = .770$, and cue type, $F(3, 69) = 7.261$, $p < .001$,

$\eta^2 = .240$. Post hoc tests showed that the TD group had a longer duration than the HFA group, $p < .001$, and the HFA had a longer duration than the LFA group, $p < .001$. This ranking of groups might be related to their differences in general performance, where successful trials required sufficient duration of gazes in the target region (>1 s to receive rewards). Post hoc tests on cue type demonstrated that the duration in the target region for nonface cues was significantly longer than that for face cues, $p < .001$. Simple main effect tests revealed that the difference between cue types was significant only for the HFA group, $F(3, 69) = 9.687$, $p < .001$, $\eta^2 = .296$, whose two gaze conditions (human and cartoon) had shorter durations than the two nongaze conditions, $ps < .001$. Again, this effect of cue type in the HFA group echoed with their general performance: They indeed performed worse with a shorter duration on the target in the two gaze cue conditions. The LFA group spent little time in the target region and did not exhibit any difference between the cue types, suggesting

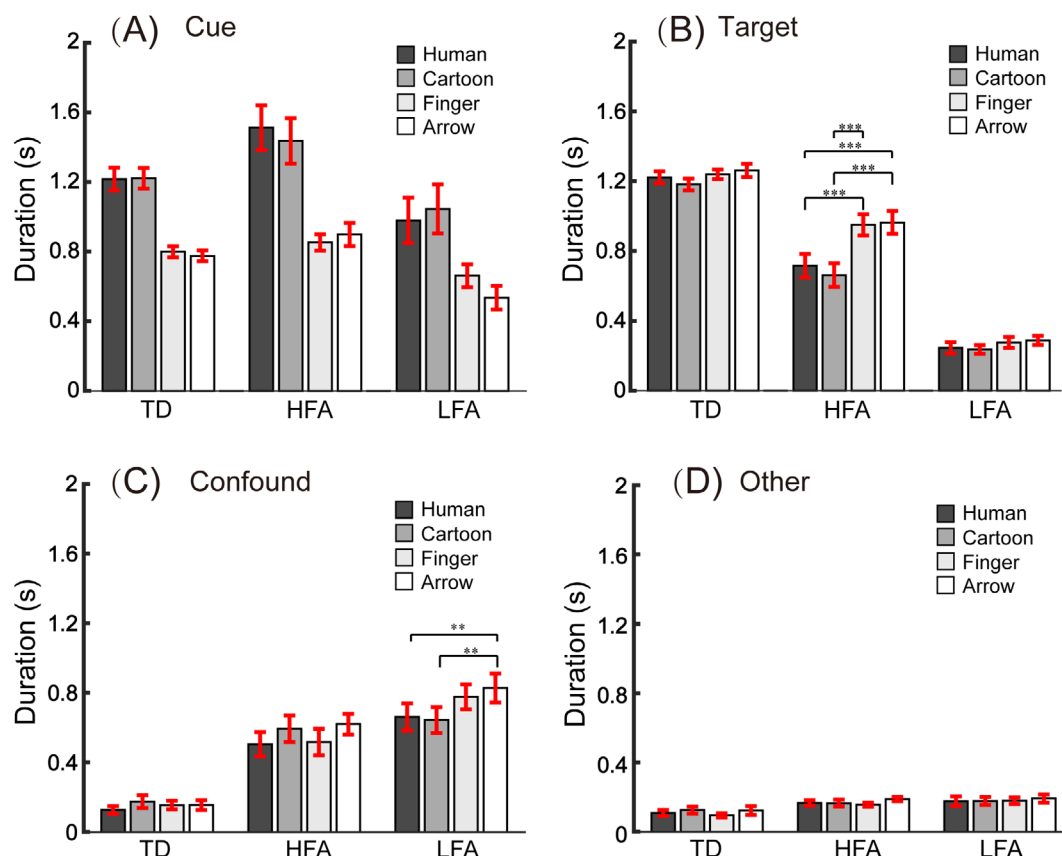


Figure 5. Duration of gaze in each area of interest, plotted separately for four different regions and four cue conditions. The cue, target, confound, and other regions are shown in Subplots A to D, respectively. The error bars denote standard error of the mean across participants. Significant results are marked only for the simple main effects. * $p < .05$. ** $p < .01$. *** $p < .001$. TD = typically developing children; HFA = high-functioning children with autism spectrum disorder; LFA = low-functioning children with autism spectrum disorder.

that their poor performance did not depend on the amount of social information contained in the cue picture.

For the confound region, the interaction effect between cue type and group was marginally significant, $F(5.098, 180.970) = 2.065$, $p = .059$, $\eta^2 = .055$. We found significant main effects of group, $F(2, 71) = 28.084$, $p < .001$, $\eta^2 = .442$, and cue condition, $F(3, 69) = 3.741$, $p < .05$, $\eta^2 = .140$. Post hoc tests on groups showed that both the HFA and the LFA groups had a longer duration in the confounding region than the TD group, $ps < .001$. For cue type, post hoc tests showed that subjects had a significantly longer duration for the arrow condition than for the human condition, $p < .05$. Simple main effect tests revealed that the cue condition differences were significant only in the LFA group, $F(3, 69) = 4.620$, $p < .01$, $\eta^2 = .167$. Post hoc tests found a significant difference between the cue conditions in the LFA group, whose two gaze cue conditions had significantly smaller duration than the arrow cue condition, $p < .01$.

For the other region, the only significant difference was the main effect of group, $F(2, 71) = 5.473$, $p < .01$, $\eta^2 = .134$. Pairwise comparisons showed that both the HFA and the LFA groups had longer durations in the uninformative region than the TD group, $ps < .05$. Thus, the duration in the confound region and the other region differed between groups, reflecting the performance difference between groups.

Classification results

Our features for the automatic classification were based on spatial distributions of gaze samples in the four AOIs (Figure 6). As there was no clear sign of learning effect across trials, we stacked gaze samples from all trials in each condition. Their spatial distributions had large overlaps between groups, but their shapes still differed. Using all the 68 features without selection, the KNN model yielded good classification performance with an overall accuracy of 81.08%, as shown by its confusion matrix (Table 3). The most common errors were mislabeling TD and LFA children as HFA children, echoing with our descriptive results that HFA's performance was in-between TD and LFA. As a result, the HFA group had the lowest specificity than the other two groups, suggesting that the classifier tended to label the other two groups as the HFA group. To further evaluate the classifier's performance, we derived a receiver operating characteristic curve (ROC) for each group and calculated its area under the curve (AUC).

The x -axis and y -axis of the ROC curve denote $(1 - \text{specificity})$ and sensitivity, respectively. A model with $\text{AUC} > .7$ is considered acceptable, and $\text{AUC} > .8$ is considered excellent. The AUC was 0.775, 0.908, and 0.927 for the HFA, the LFA, and the TD groups, respectively (Figure 7A). In sum, the classifier achieved decent performance with visual-orienting data.

We also evaluated whether the nongaze cue conditions, which had been associated with less group difference in the literature, contributed to the classification. To do so, we separated the data from the two gaze conditions (human face and cartoon face conditions) from the two nongaze conditions (finger and arrow conditions). The classification accuracy based on data from the two gaze conditions dropped to 78.4%; likewise, the accuracy based on data from the two nongaze conditions dropped to 77.0%. This reduction in classification accuracy suggested that the nongaze conditions also contain useful information to distinguish groups.

We also evaluated the improvement in classification if the model only needed to distinguish children with ASD from TD children. We combined data from the LFA and the HFA groups as a single ASD group and performed the same training and cross-validation of the KNN model. The overall accuracy and the AUC increased to 93.24% and 0.938, respectively (for details, see Table 4 and Figure 7B). Apparently, a binary classification could achieve much better overall performance with our visual-orienting data and its corresponding model.

We also assessed the performance of classification based solely on the HFA or the LFA groups, two scenarios rather typical in the literature. These previous studies recruited either HFA or LFA participants as the ASD group and performed binary classification between the ASD and the TD groups (Figure 7B). We used the data of our 24 TD children and 26 HFA children to train the KNN model and used the data of the 24 LFA children as the test data. We found that the classifier could correctly label these LFA children as children with ASD with an accuracy of 100%. On the other hand, if we used the data of our 24 TD children and 24 LFA children as the training set and used the data of the 26 HFA children as the test set, the overall accuracy dropped to 53.85%. Given that the performance of the three participant groups ranked in a descending order, these results suggested that a model trained with more similar groups (i.e., TD vs. HFA) can perform well with test data from a disparate group (i.e., LFA), and a model trained with data from more different groups

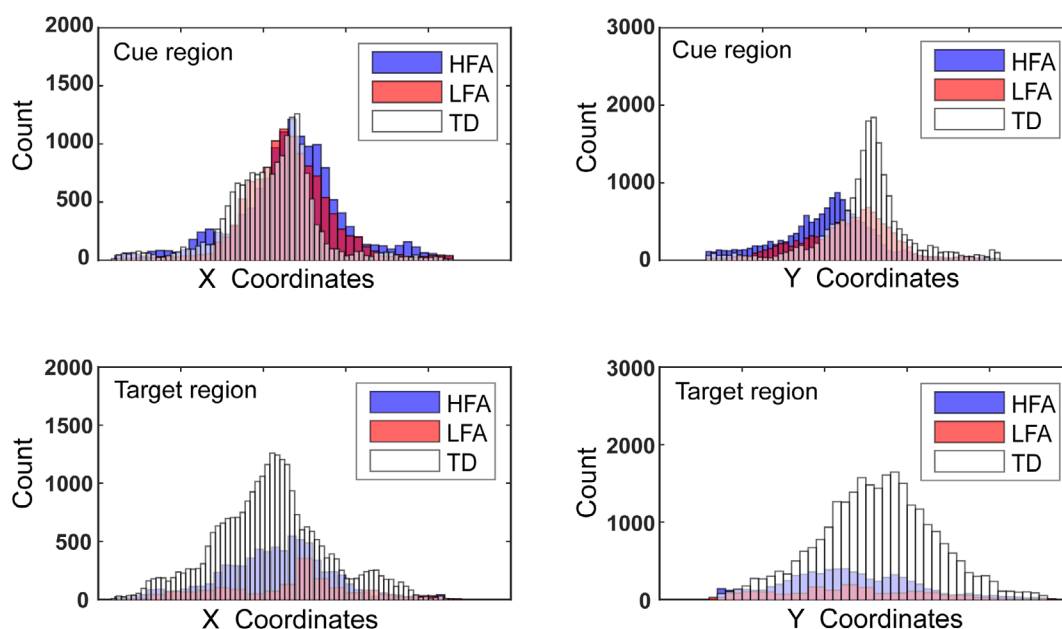


Figure 6. The spatial distributions of the gaze samples of three random participants from three different groups. Distributions in x and y coordinates and distributions in different areas of interest (Only cue and target regions are shown here.) are plotted separately. Data from the three groups are overlaid. HFA = high-functioning children with autism spectrum disorder; LFA = low-functioning children with autism spectrum disorder; TD = typically developing children.

(e.g., TD vs. LFA) will underperform with test data from an intermediate group (e.g., HFA) which might resemble both training groups.

Discussion

Using a computerized visual-orienting task, we first confirmed the deficits in the visual orienting in children with ASD based on the eye-movement data. Compared with the TD children, the children with ASD showed deficits in responding to directional cues with varying levels of social components, suggesting that their deficits in visual

orienting are not only tied to social cues but appear to be domain-general. Furthermore, the deficits in visual orienting were more severe in the children with LFA than the children with HFA, suggesting that the general cognitive ability may play an essential role in the performance of the visual-orienting task. Importantly, leveraging on the raw gaze samples during the task, the machine learning algorithm achieved excellent performance in automatic classification of ASD. Without prior assumptions on the eye-movement patterns, the algorithms were able to classify the three children's groups with a maximum classification accuracy of 81.08% and relatively high sensitivity and specificity. Classification performance also indicates that

Table 3

The Confusion Matrix of the Three-Group Classification

		Prediction results			Actual totals	% Actual error
		HFA	LFA	TD		
Actual Labels	HFA	19	5	2	26	26.9%
	LFA	5	19	0	24	20.8%
	TD	2	0	22	24	8.3%
Prediction totals		26	24	24	74	—
Prediction error (%)		26.9%	20.83%	8.3%	—	—

HFA: Sensitivity = 73.1%, Specificity = 85.4%.

LFA: Sensitivity = 79.2%, Specificity = 90.0%.

TD: Sensitivity = 91.7%, Specificity = 96.0%.

Note. HFA = high-functioning children with autism spectrum disorder; LFA = low-functioning children with autism spectrum disorder.

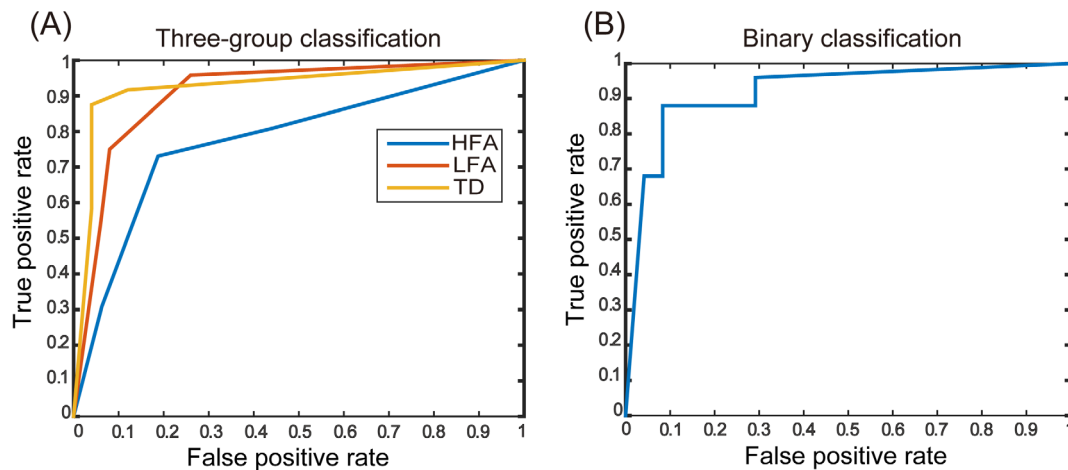


Figure 7. Receiver operating characteristic curve (ROC) curves of the group classification. (A) Three-group classification. The ROC curves are plotted separately for the high-functioning children with autism spectrum disorder (HFA) group vs. the two remaining groups, the low-functioning children with autism spectrum disorder (LFA) group vs. the two remaining groups, and the typically developing children (TD) group vs. the two remaining groups. (B) Two-group classification. The ROC curve is for classification between the autism spectrum disorder group (HFA and LFA combined) and the TD group.

conditions with fewer social cues (with nongaze directional cues) provide informative features for distinguishing participant groups.

As expected, the success rate, a general measure of the visual-orienting performance, ranked in descending order from the TD, to the HFA, and to the LFA groups. In general, the children with HFA showed significantly lower success rates in the two gaze cue conditions (human face and cartoon face) than in the nongaze cue conditions (finger and the arrow). In contrast, this trend was less obvious for TD children. These results indicate that the gaze-related information (especially the human faces) poses a greater challenge than does nongaze information for the children with HFA, as reported by many previous studies (Karttinen, Puura, Himanen, Nevalainen, & Hietanen, 2016; Kirchgessner, Chuang, Patel, & Sereno, 2015; Robic et al., 2015; Weigelt, Koldewyn, & Kanwisher, 2012).

In fact, the TD children also found the two gaze conditions more difficult: Their response latency was significantly longer in the two gaze cue conditions than in the two nongaze cue conditions. This result is consistent with previous findings demonstrating that TD individuals are slower in response to eye-gaze cues than to arrow cues (Ristic, Friesen, & Kingstone, 2002). Interestingly, the children with HFA exhibited similar between-condition differences in the response latency as TD children: They took more time to disengage from the gaze-related cues than from the nongaze cues. These differences between the gaze and the nongaze cues might result from at least two possible reasons. One possibility is that the gaze cue is a more complex source of directional information, as compared to straightforward and simplistic finger and arrow cues. They might demand more sensory and social processing and are thus more difficult for both groups. The other possibility is

Table 4
The Confusion Matrix by Binary Classification

		Prediction results		Actual totals	Actual error%
		ASD	TD		
Actual Labels	ASD	48	2	50	4%
	TD	3	21	24	12.5%
Prediction totals		51	23	74	—
Prediction error (%)		5.88%	8.7%	—	—

ASD: sensitivity = 96%, specificity = 87.5%.

TD: sensitivity = 87.5%, specificity = 96%.

Note. ASD = autism spectrum disorder; TD = typically developing children.

that the gaze cue picture, displayed as a face, is more attractive for both groups, and thus they disengaged from the gaze cue picture slower than from the nongaze cue pictures. This latter possibility is less likely than the former because it cannot explain why the success rate is also lower in these gaze conditions. Furthermore, this attractiveness explanation is unlikely for children with ASD who show more interest in nonsocial information than in social information (Senju & Johnson, 2009; Tanaka & Sung, 2016). In sum, our findings indicate that besides an overall deficit in visual orientation among children with ASD, their impairments in processing social information or more complex visual cues appear to further broaden the deficits.

It is important to note that the children with LFA performed much worse than the children with HFA in our visual-orientation task. This result is expected, given that joint attention has been related to language development (Mundy, Sigman, & Kasari, 1990; Siller & Sigman, 2002) and executive function (McEvoy, Rogers, & Pennington, 1993) in ASD. Previous studies have also found that children with ASD with lower developmental levels, reflected by intelligence quotient, usually show less accuracy in detecting gaze directions (Webster & Potter, 2008) and less frequency of the joint attention behaviors (Mundy, Sigman, & Kasari, 1994). Because most previous findings have been based on studies with high-functioning individuals with ASD, our study raises the question about the generalizability of these findings to low-functioning individuals with ASD, considering the heterogeneity of the ASD samples (Beglinger & Smith, 2001). Our results highlight the need to consider the developmental level of the samples in autism research, which could provide a more comprehensive picture of the visual-orientation patterns across the spectrum.

Although the abnormal visual orientation is evident in individuals with ASD, to our surprise, leveraging on this task to perform the automatic classification of ASD has not yet been conducted. Our results support that the visual-orientation task has impressive potentials to discern between children with HFA, children with LFA, and TD children when combined with the machine learning algorithms. We found that it is sufficient to distinguish these three groups with 81.08% accuracy by using raw gaze samples. Importantly, when we dropped data from the nongaze or the gaze conditions, the classification performance would suffer substantially, suggesting that data from the nongaze conditions carry useful information for distinguishing these

participant groups. This further confirms our behavioral findings that deficits in visual orienting in children with ASD are not specific to gaze-related cues but are more cue-independent.

Our multiclass classification achieved a similar or better accuracy than binary classification reported by previous studies using eye-movement data (Fujioka et al., 2016; Liu et al., 2016) and brain-imaging data (Anderson et al., 2011; Deshpande et al., 2013; Ecker et al., 2010; Ghiassian et al., 2016; Iidaka, 2015; Ingallhalikar et al., 2011; Jamal et al., 2014; Jiao et al., 2010; Murdaugh et al., 2012; Segovia et al., 2014; Stahl et al., 2012; Uddin et al., 2011; Wee et al., 2014; Zhou et al., 2014). Compared to previous studies, our approach advances the literature in the following aspects. First, the features used in our study are more objective because they are computed from raw eye-movement data as opposed to the subjective ratings of children's behaviors (Baranek, 1999; Clifford & Dissanayake, 2008; Osterling et al., 2002; Ozonoff et al., 2010; Teitelbaum et al., 1998). Participants performed this computerized visual-orienting task while their eye movements were continuously tracked.

Second, our features were arbitrarily specified without a prior assumption on how our participants would behave in varying experimental conditions. Previous studies on the classification of ASD typically reported their best classification results using the optimal feature set among all possible feature sets. For example, Crippa et al. (2015) asked participants to transport a ball manually and recorded hand movements. After selecting an optimal kinematic feature set, they reported a binary classification accuracy of 96.7%. However, their feature-selection process with a small sample size ($n = 15$ subjects for each of the TD and ASD groups) leads to record-wise cross-validation of model predictions, as opposed to subject-wise cross-validation (Saeb, Lonini, Jayaraman, Mohr, & Kording, 2017). Record-wise cross-validation typically inflates the prediction accuracy, and subject-wise cross-validation is a more desired and appropriate way of evaluating the performance of automatic classification. Note that feature selection with a small sample size is currently prevalent among medical studies with the machine learning techniques, and its overestimation of prediction accuracy should be cautioned (Little et al., 2017; Saeb et al., 2017; Saeb, Lonini, Jayaraman, Mohr, & Kording, 2017). Our approach circumvented this problem because we computed the features based on the raw gaze samples and their spatial distributions, and all features were used without a selection process.

Third, we included low-function individuals with ASD in our experiment and used the multiclass classification whereas previous studies primarily used the binary classification to distinguish TD individuals either from high-functioning or low-functioning individuals. Multiclass classification usually sacrifices the prediction accuracy given the lower base rate, and our findings confirm this: The accuracy increases from 81.08% to 93.24% when we combined the HFA and the LFA groups into a single ASD group and performed binary classification. Our study thus went a step further to mimic the real diagnosis situations where heterogeneity of ASD individuals poses a challenge for any automatic classification algorithms (Walsh, Elsabbagh, Bolton, & Singh, 2011). Last, our task is more child-friendly and less expensive than brain imaging to implement in clinical settings. Indeed, the whole task only lasts about 15 min and can be easily administered.

Our study shall be improved when considering the following limitations. First, our visual-orientation task is based on the still pictures whereas the dynamic directional cues are more common in daily life. Whether dynamic cues are still effective in bringing up group differences should be studied in the future. Second, our paradigm differs from previous visual-orienting tasks by adding an attention-sustaining phase, which arguably contributes to the poor performance of children with ASD. Thus, future studies should specifically address whether our machine learning algorithms can successfully distinguish individuals with ASD versus individuals with attention deficits. It is well-known that individuals with ASD often have various types of attention deficits (Lai, Lombardo, & Baron-Cohen, 2014), and the comparison between the tasks with and without an attention-sustaining phase might provide valuable features for distinguishing these subgroups in the ASD population. Third, autistic girls significantly differ from autistic boys in social attention; for example, autistic girls exhibit more typical social attention than do autistic boys (Harrop et al., 2018). Due to the limited number of girl participants, we did not explore whether our method could perform similarly for two genders. Future studies with an approximately equal number of female and male participants should test this question. Finally, the predictive power of automatic classification techniques is constrained by the base rate of ASD in the population. Therefore, extra caution should be taken if a direct comparison of the machine learning algorithms with current diagnostic methods is made.

Conclusions

Our findings demonstrate the feasibility of combining machine learning algorithms with a visual-orientation task to discern between children with HFA, children with LFA, and TD children. This computerized task, coupled with automatic classification, promises to provide an objective, easy-to-implement, and potentially less costly method for the screening of ASD.

Disclosure of conflict of interest

The authors report no biomedical financial interests or potential conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China Grants 31671168, 32071047, 61533001, 31871116; the NSFC/DFG Collaborative Research Centre SFB/TRR169 “Crossmodal Learning” II; Guangdong Key Project in “Development of new tools for diagnosis and treatment of Autism” Grant 2018B030335001; and the Beijing Municipal Science & Technology Commission Grant Z171100000117015.

Declarations of interest

There are no declarations of interest to report.

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Arlington, VA: Author. <https://doi.org/10.1176/appi.books.9780890425596>
- Anderson, J. S., Nielsen, J. A., Froehlich, A. L., DuBray, M. B., Druzgal, T. J., Ciriello, A. N., ... Lainhart, J. E. (2011). Functional connectivity magnetic resonance imaging classification of autism. *Brain*, 134(12), 3742–3754. <https://doi.org/10.1093/brain/awr263>
- Anzulewicz, A., Sobota, K., & Delafeld-Butt, J. T. (2016). Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism. *Scientific Reports*, 6, 31107. <https://doi.org/10.1038/srep31107>
- Baranek, G. T. (1999). Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age. *Journal of Autism and Developmental Disorders*, 29(3), 213–224. <https://doi.org/10.1023/a:1023080005650>
- Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T., ... BASIS team (2012). Precursors to social and

- communication difficulties in infants at-risk for autism: Gaze following and attentional engagement. *Journal of Autism and Developmental Disorders*, 42(10), 2208–2218. <https://doi.org/10.1007/s10803-012-1450-y>
- Beglinger, L. J., & Smith, T. H. (2001). A review of subtyping in autism and proposed dimensional classification model. *Journal of Autism and Developmental Disorders*, 31(4), 411–422. <https://doi.org/10.1023/a:1010616719877>
- Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *The British Journal of Psychiatry*, 175(9), 444–451. <https://doi.org/10.1192/bjp.175.5.444>
- Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 57(8), 927–937. <https://doi.org/10.1111/jcpp.12559>
- Clifford, S. M., & Dissanayake, C. (2008). The early development of joint attention in infants with autistic disorder using home video observations and parental interview. *Journal of Autism and Developmental Disorders*, 38(5), 791–805. <https://doi.org/10.1007/s10803-007-0444-7>
- Congiu, S., Fadda, R., Doneddu, G., & Striano, T. (2016). Impaired representational gaze following in children with autism spectrum disorder. *Research in Developmental Disabilities*, 57, 11–17. <https://doi.org/10.1016/j.ridd.2016.06.008>
- Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M., & Castiglioni, I. (2015). Use of machine learning to identify children with autism and their motor abnormalities. *Journal of Autism and Developmental Disorders*, 45(7), 2146–2156. <https://doi.org/10.1007/s10803-015-2379-8>
- Dawson, G., Meltzoff, A. N., Osterling, J., Cleary, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, 28(6), 479–485. <https://doi.org/10.1023/a:1026043926488>
- Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology*, 40(2), 271–283. <https://doi.org/10.1037/0012-1649.40.2.271>
- Dawson, J. E., Matson, J. L., & Cherry, K. E. (1998). An analysis of maladaptive behaviors in persons with autism, PDD-NOS, and mental retardation. *Research in Developmental Disabilities*, 19(5), 439–448. [https://doi.org/10.1016/s0891-4222\(98\)00016-x](https://doi.org/10.1016/s0891-4222(98)00016-x)
- Deshpande, G., Libero, L. E., Sreenivasan, K. R., Deshpande, H. D., & Kana, R. K. (2013). Identification of neural connectivity signatures of autism using machine learning. *Frontiers in Human Neuroscience*, 7, 670. <https://doi.org/10.3389/fnhum.2013.00670>
- Duda, M., Ma, R., Haber, N., & Wall, D. P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Translational Psychiatry*, 6, e732. <https://doi.org/10.1038/tp.2015.221>
- Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(4), 325–327. <https://doi.org/10.1109/TSMC.1976.5408784>
- Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E. M., ... MRC AIMS Consortium (2010). Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. *NeuroImage*, 49(1), 44–56. <https://doi.org/10.1016/j.neuroimage.2009.08.024>
- Falck-Ytter, T., Thorup, E., & Bölte, S. (2014). Lack of processing bias for the objects other people attend to in 3-year-olds with autism. *Journal of Autism and Developmental Disorders*, 45(6), 1897–1904. <https://doi.org/10.1007/s10803-014-2278-4>
- Fujioka, T., Inohara, K., Okamoto, Y., Masuya, Y., Ishitobi, M., Saito, D. N., ... Kosaka, H. (2016). Gazefinder as a clinical supplementary tool for discriminating between autism spectrum disorder and typical development in male adolescents and adults. *Molecular Autism*, 7(1), 1–14. <https://doi.org/10.1186/s13229-016-0083-y>
- Ghiassian, S., Greiner, R., Jin, P., & Brown, M. R. G. (2016). Using functional or structural magnetic resonance images and personal characteristic data to identify ADHD and autism. *PLoS One*, 11(12), e0166934. <https://doi.org/10.1371/journal.pone.0166934>
- Goldberg, M. C., Mostow, A. J., Vecera, S. P., Larson, J. C. G., Mostofsky, S. H., Mahone, E. M., & Denckla, M. B. (2008). Evidence for impairments in using static line drawings of eye gaze cues to orient visual-spatial attention in children with high functioning autism. *Journal of Autism and Developmental Disorders*, 38(8), 1405–1413. <https://doi.org/10.1007/s10803-007-0506-x>
- Harrop, C., Jones, D., Zheng, S., Nowell, S. W., Boyd, B. A., & Sasson, N. (2018). Sex differences in social attention in autism spectrum disorder. *Autism Research*, 11(9), 1264–1275. <https://doi.org/10.1002/aur.1997>
- Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., ... Analysis, S. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, 542(7641), 348–351. <https://doi.org/10.1038/nature21369>
- Iidaka, T. (2015). Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex*, 63, 55–67. <https://doi.org/10.1016/j.cortex.2014.08.011>
- Ingalhalikar, M., Parker, D., Bloy, L., Roberts, T. P. L., & Verma, R. (2011). Diffusion based abnormality markers of pathology: Toward learned diagnostic prediction of ASD. *NeuroImage*, 57(3), 918–927. <https://doi.org/10.1016/j.neuroimage.2011.05.023>
- Jamal, W., Das, S., Oprescu, I.-A., Maharatna, K., Apicella, F., & Sicca, F. (2014). Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates. *Journal of Neural Engineering*, 11(4), 046019. <https://doi.org/10.1088/1741-2560/11/4/046019>
- Jiao, Y., Chen, R., Ke, X., Chu, K., Lu, Z., & Herskovits, E. H. (2010). Predictive models of autism spectrum disorder based on brain regional cortical thickness. *NeuroImage*, 50(2), 589–599. <https://doi.org/10.1016/j.neuroimage.2009.12.047>
- Kaartinen, M., Puura, K., Himanen, S.-L., Nevalainen, J., & Hietanen, J. K. (2016). Autonomic arousal response habituation to social stimuli among children with ASD. *Journal of Autism*

- and *Developmental Disorders*, 46(12), 3688–3699. <https://doi.org/10.1007/s10803-016-2908-0>
- Kirchessner, M. A., Chuang, A. Z., Patel, S. S., & Sereno, A. B. (2015). Intact reflexive but deficient voluntary social orienting in autism spectrum disorder. *Frontiers in Neuroscience*, 9, 453. <https://doi.org/10.3389/fnins.2015.00453>
- Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational Psychiatry*, 5, e514. <https://doi.org/10.1038/tp.2015.7>
- Lai, M.-C., Lombardo, M. V., & Baron-Cohen, S. (2014). Autism. *Lancet*, 383(9920), 896–910. [https://doi.org/10.1016/S0140-6736\(13\)61539-1](https://doi.org/10.1016/S0140-6736(13)61539-1)
- Lewy, A. L., & Dawson, G. (1992). Social stimulation and joint attention in young autistic children. *Journal of Abnormal Child Psychology*, 20(6), 555–566. <https://doi.org/10.1007/BF00911240>
- Little, M. A., Varoquaux, G., Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., & Kording, K. P. (2017). Using and understanding cross-validation strategies. Perspectives on Saeb et al. *GigaScience*, 6(5), 1–6. <https://doi.org/10.1093/gigascience/gix020>
- Liu, W., Li, M., & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, 9(8), 888–898. <https://doi.org/10.1002/aur.1615>
- Lord, C., Risi, S., Forssman, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223. <https://doi.org/10.1023/A:1005592401947>
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659–685. <https://doi.org/10.1007/BF02172145>
- McEvoy, R. E., Rogers, S. J., & Pennington, B. F. (1993). Executive function and social communication deficits in young autistic children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 34(4), 563–578. <https://doi.org/10.1111/j.1469-7610.1993.tb01036.x>
- Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, 20(1), 115–128. <https://doi.org/10.1007/BF02206861>
- Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology*, 6(3), 389–401. <https://doi.org/10.1017/S0954579400006003>
- Murdaugh, D. L., Shinkareva, S. V., Deshpande, H. R., Wang, J., Pennick, M. R., & Kana, R. K. (2012). Differential deactivation during mentalizing and classification of autism based on default mode network connectivity. *PLoS One*, 7(11), e50064. <https://doi.org/10.1371/journal.pone.0050064>
- Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders*, 24(3), 247–257. <https://doi.org/10.1007/BF02172225>
- Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology*, 14(2), 239–251. <https://doi.org/10.1017/s0954579402002031>
- Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., ... Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49(3), 256–266.
- Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain: A Journal of Neurology*, 128(5), 1038–1048. <https://doi.org/10.1093/brain/awh404>
- Povinelli, D. J., & Povinelli, T. J. (1996). Mindblindness. An essay on autism and theory of mind. *Trends in Neurosciences*, 19(7), 299–300. [https://doi.org/10.1016/S0166-2236\(96\)60013-7](https://doi.org/10.1016/S0166-2236(96)60013-7)
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. London, England: Oxford Psychologists Press.
- Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, 9(3), 507–513. <https://doi.org/10.3758/bf03196306>
- Robic, S., Sonié, S., Fonlupt, P., Henaff, M.-A., Touil, N., Coricelli, G., ... Schmitz, C. (2015). Decision-making in a changing world: A study in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 45(6), 1603–1613. <https://doi.org/10.1007/s10803-014-2311-7>
- Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., & Kording, K. P. (2017). The need to approximate the use-case in clinical machine learning. *GigaScience*, 6(5), gix019. <https://doi.org/10.1101/059774>
- Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., & Kording, K. P. (2017). The need to approximate the use-case in clinical machine learning. *GigaScience*, 6(5), 1–9. <https://doi.org/10.1093/gigascience/gix019>
- Sang, B., & Miao, X. (1990). The revision of trail norm of Peabody picture vocabulary test revised (PPVT-R) in Shanghai proper. *Psychological Science*, 14(5), 20–25. (in Chinese)
- Segovia, F., Holt, R., Spencer, M., Górriz, J. M., Ramírez, J., Puntonet, C. G., ... Suckling, J. (2014). Identifying endophenotypes of autism: A multivariate approach. *Frontiers in Computational Neuroscience*, 8, 60. <https://doi.org/10.3389/fncom.2014.00060>
- Senju, A., & Johnson, M. H. (2009). Atypical eye contact in autism: Models, mechanisms and development. *Neuroscience and Biobehavioral Reviews*, 33(8), 1204–1214. <https://doi.org/10.1016/j.neubiorev.2009.06.001>
- Senju, A., Tojo, Y., Dairoku, H., & Hasegawa, T. (2004). Reflexive orienting in response to eye gaze and an arrow in children with and without autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 45(3), 445–458. <https://doi.org/10.1111/j.1469-7610.2004.00236.x>
- Siller, M., & Sigman, M. (2002). The behaviors of parents of children with autism predict the subsequent development of their children's communication. *Journal of Autism and Developmental Disorders*, 32(2), 77–89. <https://doi.org/10.1023/a:1014884404276>

- Stahl, D., Pickles, A., Elsabbagh, M., Johnson, M. H., & Team, B. A. S. I. S. (2012). Novel machine learning methods for ERP analysis: A validation from research on infants at risk for autism. *Developmental Neuropsychology*, 37(3), 274–298. <https://doi.org/10.1080/87565641.2011.650808>
- Stauder, J. E. A., Bosch, C. P. A., & Nuij, H. A. M. (2011). Atypical visual orienting to eye gaze and arrow cues in children with high functioning autism Spectrum disorder. *Research in Autism Spectrum Disorders*, 5(2), 742–748. <https://doi.org/10.1016/j.rasd.2010.08.008>
- Tanaka, J. W., & Sung, A. (2016). The “eye avoidance” hypothesis of autism face processing. *Journal of Autism and Developmental Disorders*, 46(5), 1538–1552. <https://doi.org/10.1007/s10803-013-1976-7>
- Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., & Maurer, R. G. (1998). Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences of the United States of America*, 95(23), 13982–13987. <https://doi.org/10.1073/pnas.95.23.13982>
- Uddin, L. Q., Menon, V., Young, C. B., Ryali, S., Chen, T., Khouzam, A., ... Hardan, A. Y. (2011). Multivariate search-light classification of structural magnetic resonance imaging in children and adolescents with autism. *Biological Psychiatry*, 70(9), 833–841. <https://doi.org/10.1016/j.biopsych.2011.07.014>
- Vance, H. R., & Stone, J. E. (1990). Peabody picture vocabulary test-revised (PPVT-R). *Diagnostique*, 15, 149–160.
- Wall, D. P., Kosmicki, J., Deluca, T. F., Harstad, E., & Fusaro, V. A. (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism. *Translational Psychiatry*, 2, e100. <https://doi.org/10.1038/tp.2012.10>
- Walsh, P., Elsabbagh, M., Bolton, P., & Singh, I. (2011). In search of biomarkers for autism: Scientific, social and ethical challenges. *Nature Reviews. Neuroscience*, 12(10), 603–612. <https://doi.org/10.1038/nrn3113>
- Wang, Q., Hoi, S. P., Wang, Y., Lam, C. M., Fang, F., & Yi, L. (2020). Gaze response to others' gaze following in children with and without autism. *Journal of Abnormal Psychology*, 129(3), 320–329. <https://doi.org/10.1037/abn0000498>
- Webster, S., & Potter, D. D. (2008). Brief report: Eye direction detection improves with development in autism. *Journal of Autism and Developmental Disorders*, 38(6), 1184–1186. <https://doi.org/10.1007/s10803-008-0539-9>
- Wedyan, M., & Al-Jumaily, A. (2016). *Upper limb motor coordination based early diagnosis in high risk subjects for Autism*. Paper presented at the computational intelligence (SSCI), 2016 IEEE symposium series on.
- Wee, C. Y., Wang, L., Shi, F., Yap, P. T., & Shen, D. (2014). Diagnosis of autism spectrum disorders using regional and interregional morphological features. *Human Brain Mapping*, 35(7), 3414–3430. <https://doi.org/10.1002/hbm.22411>
- Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience and Biobehavioral Reviews*, 36(3), 1060–1084. <https://doi.org/10.1016/j.neubiorev.2011.12.008>
- Zhou, Y., Yu, F., & Duong, T. (2014). Multiparametric MRI characterization and prediction in autism spectrum disorder using graph theory and machine learning. *PLoS One*, 9(6), e90405. <https://doi.org/10.1371/journal.pone.0090405>