RESEARCH ARTICLE

Digital Behavioral Phenotyping Detects Atypical Pattern of Facial Expression in Toddlers with Autism

Kimberly L. H. Carpenter ⓘ, Jordan Hahemi, Kathleen Campbell, Steven J. Lippmann, Jeffrey P. Baker, Helen L. Egger, Steven Espinosa, Saritha Vermeer, Guillermo Sapiro, and Geraldine Dawson

Commonly used screening tools for autism spectrum disorder (ASD) generally rely on subjective caregiver questionnaires. While behavioral observation is more objective, it is also expensive, time-consuming, and requires significant expertise to perform. As such, there remains a critical need to develop feasible, scalable, and reliable tools that can characterize ASD risk behaviors. This study assessed the utility of a tablet-based behavioral assessment for eliciting and detecting one type of risk behavior, namely, patterns of facial expression, in 104 toddlers (ASD $N = 22$) and evaluated whether such patterns differentiated toddlers with and without ASD. The assessment consisted of the child sitting on his/her caregiver's lap and watching brief movies shown on a smart tablet while the embedded camera recorded the child's facial expressions. Computer vision analysis (CVA) automatically detected and tracked facial landmarks, which were used to estimate head position and facial expressions (Positive, Neutral, All Other). Using CVA, specific points throughout the movies were identified that reliably differentiate between children with and without ASD based on their patterns of facial movement and expressions (area under the curves for individual movies ranging from 0.62 to 0.73). During these instances, children with ASD more frequently displayed Neutral expressions compared to children without ASD, who had more All Other expressions. The frequency of All Other expressions was driven by non-ASD children more often displaying raised eyebrows and an open mouth, characteristic of engagement/interest. Preliminary results suggest computational coding of facial movements and expressions via a tablet-based assessment can detect differences in affective expression, one of the early, core features of ASD. *Autism Res* 2020, 00: 1–12. © 2020 International Society for Autism Research and Wiley Periodicals LLC

Lay Summary: This study tested the use of a tablet in the behavioral assessment of young children with autism. Children watched a series of developmentally appropriate movies and their facial expressions were recorded using the camera embedded in the tablet. Results suggest that computational assessments of facial expressions may be useful in early detection of symptoms of autism.

Keywords: autism; risk behaviors; facial expressions; computer vision; early detection

Introduction

Autism spectrum disorder (ASD) can be reliably diagnosed as early as 24 months old and the risk signs can be detected as early as 6–12 months old [Dawson & Bernier, 2013; Luyster et al., 2009]. Despite this, the average age of diagnosis in the United States remains around 4 years of age [Christensen et al., 2016]. While there is mixed evidence for the stability of autism traits over early childhood [Bieleninik et al., 2017; Waizbard-Bartov et al., 2020], the delay in diagnosis can still impact timely intervention during a critical window of development. In response to this, in 2007 the American Academy of Pediatrics published guidelines supporting the need for all children to be screened for ASD between 18- and 24-months of age as part of their well-child visits [Myers, Johnson, & Council on Children With Disabilities, 2007]. Current screening typically relies on caregiver report, such as the Modified Checklist of ASD in Toddlers—Revised with Follow-up (M-CHAT-R/F) (Robins et al., 2014). Evidence suggests that a two-tiered screening approach, including direct observational assessment of the child, improves the positive predictive value of M-CHAT screening by 48% [Khowaja, Robins, &

Adamson, 2017] and may reduce ethnic/racial disparities in general screening [Guthrie et al., 2019]. Current tools for observational assessment of ASD signs in infants and toddlers, such as the Autism Observation Scale for Infants (AOSI) and Autism Diagnostic Observation Schedule (ADOS), take substantial time and training to administer, resulting in a shortage of qualified diagnosticians to perform these observational assessments. As such, there remains a critical need to develop feasible, scalable, and reliable tools that can characterize ASD risk behaviors and identify those children who are most in need of follow-up by an ASD specialist. In an effort to address this critical need, we have embarked on a program of research using computer vision analysis (CVA) to develop tools for digitally phenotyping early emerging risk behaviors for ASD [Dawson & Sapiro, 2019]. If successful, such digital screening tools have the opportunity to help existing practitioners reach more children and assist in triaging boundary cases for review by specialists.

One of the early emerging signs of ASD is a tendency to more often display a neutral facial expression. This pattern is evident in the quality of facial expressions and in sharing emotional expressions with others [Adrien et al., 1993; Baranek, 1999; S. Clifford & Dissanayake, 2009; S. Clifford, Young, & Williamson, 2007; S. M. Clifford & Dissanayake, 2008; Maestro et al., 2002; Osterling, Dawson, & Munson, 2002; Werner, Dawson, Osterling, & Dinno, 2000]. A restricted range of emotional expression and its integration with eye gaze (e.g., during social referencing) have been found to differentiate children with ASD from typically developing children, as well as those who have other developmental delays, as early as 12 months of age [Adrien et al., 1991; S. Clifford et al., 2007; Filliter et al., 2015; Gangi, Ibanez, & Messinger, 2014; Nichols, Ibanez, Foss-Feig, & Stone, 2014]. While core features of ASD vary by age, cognitive ability, and language, one of the most stable symptoms from early childhood through adolescences is increased frequency of neutral expression [Bal, Kim, Fok, & Lord, 2019]. As such, differences in facial affect may show utility in assessing early risk for ASD.

A recent meta-analysis of facial expression production in autism found that individuals with ASD display facial expressions less often than non-ASD participants and that, when they did display facial expressions, the expressions occurred for shorter durations and were of different quality than non-ASD individuals (Trevisan, Hoskyn, & Birmingham, 2018). Decreases in the frequency of both emotional facial expressions and the sharing of those expressions with others has been demonstrated across naturalistic interactions [Bieberich & Morgan, 2004; Czapinski & Bryson, 2003; Dawson, Hill, Spencer, Galpert, & Watson, 1990; Mcgee, Feldman, & Chernin, 1991; Snow, Hertzig, & Shapiro, 1987; Tantam,

Holmes, & Cordess, 1993], in lab-based assessments such as during the ADOS [Owada et al., 2018] or the AOSI [Filliter et al., 2015] and in response to emotion-eliciting videos [Trevisan, Bowering, & Birmingham, 2016]. Furthermore, higher frequency of neutral expressions correlates with social impairment in children with ASD [Owada et al., 2018] and differentiates them from children with other delays [Bieberich & Morgan, 2004; Yirmiya, Kasari, Sigman, & Mundy, 1989]. As such, frequency and duration of facial affect is a promising early risk marker for young children with autism.

Previous research on atypical facial expressions in children with ASD has relied on hand coding of facial expressions, which is time intensive and often requires significant training [Bieberich & Morgan, 2004; S. Clifford et al., 2007; Dawson et al., 1990; Gangi et al., 2014; Mcgee et al., 1991; Nichols et al., 2014; Snow et al., 1987]. This approach is not scalable for use in general ASD risk screening or as a behavioral biomarker or outcome assessment for use in large clinical trials. As such, the field has moved toward automating the coding of facial expressions. In one of the earliest studies of this approach, Guha and colleagues demonstrated that children with ASD have atypical facial expressions when mimicking others. However, their technology required the children to wear markers on their face for data capture [Guha, Yang, Grossman, & Narayanan, 2018; Guha et al., 2015], which is both invasive and not scalable. More recently, several groups have applied non-invasive CVA technology to measuring affect in older children and adults with ASD within the laboratory setting [Capriola-Hall et al., 2019; Owada et al., 2018; Samad et al., 2018]. This represents an important move toward scalability as CVA approaches do not rely on the presence of physical markers on the face to extract emotion information. Rather, CVA relies on videos of the individual in which features around specific regions on a face (e.g., the mouth and eyes) are extracted. Notably, these features mirror those used by the manually rated facial affect coding system (FACS) [Ekman, 1997]. Both our earlier work [Hashemi et al., 2018] and that of others [Capriola-Hall et al., 2019] have shown good concordance between human coding and CVA rating of facial emotions. Furthermore, previous research in adults has demonstrated that CVA can detect neutral facial expressions more reliably than human coders [Lewinski, 2015].

Building on previous work applying CVA in laboratory settings, we have developed a portable tablet-based technology that uses the embedded camera and automatic CVA to code ASD risk behaviors in <10 min across a range of non-laboratory settings (e.g., pediatric clinics, at home, etc.). We developed a series of movies designed to capture children's attention, elicit emotion in response to novel and interesting events, and assess the toddler's ability to sustain attention and share it with others. By embedding

these movies in a fully automated system on a cost-effective tablet whereby the elicited behaviors, in this case the frequency of different patterns of facial affect, are automatically encoded with CVA, we aim to create a tool that is objective, efficient, and accessible. The current analysis focuses on preliminary results supporting the utility of this tablet-based assessment for the detection of facial movement and affect in young children and the use of facial affect to differentiate children with and without ASD. Though facial affect is the focus of the current analysis, the ultimate goal is to combine information across autism risk features collected through the current digital screening tool [e.g., delayed response to name as described in Campbell et al., 2019], to develop a risk score based on multiple behaviors [Dawson & Sapiro, 2019]. This information could then be combined with additional measures of risk to enhance screening for ASD.

Methods

Participants

Participants were 104 children 16–31 months of age (Table 1). Children were recruited at their pediatric primary care visit by a research assistant embedded within the clinic or via referral from their physician, as well as through community advertisement ($N = 4$ in the non-ASD group and $N = 15$ in the ASD group). For children recruited within the pediatric clinics, recruitment occurred at the 18- or 24-month well-child visit at the same time as they received standard screening for ASD with the M-CHAT-R/F. A total of 76% of the participants recruited in the clinic by a research assistant chose to participate. Of the participants who chose not to participate, 11% indicated that they were not interested in the study,

whereas the remainder declined due to not having enough time, having another child to take care of, wanting to discuss with their partner, or their child was already too distressed after the physician visit. All children who enrolled in the study found the procedure engaging enough that they were able to provide adequate data for analysis. Because the administration is very brief and non-demanding, data loss was not a significant problem.

Exclusionary criteria included known vision or hearing deficits, lack of exposure to English at home, or insufficient English language skills for caregiver's informed consent. Twenty-two children were diagnosed with ASD. The non-ASD comparison group ($N = 82$) was comprised of 74 typically developing children and 8 children with a non-ASD delay, which was defined by a diagnosis of language delay or developmental delay of clinical significance sufficient to qualify for speech or developmental therapy as recorded in the electronic medical record. All caregivers/legal guardians gave written informed consent, and the study protocol was approved by the Duke University Health System IRB.

Children recruited from the pediatric primary care settings received screening with a digital version of the M-CHAT-R/F as part of a quality improvement study ongoing in the clinic [Campbell et al., 2017]. Participants recruited from the community received ASD screening with the digital M-CHAT-R/F prior to the tablet assessment. As part of their participation in the study, children who either failed the M-CHAT-R/F or for whom there was caregiver or physician concern about possible ASD underwent diagnostic testing using the ADOS-Toddler (ADOS-T) [Luyster et al., 2009] conducted by a licensed psychologist or research-reliable examiner supervised by a licensed

Table 1. Sample Demographics

	Typically developing (N = 74; 71%)	Non-ASD delay (N = 8; 8%)	ASD (N = 22; 21%)
Age			
Months [mean (SD)]	21.7 (3.8)	23.9 (3.7)	26.2 (4.1)
Sex			
Female	31 (42)	3 (38)	5 (23)
Male	43 (58)	5 (62)	17 (77)
Ethnicity/race			
African American	10 (14)	1 (13)	3 (14)
Caucasian	46 (62)	2 (25)	10 (45)
Hispanic	1 (1)	0 (0)	1 (4)
Other/unknown	17 (23)	5 (62)	8 (37)
Insurance[a]			
Medicaid	11 (15)	1 (14)	6 (67)
Non-Medicaid	60 (85)	6 (86)	3 (33)
MCHAT result[b]			
Positive	1 (1)	0 (0)	18 (82)
Negative	73 (99)	8 (100)	4 (18)

[a]Insurance status was unknown for 17 (16%) of participants in this study.

[b]Children for whom the MCHAT was negative but received and ASD diagnosis were referred for assessment due to concerns by either the parent or the child's physician.

psychologist. The mean ADOS-T score was 18.00 (SD = 4.67). A subset of the ASD children (*N* = 13) also received the Mullen Scales of Early Learning (Mullen, 1995). The mean IQ based on the Early Learning Composite Score for this subgroup was 63.58 (SD = 25.95). None of the children in the non-ASD comparison group was administered the ADOS or Mullen.

Children's demographic information was extracted from the child's medical record or self-reported by the caregiver at their study visit. Children in the ASD group were, on average, 4 months younger than the comparison group (*t*[102] = 4.64, *P* < 0.0001). Furthermore, as would be expected, there was a higher proportion of males in the ASD group than in the comparison group, though the difference was not statistically significant (χ^2 [1, 104] = 2.60, *P* = 0.11). There were no differences in the proportion of racial/ethnic minority children between the two groups (χ^2 [1, 104] = 1.20, *P* = 0.27). When looking only at the children for which Medicaid status was known, there was no difference in the proportion of children on Medicaid in the ASD and the non-ASD group (χ^2 [1, 87] = 1.82, *P* = 0.18).

Stimuli and Procedure

A series of developmentally appropriate brief movies designed to elicit affect and engage the child's attention were shown on a tablet while the child sat on a caregiver's lap. The tablet was placed on a stand approximately 3 ft away from the child to prevent the child from touching the screen as depicted in previous publications [Campbell et al., 2019; Dawson et al., 2018; Hashemi et al., 2015; Hashemi et al., 2018]. Movies consisted of cascading bubbles (2 × 30 sec), a mechanical bunny (66 sec), animal puppets interacting with each other (68 sec), and a split screen showing a woman singing nursery rhymes on one side and dynamic, noise-making toys on the other side (60 sec; Fig. 1). These movies included stimuli that have been used in previous studies of ASD symptomatology [Murias et al., 2018], as well as developed specifically for the current tablet-based technology to elicit autism symptoms, based on Dawson et al. [2004], Jones, Dawson, Kelly, Estes, and Webb [2017], Jones et al. [2016], and Luyster et al. [2009]. At three points during the movies, the examiner located behind the child called the child's name.

Prior to the administration of the app, caregivers were clearly instructed not to direct their child's attention or in any way try to influence the child's behavior during the assessment. Furthermore, if the caregiver began to try to direct their child's attention, the examiner in the room immediately asked the caregiver to refrain from doing so. If the caregiver persisted, this was noted on our validity form and the administration would have been considered invalid. Researchers stopped the task for one comparison participant due to crying. Researchers restarted the task for three participants with ASD due to difficulty remaining in view of the tablet's camera for more than half of the first video stimulus. If other family members were present during the well-child visit, they were asked to stand behind the caregiver and child so as to not distract the child during the assessment. Additionally, for children assessed during a well-child visit, research assistants were instructed to collect data prior to any planned shots or blood draws.

Computer Vision Analysis

The frontal camera in the tablet recorded video throughout the experiment at 1280 × 720 resolution and 30 frames per second. The CVA algorithm [Hashemi et al., 2018] first automatically detected and tracked 49 facial landmarks on the child's face [De la Torre et al., 2015]. Head positions relative to the camera were estimated by computing the optimal rotation parameters between the detected landmarks and a 3D canonical face model [Fischler & Bolles, 1981]. A "not visible" tag was assigned to frames where the face was not detected or the face exhibited drastic yaw (>45° from center). We acknowledge that the current method used only indicates whether the child is oriented toward the stimulus and does not track eye movements. For each "visible" frame, the probability of expressing three standard categories of facial expressions, Positive, Neutral (i.e., no active facial

Figure 1. Example of movie stimuli. Developmentally appropriate movies consisted of cascading bubbles, a mechanical bunny, animal puppets interacting with each other, and a split screen showing a woman singing nursery rhymes on one side and dynamic, noise-making toys on the other side.

action unit), or Other (all other expressions), was assigned [Hashemi et al., 2018]. The model for automatic facial expression is an extension of the pose-invariant and cross-modal dictionary learning approach originally described in Hashemi et al. [2015]. During training, the dictionary representation is setup to map facial information between 2D and 3D modalities and is then able to infer discriminative facial information for facial expressions recognition even when only 2D facial images are available at deployment. For training, data from Binghamton University 3D Facial Expression database [Yin, Wei, Sun, Wang, & Rosato, 2006] were used, along with synthesized faces images with varying poses [see Hashemi et al., 2015 for synthesis details]. Extracted image features and distances between a subset of facial landmarks were used as facial features to learn the robust dictionary. Lastly, using the inferred discriminative 3D and frontal 2D facial features, a multiclass support vector machine [Chang & Lin, 2011] was trained to classify the different facial expressions.

In recent years, there has been progress on automatic facial expression analysis of both children and toddlers [Dys & Malti, 2016; Gadea, Aliño, Espert, & Salvador, 2015; Haines et al., 2019; LoBue & Thrasher, 2014; Messinger, Mahoor, Chow, & Cohn, 2009]. In addition to this, we have previously validated our CVA algorithm against expert human rater coding of facial affect in a subsample of 99 video recordings across 33 participants (ASD = 15, non-ASD = 18). This represents 20% of the non-ASD sample and a matched group from the ASD sample. The selection of participants for this previously published validation study was based on age distribution to ensure representation across the range of ages for both non-ASD and ASD groups. This previous work showed strong concordance between CVA- and human-rated coding of facial emotion in this data set, with high precision, recall, and $F1$ scores of 0.89, 0.90, and 0.89, respectively [Hashemi et al., 2018].

Statistical Approach

For each video frame, the CVA algorithm produces a probability value for each expression (Positive, Neutral, Other). We calculated the mean of the probability values for each of the three expression types within non-overlapping 90-frame (3 sec) intervals, excluding frames when the face was not visible. A 3-sec window was selected as it provided us with a continuous distribution of the emotion probabilities, while still being within the 0.5–4-sec window of a macroexpression (Ekman, 2003). Additionally, for each of the name call events we removed the time window starting at cue for the name call prompt through the point where 75% of the audible name calls actually occurred, plus 150 frames (5 sec). This window was selected based on our previous study

[Campbell et al., 2019] showing that orienting tended to occur within a few seconds after a name call. We calculated the proportion of frames the child was not attending to the movie stimulus, based on the "visible" and "not visible" tags described above, within each 90-frame interval, excluding name call periods. Thus, for each child, we generated four variables (mean probabilities of Positive, Neutral, or Other; and proportion of frames not attending) for each 90-frame interval within each of the five movies.

To evaluate differences between ASD and non-ASD children at regular intervals throughout the assessment, we fit a series of bivariate logistic regressions to obtain the odds ratios for the associations between the mean expression probability or attending proportion during a given interval, parameterized as increments of 10% points, and ASD diagnosis. We then fit a series of multivariable logistic regression models, separately for each movie and variable, which included parameters for each of the 3-sec intervals within the movie to predict ASD diagnosis. Given the large number of intervals relative to the small sample size, we used a Least Absolute Shrinkage and Selection Operator (LASSO) penalized regression approach [Tibshirani, 1996] to select a parsimonious set of parameters representing the intervals within each movie and expression type that were most predictive of ASD diagnosis.

For each of the five movies, we then combined the LASSO-selected interval parameters into a full logistic model. When more than one expression parameter was selected for a given interval, we selected the one with the stronger odds ratio estimate. Analyses were conducted with and without age as a covariate. Since the small study size precluded having separate training and validation sets, we used leave-one-out cross-validation to assess model performance. Receiver-operating characteristics (ROC) curves were plotted and the *c*-statistic for the area under the ROC curve was calculated for each movie.

Results

Figure 2 depicts the odds ratio analysis using the "Rhymes and Toys" movie as one illustrative example. As shown by the variability in the odds ratio estimates, some parts of the movies elicited strongly differential responses in certain patterns of expression (blue window, Fig. 2), while in other sections, there were not substantial differences between the two groups (green window, Fig. 2). Overlaid on the plot are the odds ratios and confidence bands for the interval parameters selected by the expression-specific LASSO models. These selected parameters were then used in the movie-level logistic models for which we calculated classification metrics.
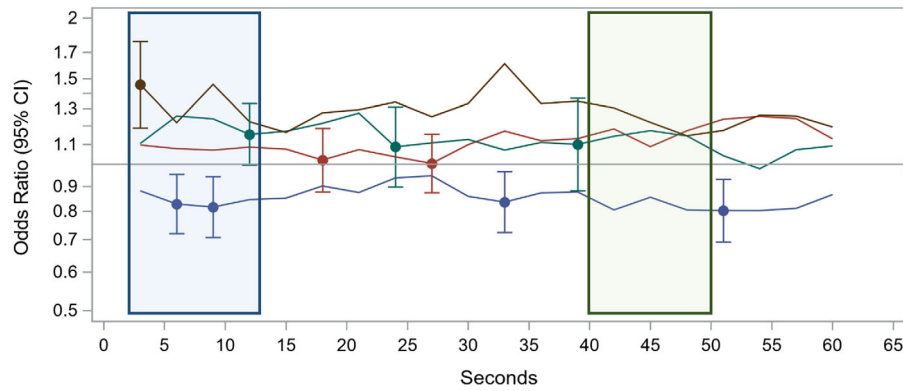
Figure 2. Time series of odds ratios for the association between the mean expression probability or proportion attending and ASD diagnosis. Using the "Rhymes" movie as one illustrative example, lines depict the odds of meeting criteria for ASD (OR > 1) or being in the non-ASD comparison group (OR < 1) for each of the outcomes of interest for each 3 sec time bin across the movie. Points with error bars are intervals that were selected by the LASSO regression models and included in the final logistic model. The blue window depicts a segment of the movie where there were differential emotional responses between the ASD and non-ASD children. The green window depicts a segment of the movie in which there was no difference in emotional responses between the groups.
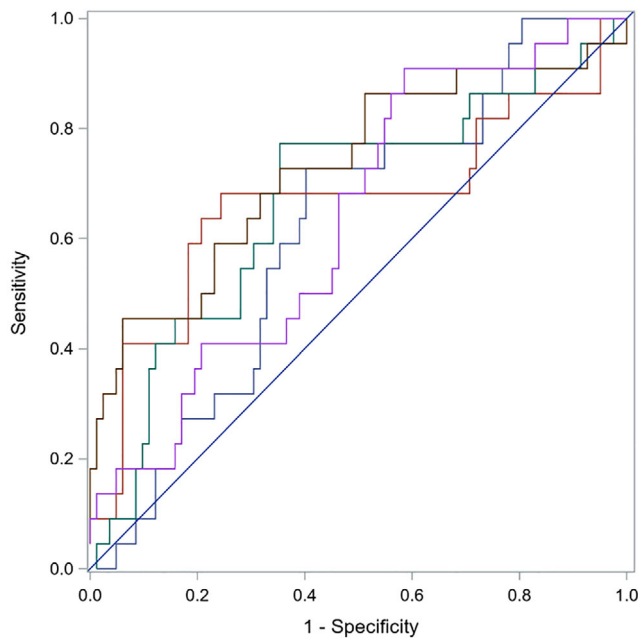

Figure 3. Receiver-operating characteristics (ROC) curves. ROC curves were calculated for predictive ability of expression-specific LASSO selected interval parameters for facial expressions and attention to stimulus for each movie independently.

Figure 3 compares the ROC curves for the five final movie-level logistic models after leave-one-out cross-validation. ROC curves analyses were performed for each video individually. The model for the "Rhymes" movie yielded the strongest predictive ability, with an area under the curve (AUC) of 0.73 (95% confidence interval [CI] 0.59–0.86), followed by the "Puppets" (AUC = 0.67; 95% CI 0.53–0.80) and the "Bunny" (AUC = 0.66; 95%

CI 0.51–0.82) videos. Finally, the two "Bubbles" movies that bookend the stimulus sets were the least predictive, with AUCs of 0.62 (95% CI 0.49–0.74) and 0.64 (95% CI 0.51–0.76), respectively. Because there was a significant difference in age between the ASD and non-ASD comparison groups, we ran a second set of ROC analyses where age was included as a covariate, shown in Table 2. Results remained significant after including the age covariate.

Given the preponderance for the Other emotional category in our non-ASD comparison group, we explored: (a) what specific facial movements are driving this category of Other expressions and (b) how it differs from the Neutral expression category. We focused on analyzing movements of facial landmarks and head pose angles and how they differ between the facial expression categories. Our CVA algorithm aligns the facial landmarks of the child to a canonical face model through an affine transformation, which normalizes the landmark locations across all video frames to a common space. This normalization process is commonly used across CVA tasks related to facial analysis as it allows one to analyze/compare landmark locations across different frames or participants. With this alignment step, we were able to quantify the distances between the eye corners and the corners of the eyebrows, the vertical distance between the inner lip points, and the vertical distance between the outer lip points (Fig. 4, right). To interpret features that differentiated the Neutral from the Other facial expression category, we assessed differences between these facial landmark distances of a given child when they were predominately expressing Neutral versus Other facial expressions. We also included yaw and pitch head pose angles since they may play a role in the alignment process.

Table 2. Comparison of ASD Versus Non-ASD: Area Under the Curve (AUC) Analyses

	AUC without covariates	AUC with age in the model
Bubbles 1	0.62	0.75
Bunny	0.66	0.81
Puppets	0.67	0.78
Rhymes	0.73	0.83
Bubbles 2	0.64	0.79

We focused in on three stimuli in which participants exhibited high probabilities of Other expressions, namely, the first Bubbles, Puppets, and Rhymes and Toys videos. Out of the 104 participants, all exhibited frames where both Neutral and Other facial expressions were dominant (probability of expression over 60%). A Wilson signed-ranks test, reported as (median difference, P value), indicated that within individual participants for each diagnostic group, the median differences between the distances of Other versus Neutral facial expressions were significantly higher for inner right eyebrows (non-ASD: diff = 2.5, $P < 8.1e-10$; ASD: diff = 4.0, $P < 1.1e-4$), inner left eyebrows (non-ASD: diff = 2.4, $P < 1.5e-9$; ASD: diff = 3.6, $P < 1.4e-4$), outer right eyebrows (non-ASD: diff = 1.2, $P < 1.0e-5$; ASD: diff = 2.3, $P < 1.6e-4$), outer left eyebrows (non-ASD: diff = 0.9, $P < 3.4e-6$; ASD: diff = 1.4, $P < 1.4e-3$), and mouth height (non-ASD: diff = 1.5, $P < 1.0e-3$), as well as for pitch head pose angles (non-ASD: diff = 2.8, $P < 8.7e-10$; ASD: diff = 4.4, $P < 1.2e-3$); but not for eye heights, lips parting, nor yaw head pose angles.

Discussion

The present study evaluated an application administered on a tablet that was comprised of carefully designed movies that elicited affective expressions combined with CVA of recorded behavioral responses to identify patterns of facial movement and emotional expression that differentiate toddlers with ASD from those without ASD. We demonstrated that the movies elicited a range of affective facial expressions in both groups. Furthermore, using CVA we found children with ASD were more likely to display a neutral expression than children without ASD when watching this series of videos, and the patterns of facial expressions elicited during specific parts of the movies differed between the two groups. We believe this finding has strong face validity that rests on both research and clinical observations of a restricted range of facial expression in children with autism. Furthermore, this replicates a previous finding of our group reporting increased frequency of neutral expression in young children who screened positive on the M-CHAT [Egger et al., 2018]. Together, these preliminary results

support the use of engaging brief movies shown on a cost-effective tablet, combined with automated CVA behavioral coding, as an objective and feasible tool for measuring an early emerging symptom of ASD, namely, increased frequency of neutral facial expressions.

While the predictive power of emotional expression in some of the videos varied, all but one represent a medium effect, equivalent to Cohen's $d = 0.5$ or greater [Rice & Harris, 2005]. Overall, the best predictor from our battery of videos is the "Rhymes" video, which had an AUC with a large effect size (equivalent to $d > 0.8$). While this may suggest that presenting the "Rhymes" video alone is sufficient for differentiating between the ASD and non-ASD groups, we caution readers from coming to this conclusion for two reasons: First, it is possible that, had we had a larger sample, the other videos would have had a larger effect. Second, we anticipate that there will be variability in the ASD group with regard to which features a single child will express and different videos may be better suited to elicit different features in any given individual. As such, we believe that it is important to understand how each independent feature, in this case facial affect, performs across the different videos so that we can being to build better predictive models from combinations of features [e.g., facial affect and postural sway as described in Dawson et al., 2018].

To further understand the difference of facial expression in the non-ASD group as compared to our ASD sample, we explored the facial landmarks differentiating between the Other facial expression category that dominated the non-ASD control group versus the Neutral facial expression, which was more common in the ASD group. Through this analysis, we identified the features of raised eyebrows and open mouth to play a role in discriminating between the Other vs. Neutral categories. This facial pattern is consistent with an engaged/interested look displayed when a child is actively watching, as described in young children by Sullivan and Lewis [2003]. It is interesting to note a raised pitch angle was also statistically significant. Since the median difference of this angle between the two facial expression is small (3.2°), this may be a natural movement of raising one's eyebrows.

Our results need to be considered in light of several limitations. First, the CVA models of facial expressions used in the current study were trained on adult faces [Hashemi et al., 2018]. Despite this, our previous findings with young children demonstrate good concordance between human and CVA coding on the designation of facial expressions [Hashemi et al., 2018]. Furthermore, the Other facial expression category includes all non-positive or negative expressions. As such, even though we were able to determine the predominant feature driving those expressions was the raised eyebrows, which is in line with our observations from watching the movies,
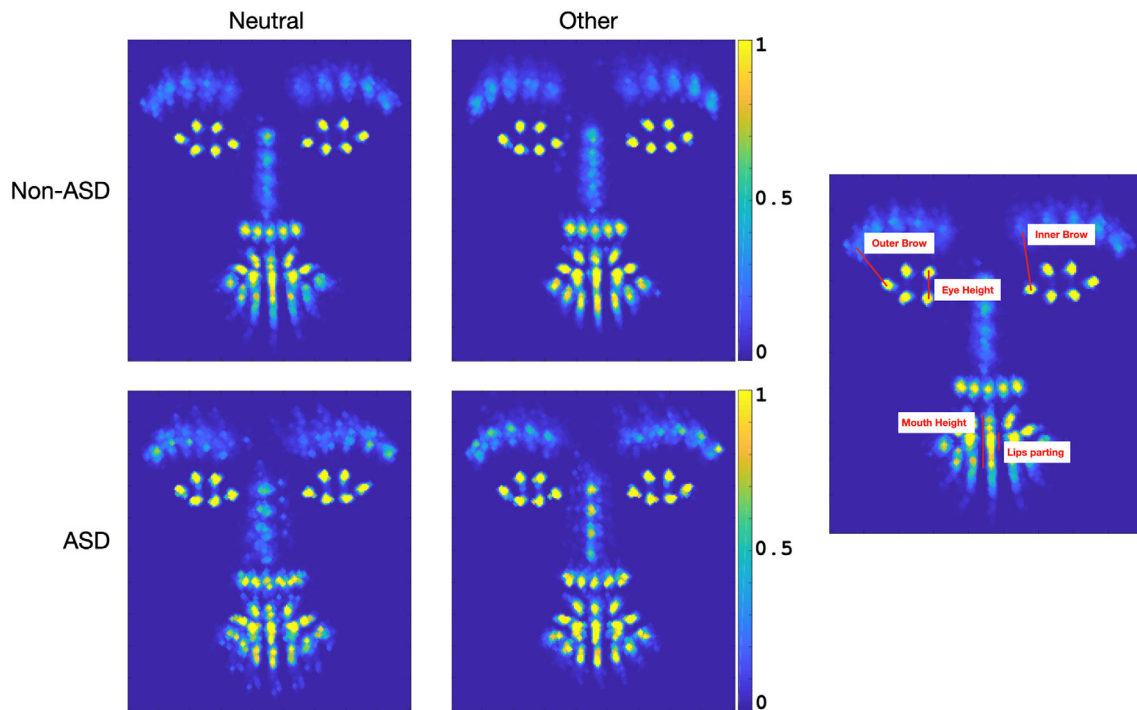
Figure 4. Analysis of Other Facial Expression. The 4 panels on the left depict heat maps of aligned landmarks across ASD and non-ASD participants when they were exhibiting Neutral and other facial expressions (the color bar indicates proportion of frames where landmarks were displayed in a given image location). The single panel on the right is an example of the landmark distances explored.

it is possible that there are a combination of facial expressions in the non-ASD group driving this designation. Future studies will need to train on the engaged/interested facial expression specifically and test the robustness of this finding. Additionally, though we have previously demonstrated good reliability between our CVA algorithms and human coding of emotions [Hashemi et al., 2015], future validation of our CVA analysis of emotional facial expressions in larger datasets is currently underway. Second, using the LASSO statistical approach means our model may not select all features that have differentiating information. However, we selected this approach because it minimizes over fitting the model. Third, our sample size was relatively small and we do not have separate training and testing samples. To account for this, we applied cross-validation on the ROC curves even though this decreased the performance metrics of the model. This suggests that our ROC results are potentially conservative. Finally, our comparison group contains both children with typical development and children with non-ASD developmental delays, a factor that can be viewed as both a weakness and a strength. Previous research has demonstrated that increased frequency of neutral expressions does differentiate children with ASD from those with other developmental delays [Bieberich & Morgan, 2004; Yirmiya et al., 1989]. However, due to the small sample size of children with non-

ASD developmental delays, we were unable to directly test this in our data. Furthermore, because only a subset of the sample received an assessment of cognitive ability, it is possible that there were additional children in the non-ASD comparison group that also had a developmental delay that was undetected. Ongoing research in a prospective, longitudinal study with larger samples is underway to further parse the ability of our CVA tools to differentiate between children with ASD, children with a non-ASD developmental delay and/or attention deficit hyperactivity disorder, and typically developing children.

While a difference in facial expression is one core feature of ASD, the heterogeneity in ASD means we do not expect all children with ASD to display this sign of ASD. As such, our next step is to combine the current results with other measures of autism risk assessed through the current digital screening tool, including response to name [Campbell et al., 2019], postural sway [Dawson et al., 2018], and differential vocalizations [Tenenbaum et al., 2020], among other features, to develop a risk score based on multiple behaviors [Dawson & Sapiro, 2019]. Since no one child is expected to display every risk behavior, a goal is to determine thresholds based on the total number of behaviors, regardless of which combination of behaviors, to asses for risk. This is similar to what is done in commonly used screening and diagnostic tools, such as the M-CHAT [Robins et al., 2014], Autism Diagnostic

Interview [Lord, Rutter, & Le Couteur, 1994], and ADOS [Gotham, Risi, Pickles, & Lord, 2007; Lord et al., 2000].

In summary, we evaluated an integrated, objective tool for the elicitation and measurement of facial movements and expressions in toddlers with and without ASD. The current study adds to a body of research supporting digital behavioral phenotyping as a viable method for assessing autism risk behaviors. Our goal is to further develop and validate this tool so that it can eventually be used within the context of current standard of care to enhance autism screening in pediatric populations.

Acknowledgments

Conflict of interests

Guillermo Sapiro has received basic research gifts from Amazon, Google, Cisco, and Microsoft and is a consultant for Apple and Volvo. Geraldine Dawson is on the Scientific Advisory Boards of Janssen Research and Development, Akili, Inc., LabCorp, Inc., Tris Pharma, and Roche Pharmaceutical Company, a consultant for Apple, Inc, Gerson Lehrman Group, Guidepoint, Inc., Teva Pharmaceuticals, and Axial Ventures, has received grant funding from Janssen Research and Development, and is CEO of DASIO, LLC (with Guillermo Sapiro). Dawson receives royalties from Guilford Press, Springer, and Oxford University Press. Dawson, Sapiro, Carpenter, Hashemi, Campbell, Espinosa, Baker, and Egger helped develop aspects of the technology that is being used in the study. The technology has been licensed and Dawson, Sapiro, Carpenter, Hashemi, Espinosa, Baker, Egger, and Duke University have benefited financially.

References

Adrien, J. L., Faure, M., Perrot, A., Hameury, L., Garreau, B., Barthelemy, C., & Sauvage, D. (1991). Autism and family home movies: Preliminary findings. Journal of Autism and Developmental Disorders, 21(1), 43–49.

Adrien, J. L., Lenoir, P., Martineau, J., Perrot, A., Hameury, L., Larmande, C., & Sauvage, D. (1993). Blind ratings of early symptoms of autism based upon family home movies. Journal of the American Academy of Child and Adolescent Psychiatry, 32(3), 617–626. https://doi.org/10.1097/00004583-199305000-00019

Bal, V. H., Kim, S. H., Fok, M., & Lord, C. (2019). Autism spectrum disorder symptoms from ages 2 to 19 years: Implications for diagnosing adolescents and young adults. Autism Research, 12(1), 89–99. https://doi.org/10.1002/aur.2004

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.

Bieberich, A. A., & Morgan, S. B. (2004). Self-regulation and affective expression during play in children with autism or Down syndrome: A short-term longitudinal study. Journal of Autism and Developmental Disorders, 34(4), 439–448. https://doi.org/10.1023/b:jadd.0000037420.16169.28

Bieleninik, Ł., Posserud, M.-B., Geretsegger, M., Thompson, G., Elefant, C., & Gold, C. (2017). Tracing the temporal stability of autism spectrum diagnosis and severity as measured by the Autism Diagnostic Observation Schedule: A systematic review and meta-analysis. PLoS One, 12(9), e0183160–e0183160. https://doi.org/10.1371/journal.pone.0183160

Campbell, K., Carpenter, K. L., Hashemi, J., Espinosa, S., Marsan, S., Borg, J. S., … Dawson, G. (2019). Computer vision analysis captures atypical attention in toddlers with autism. Autism, 23(3), 619–628. https://doi.org/10.1177/1362361318766247

Campbell, K., Carpenter, K. L. H., Espinosa, S., Hashemi, J., Qiu, Q., Tepper, M., … Dawson, G. (2017). Use of a Digital Modified Checklist for Autism in Toddlers—Revised with follow-up to improve quality of screening for autism. The Journal of Pediatrics, 183, 133–139.e1. https://doi.org/10.1016/j.jpeds.2017.01.021

Capriola-Hall, N. N., Wieckowski, A. T., Swain, D., Tech, V., Aly, S., Youssef, A., … White, S. W. (2019). Group differences in facial emotion expression in autism: Evidence for the utility of machine classification. Behavior Therapy, 50(4), 828–838. https://doi.org/10.1016/j.beth.2018.12.004

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 1–27.

Christensen, D. L., Baio, J., Braun, K. V., Bilder, D., Charles, J., Constantino, J. N., … Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. MMWR Surveillance Summaries, 65(SS-3), 1–23.

Clifford, S., & Dissanayake, C. (2009). Dyadic and triadic behaviours in infancy as precursors to later social responsiveness in young children with autistic disorder. Journal of Autism and

Developmental Disorders, 39(10), 1369–1380. https://doi.org/10.1007/s10803-009-0748-x

Clifford, S., Young, R., & Williamson, P. (2007). Assessing the early characteristics of autistic disorder using video analysis. Journal of Autism and Developmental Disorders, 37(2), 301–313. https://doi.org/10.1007/s10803-006-0160-8

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

Czapinski, P., & Bryson, S. (2003). Reduced facial muscle movements in autism: Evidence for dysfunction in the neuromuscular pathway? Brain and Cognition, 51(2), 177–179.

Dawson, G., & Bernier, R. (2013). A quarter century of progress on the early detection and treatment of autism spectrum disorder. Development and Psychopathology, 25(4 Pt 2), 1455–1472. https://doi.org/10.1017/S0954579413000710

Dawson, G., Campbell, K., Hashemi, J., Lippmann, S. J., Smith, V., Carpenter, K., … Sapiro, G. (2018). Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. Scientific Reports, 8 (1), 17008. https://doi.org/10.1038/s41598-018-35215-8

Dawson, G., Hill, D., Spencer, A., Galpert, L., & Watson, L. (1990). Affective exchanges between young autistic-children and their mothers. Journal of Abnormal Child Psychology, 18 (3), 335–345. https://doi.org/10.1007/Bf00916569

Dawson, G., & Sapiro, G. (2019). Potential for digital behavioral measurement tools to transform the detection and diagnosis of autism spectrum disorder. JAMA Pediatrics, 173(4), 305–306. https://doi.org/10.1001/jamapediatrics.2018.5269

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

De la Torre, F., Chu, W. S., Xiong, X., Vicente, F., Ding, X., & Cohn, J. (2015). IntraFace. IEEE International Conference on Automatic Face Gesture Recognition Workshops. (pp. 1–8). Ljubljana, Slovenia: IEEE. https://doi.org/10.1109/fg.2015.7163082

Dys, S. P., & Malti, T. (2016). It's a two-way street: Automatic and controlled processes in children's emotional responses to moral transgressions. Journal of Experimental Child Psychology, 152, 31–40. https://doi.org/10.1016/j.jecp.2016.06.011

Egger, H. L., Dawson, G., Hashemi, J., Carpenter, K. L. H., Espinosa, S., Campbell, K., … Sapiro, G. (2018). Automatic emotion and attention analysis of young children at home: A ResearchKit autism feasibility study. NPJ Digital Medicine, 1 (1), 20. https://doi.org/10.1038/s41746-018-0024-6

Ekman, P. (2003). Emotions revealed (2nd ed.). New York, NY: Times Books.

Ekman, R. (1997). What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). New York, NY: Oxford University Press.

Filliter, J. H., Longard, J., Lawrence, M. A., Zwaigenbaum, L., Brian, J., Garon, N., … Bryson, S. E. (2015). Positive affect in infant siblings of children diagnosed with autism spectrum disorder. Journal of Abnormal Child Psychology, 43(3), 567–575. https://doi.org/10.1007/s10802-014-9921-6

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the Association for Computing Machinery, 24(6), 381–395. https://doi.org/10.1145/358669.358692

Gadea, M., Aliño, M., Espert, R., & Salvador, A. (2015). Deceit and facial expression in children: The enabling role of the "poker face" child and the dependent personality of the detector. Frontiers in Psychology, 6, 1089–1089. https://doi.org/10.3389/fpsyg.2015.01089

Gangi, D. N., Ibanez, L. V., & Messinger, D. S. (2014). Joint attention initiation with and without positive affect: Risk group differences and associations with ASD symptoms. Journal of Autism and Developmental Disorders, 44(6), 1414–1424. https://doi.org/10.1007/s10803-013-2002-9

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. Journal of Autism and Developmental Disorders, 37(4), 613–627. https://doi.org/10.1007/s10803-006-0280-1

Guha, T., Yang, Z., Grossman, R. B., & Narayanan, S. S. (2018). A computational study of expressive facial dynamics in children with autism. IEEE Transactions on Affective Computing, 9(1), 14–20. https://doi.org/10.1109/taffc.2016.2578316

Guha, T., Yang, Z., Ramakrishna, A., Grossman, R. B., Darren, H., Lee, S., & Narayanan, S. S. (2015). On quantifying facial expression-related atypicality of children with autism spectrum disorder. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2015, 803–807. https://doi.org/10.1109/icassp.2015.7178080

Guthrie, W., Wallis, K., Bennett, A., Brooks, E., Dudley, J., Gerdes, M., … Miller, J. S. (2019). Accuracy of autism screening in a large pediatric network. Pediatrics, 144(4), e20183963. https://doi.org/10.1542/peds.2018-3963

Haines, N., Bell, Z., Crowell, S., Hahn, H., Kamara, D., McDonough-Caplan, H., … Beauchaine, T. P. (2019). Using automated computer vision and machine learning to code facial expressions of affect and arousal: Implications for emotion dysregulation research. Development and Psychopathology, 31(3), 871–886. https://doi.org/10.1017/S0954579419000312

Hashemi, J., Campbell, K., Carpenter, K., Harris, A., Qiu, Q., Tepper, M., … Calderbank, R. (2015). A scalable app for measuring autism risk behaviors in young children: A technical validity and feasibility study. Paper presented at the Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare, Dublin, Ireland.

Hashemi, J., Dawson, G., Carpenter, K. L. H., Campbell, K., Qiu, Q., Espinosa, S., … Sapiro, G. (2018). Computer vision analysis for quantification of autism risk behaviors. IEEE Transactions on Affective Computing, 1–1. https://doi.org/10.1109/taffc.2018.2868196

Jones, E. J., Dawson, G., Kelly, J., Estes, A., & Webb, S. J. (2017). Parent-delivered early intervention in infants at risk for ASD: Effects on electrophysiological and habituation measures of social attention. Autism Research, 10(5), 961–972.

Jones, E. J., Venema, K., Earl, R., Lowy, R., Barnes, K., Estes, A., … Webb, S. (2016). Reduced engagement with social stimuli in 6-month-old infants with later autism spectrum disorder: A longitudinal prospective study of infants at high familial risk. Journal of Neurodevelopmental Disorders, 8(1), 7.

Khowaja, M., Robins, D. L., & Adamson, L. B. (2017). Utilizing two-tiered screening for early detection of autism spectrum disorder. Autism, 22, 881–890. https://doi.org/10.1177/1362361317712649

Lewinski, P. (2015). Automated facial coding software outperforms people in recognizing neutral faces as neutral from standardized datasets. Frontiers in Psychology, 6, 1386. https://doi.org/10.3389/fpsyg.2015.01386

LoBue, V., & Thrasher, C. (2014). The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. Frontiers in Psychology, 5, 1532. https://doi.org/10.3389/fpsyg.2014.01532

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., 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.

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.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., … Lord, C. (2009). The Autism Diagnostic Observation Schedule-toddler module: A new module of a standardized diagnostic measure for autism spectrum disorders. Journal of Autism and Developmental Disorders, 39(9), 1305–1320. https://doi.org/10.1007/s10803-009-0746-z

Maestro, S., Muratori, F., Cavallaro, M. C., Pei, F., Stern, D., Golse, B., & Palacio-Espasa, F. (2002). Attentional skills during the first 6 months of age in autism spectrum disorder. Journal of the American Academy of Child and Adolescent Psychiatry, 41(10), 1239–1245. https://doi.org/10.1097/00004583-200210000-00014

Mcgee, G. G., Feldman, R. S., & Chernin, L. (1991). A comparison of emotional facial display by children with autism and typical preschoolers. Journal of Early Intervention, 15(3), 237–245. https://doi.org/10.1177/105381519101500303

Messinger, D. S., Mahoor, M. H., Chow, S. M., & Cohn, J. F. (2009). Automated measurement of facial expression in infant-mother interaction: A pilot study. Infancy, 14(3), 285–305. https://doi.org/10.1080/15250000902839963

Mullen, E. M. (1995). Mullen scales of early learning. Circle Pines, MN: American Guidance Service Inc.

Murias, M., Major, S., Compton, S., Buttinger, J., Sun, J. M., Kurtzberg, J., & Dawson, G. (2018). Electrophysiological biomarkers predict clinical improvement in an open-label trial assessing efficacy of autologous umbilical cord blood for treatment of autism. Stem Cells Translational Medicine, 7(11), 783–791. https://doi.org/10.1002/sctm.18-0090

Myers, S. M., Johnson, C. P., & Council on Children With Disabilities. (2007). Management of children with autism spectrum disorders. Pediatrics, 120(5), 1162–1182. https://doi.org/10.1542/peds.2007-2362

Nichols, C. M., Ibanez, L. V., Foss-Feig, J. H., & Stone, W. L. (2014). Social smiling and its components in high-risk infant siblings without later ASD symptomatology. Journal of Autism and Developmental Disorders, 44(4), 894–902. https://doi.org/10.1007/s10803-013-1944-2

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.

Owada, K., Kojima, M., Yassin, W., Kuroda, M., Kawakubo, Y., Kuwabara, H., … Yamasue, H. (2018). Computer-analyzed facial expression as a surrogate marker for autism spectrum social core symptoms. PLoS One, 13(1), e0190442. https://doi.org/10.1371/journal.pone.0190442

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. Law and Human Behavior, 29(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). Pediatrics, 133(1), 37–45. https://doi.org/10.1542/peds.2013-1813

Samad, M. D., Diawara, N., Bobzien, J. L., Harrington, J. W., Witherow, M. A., & Iftekharuddin, K. M. (2018). A feasibility study of autism behavioral markers in spontaneous facial, visual, and hand movement response data. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(2), 353–361. https://doi.org/10.1109/tnsre.2017.2768482

Snow, M. E., Hertzig, M. E., & Shapiro, T. (1987). Expression of emotion in young autistic children. Journal of the American Academy of Child and Adolescent Psychiatry, 26 (6), 836–838. https://doi.org/10.1097/00004583-198726060-00006

Sullivan, M. W., & Lewis, M. (2003). Emotional expressions of young infants and children—A practitioner's primer. Infants and Young Children, 16(2), 120–142. https://doi.org/10.1097/00001163-200304000-00005

Tantam, D., Holmes, D., & Cordess, C. (1993). Nonverbal expression in autism of Asperger type. Journal of Autism and Developmental Disorders, 23(1), 111–133.

Tenenbaum, E. J., Carpenter, K. L. H., Sabatos-DeVito, M., Hashemi, J., Vermeer, S., Sapiro, G., & Dawson, G. (2020). A six-minute measure of vocalizations in toddlers with autism spectrum disorder. Autism Research, 13(8), 1373–1382. https://doi.org/10.1002/aur.2293

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Trevisan, D. A., Bowering, M., & Birmingham, E. (2016). Alexithymia, but not autism spectrum disorder, may be related to the production of emotional facial expressions. Molecular Autism, 7, 46. https://doi.org/10.1186/s13229-016-0108-6

Trevisan, D. A., Hoskyn, M., & Birmingham, E. (2018). Facial expression production in autism: A meta-analysis. Autism Research, 11(12), 1586–1601. https://doi.org/10.1002/aur.2037

Waizbard-Bartov, E., Ferrer, E., Young, G. S., Heath, B., Rogers, S., Wu Nordahl, C., … Amaral, D. G. (2020). Trajectories of autism symptom severity change during early childhood. Journal of Autism and Developmental Disorders. https://doi.org/10.1007/s10803-020-04526-z

Werner, E., Dawson, G., Osterling, J., & Dinno, N. (2000). Brief report: Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. Journal of Autism and Developmental Disorders, 30(2), 157–162.

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006). *A 3D facial expression database for facial behavior research*. Paper presented at the 7th International Conference on automatic face and gesture recognition (FGR06), University of Southampton, Southampton, UK.

Yirmiya, N., Kasari, C., Sigman, M., & Mundy, P. (1989). Facial expressions of affect in autistic, mentally retarded and normal children. Journal of Child Psychology and Psychiatry, 30(5), 725–735.