Characterizing autistic features among preschoolers interacting with social robots

Marie D. Manner[1], **Maria Gini**[1], and **Jed Elison**[2]
[1]Department of Computer Science and Engineering
[2]Institute of Child Development
University of Minnesota
manne044, gini, jtelison @umn.edu

Abstract

Emerging evidence suggests that individuals with autism spectrum disorder show interests or preferences for specific types of nonsocial information, including various forms of technology such as autonomous robots. It is unknown whether direct interaction with a humanoid robot could provide novel insight regarding characteristic variability among young children with autism spectrum disorder and/or at-risk for a subsequent diagnosis. We recruited and assessed n = 59 preschool aged children (2 – 4 year-olds), some of whom are considered at risk for developing autism spectrum disorder. We characterized variability in social responsiveness with standardized parent-report measures and an automated characterization of proxemics, or social distance, during a semi-structured interaction between the child and robot. We clustered participants based on proxemics and two dimensions of autistic features, one that quantifies ritualized/routinized behaviors and one that quantifies socialization. We then validated the clinical utility of this grouping strategy with two other dimensions of autism, one that quantifies restricted patterns of play and one that quantifies communication abilities. Results show that this social distance measure, captured in this unique context, can be used with existing methods to cluster participants into clinically meaningful subgroups.

1 Introduction

Autism spectrum disorder (ASD) is a highly variable, developmental disorder characterized by a variety of deficits in social communication and interaction, such as lack of eye contact, and restricted or repetitive behaviors, interests, and activities [Association, 2013]. These behaviors are thus considered early markers for autism. Research has also shown that children with autism have a strong interest in technology such as autonomous robots, as established in [Scassellati, 2007]. Given that the autism phenotype, or observable characteristics, varies from child to child, this work seeks to combine the frequent interests of children with autism in technology and robots, social behaviors of robots, and automated proxemics detection to add more nuanced characterization to the autism phenotype.

This work combines and validates a combination of important metrics of autism: social distances, ritual behaviors, restricted behaviors, socialization capabilities, and communication capabilities. Social distances, or proxemics, has been shown to differ between typically developed people and people with autism; this personal space differs between individuals and also between objects and people [Asada *et al.*, 2016]. Restricted and ritual behaviors can include intense interest in a very narrow range of items (circumscribed interests), repetitive movements like hand flapping or rocking, ritual behaviors like lining up toys, and/or insistence on sameness of environment or routines. Communication and socialization skills can be measured to verify a child is hitting milestones at a similar rate as typically developing peers. We collected parent-reported ratings of repetitive and restricted behaviors using the Repetitive Behavior Scales for Early Childhood [Wolff *et al.*, 2016]. as well as information on adaptive behavior (Vineland Adaptive Behavior Scales) and general developmental level (Mullen Scales of Early Learning). Relating these metrics to our new metric, a normalized ratio using the child's distance to the social robot and the child's distance to their closest caregiver, shows us distinct groups of participants. This research shows that children clustered on some sub-scales of these tests, combined with our distance metric, show differences in other development sub-scales.

We begin with related work in Section 2, give a high-level description of the human-robot interaction experiment and raw data taken therefrom in Section 3, describe the data used and preliminary analysis in Section 4, and end with a summary in Section 5.

2 Related Work

Socially Assistive Robotics is a recent area of robotics research aimed at helping populations with special needs; it includes research for children or adults, such as robots as tools for children with pervasive developmental disorders or robots for adults as tools, companions, or helpers. Socially assistive robotics research in autism is over a decade old, yet does not currently meet standards of psychology and child development researchers [Diehl *et al.*, 2012; Scassellati *et al.*, 2012; Pennisi *et al.*, 2016]. Robotics research with children with autism stems from the fact that afflicted children tend to especially enjoy autonomous (or seemingly autonomous) robots [Dautenhahn and Werry, 2004], and researchers have used a wide variety of robot appearances and abilities in this area [Scassellati *et al.*, 2012]. While the reason for this high level of interest is unknown, researchers clearly have the potential to leverage robotics for autism diagnosis or treatment [Scassellati, 2005].

Automatically detecting autism or autistic traits is a current research area in computer vision, and much work uses as much data as possible. For example, Hashemi et al. [2012] analyzed footage from a non-intrusive GoPro camera placed on a table, two to four feet from a clinician-child pair in which the clinician was testing the child with a disengagement of attention task and a visual tracking task. The authors went even further in [Hashemi *et al.*, 2014], in which they analyzed interest sharing and atypical motor behavior by estimating head motions from facial features and motor behavior by arm symmetry. Fasching et al. [2015] automatically coded activities of people with obsessive-compulsive disorders from overhead video footage in a structured lab, tracking how many times participants touched various objects. These objects are statically located, such as faucets and handles, and easier to locate in a static environment. In contrast, our laboratory works with very young children in a play-based interaction, which adds difficulties in instrumenting the room and reliably tracking an active, potentially non-cooperative child.

Our work is informed by a common need of children on the autism spectrum for sameness of routine; therefore, we do not want to instrument the child's clothing, or interrupt their autonomy and routine more than already necessary to play with a novel, social humanoid robot. We want to avoid adding fiducial markers and maintain a structured and reproducible play interaction. All interactions are thus initiated by a novel, humanoid robot, while the entire experiment is recorded from several different perspectives. As our own studies evidence, the only perspective that can reliably capture the participant's location at all times is from the ceiling, so we mount a GoPro camera on the ceiling in the center of the room. Fig. 1 shows this vantage point, which allows for tracking people's position over time.

Much research in socially assistive robotics studies child-robot interactions. Feil-Seifer and Matarić [2011] created a short free-play interaction with children and a robot, with the future intention of allowing a robot to adjust its own behavior based on the child's reaction. This work tracked the child in relation to the robot to automatically determine if the child was having a positive or negative reaction to the robot. The authors manually coded for the child avoiding the robot, interacting with the robot or playing with bubbles the robot generated, staying still, being near parent, being against the wall, or none of those. Results showed that children with a positive reaction to the robot spent over 80% of time interacting, whereas children with a negative reaction spent less than 20% of time interaction with the robot. Mead et al. [2013] also investigated proxemics (the study of social distances), by placing a participant and researcher in discussion about a static humanoid robot. Using a video camera and depth data, they studied body pose during the experiment, training Hidden Markov Models on sensory experiences (such as voice loudness and a variety of distances to other people and environment objects) to correctly annotate initiation and termination of conversation.

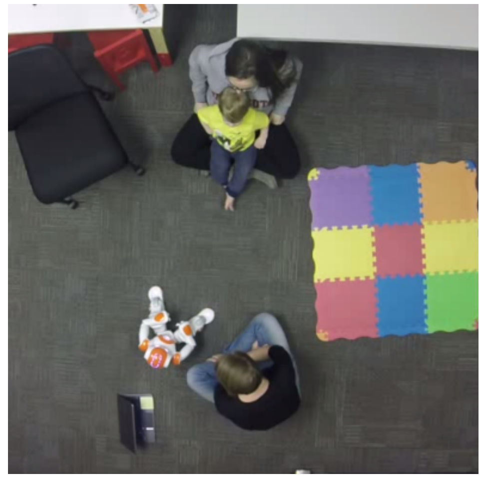

Figure 1: Sample overhead view of the human-robot interaction experiment.

Current research has therefore touched on proxemics and how these measures change between typically and atypically developing people, automatically detecting symptoms of autism, and how children react to autonomous robots. We contribute a much larger sample size of participants, a far more characterized description of our participants, and a novel metric of our participants based on how they interact with a social humanoid robot. We leverage these details to compare and contrast our participants, which ultimately allows us to characterize and group our participants into different proxemic behaviors.

3 Research Method

3.1 Experimental Paradigm

The overarching goal of the robot interaction study with toddlers (the age group of roughly 2–3 years old) is to better characterize quantitative autistic traits in order to improve characterization of homogeneous subgroups of children with autism or at risk for autism. If we can better determine subsets of autism phenotypes, we may be able to identify children at high risk for autism spectrum disorder (ASD) when they are very young. We first needed to demonstrate that the response when interacting with a social robot varies between children at all. For example, if the only responses to playing with the robot were to sit on one's parent or to sit in front of the robot, then the play scenario does not generate varied enough response to merit recruiting children diagnosed with autism and wasting an affected family's valuable and constrained time. Therefore, we first recruited a large number of fairly neurotypical children for a proof-of-concept human-robot interaction experiment.

Children were recruited from a laboratory-maintained database at the University of Minnesota's Institute of Child Development. The participants in this study are low to medium risk for ASD (because autism is more common among siblings, some participants are considered at higher risk); none of the participants are clinically diagnosed with a spectrum disorder. Thus, the data shown in this paper constitutes a reliable interaction baseline of varied but neurotypical children, and the variability in the participants shown in our results can be compared to future interactions between our robot and very high risk or clinically diagnosed toddlers. Written informed parental consent was ensured in advance of all testing; all research was approved by the university's Institutional Review Board.

We collected multiple data sets from each participant, including standardized and novel assessments and video footage. We introduce and use a proxemics metric automatically generated using the overhead video footage, as well as scores from classic child development assessments, the Vineland Adaptive Behavior Scales [Sparrow *et al.*, 1984], Mullen Scales of Early Learning [Mullen, 1995], Social Responsiveness Scale [Constantino and Gruber, 2002], and the Video-Referenced Ratings of Reciprocal Social Behavior [Marrus *et al.*, 2015]. We first wrote software to track actors of interest throughout the video; these actors include the child, the robot, experimenters, and caregivers. The actor tracking software we wrote uses the AdaBoost [Freund and Schapire, 1999] algorithm implemented in the open source computer vision library OpenCV3.1 and Python3.4. The user first chooses the objects or people to track, then the software tracks the item over time from start to end of the robot interaction video, frame by frame. We record the center of the raw coordinates of each actor, calculate the distance between them (accounting for size of the actor), and convert the resulting pixel distance to distance in feet.

During the experiments, we introduce the child to Robbie the Robot (a NAO from Softbank Robotics). Robbie plays different games such as looking games, imitation games, and dances. The games include "I Spy" (a looking game that encourages the child to find objects in the room), "Simon Says" (an imitation game that encourages the child to copy motions like clapping and waving), and several dances set to music. The set of games is presented in the same order for every child. The experimenter controlling the robot imitates some of the robot's movements and plays along during some of the looking games, encouraging the child to do the same. The interaction is recorded from up to four perspectives, most notably from a GoPro mounted on the ceiling. The GoPro is the only camera that is always located in the same place, is impossible to reach by participants, and from which we can almost always see all actors in the room.

We attempted to keep the same location for all actors in the room across experiments. The experimenter that controlled the robot (hereafter called simply the experimenter) sat next to the robot slightly off-center in the room. If the child did not need comfort or attention from their caregiver, they were seated or standing on the floor near the robot, facing the robot. In this case, the caregiver sat near the edge of the room with another researcher, answering questions from a development assessment. If the child needed constant or frequent attention from the parent, the child might be seated on or near their parent during all or part of the interaction, usually closer to the robot than the parent would be if the child did not need attention. Fig. 1 shows part of an overhead video frame; the child faces the robot, seated on their caregiver, and the experimenter is next to the robot.

3.2 Data

In all, 65 participants were recruited for this study, of which 60 contributed video footage of a robot interaction. In one case, the experimenter, a second researcher, and the parent all sat on the floor and attempted to draw the child's attention to the robot; as this was a highly unusual configuration, this video was not analyzed for this paper, leaving us with 59 videos. These 59 participants (31 males, 28 females), were aged 25–45 months (mean 32.9 months, standard deviation 4.6 months). The interactions and therefore videos range from roughly nine to 15 minutes long, depending on the child's willingness or ability to continue interacting with the robot. The original video is slightly distorted, thus we first undistort it and use the undistorted video for later analysis.

There are seven total parts to the interaction (games and dances), which we call presses for attention, or *presses*. Each press varies in time, and there is a one minute buffer between each press to give the child time to re-engage if they were not interested or took a break from playing for some reason (e.g. requested a snack or needed the rest-room). If the experimenter did not trigger the next press, after one minute passed the robot started the next press anyway. Some interactions were not completed due to equipment malfunction or participant choice, giving us some children who did not complete all seven presses. Where possible, these data are included in the analyses.

The proxemics data used is generated from the raw coordinates of each actor in the room. First, we find the Euclidean distance between actors, giving us three channels of data: the distances between the child and robot, the child and caregiver, and the child and experimenter. In some experiments, two caregivers were present during the interaction; the minimum distance between the child and either parent was used in our data, ensuring that we can reasonably compare children with one or two caregivers present. The interactions, which are recorded at 33 frames/second, were reduced by averaging the Euclidean distances in one second windows to smooth the data slightly; we use this averaged dataset for data analysis.

The assessment data used is generated by three exams – two parent-report questionnaires and an administered test. The first test is the Vineland Adaptive Behavior Scales (2nd edition); this assessment asks parents to report on the child's behavior in five areas. These areas are communication, daily living skills, socialization, motor skills, and maladaptive behavior. Our research focuses on the socialization, communication, and maladaptive behavior scales. The second test is the Mullen Scales of Early Learning, which is a test administered by a trained professional to assess cognitive functioning in children aged 0 to 68 months (5.6 years). The areas tested in this exam include gross motor skills, visual reception, fine motor, expressive language, and receptive language. The third test varies by child. For children around 31 months or older (2.5 years), we use the Social Responsiveness Scale, which is a parent report questionnaire measuring social abilities. For younger children, which in this study are approximately 24–30 months of age (2 years to 2.5 years), we use the Video-Referenced Ratings of Reciprocal Social Behavior, a parent report questionnaire measuring social abilities.

4 Data Analysis and Results

Given the locations and distances over time between actors in the interaction, we have three distinct time series: the distances between child and NAO, child and caregiver, and child and experimenter. An example is shown in Fig. 2

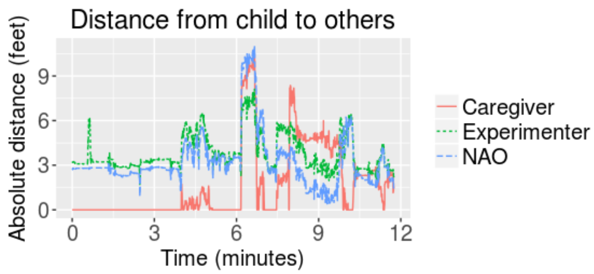

Figure 2: The Euclidean distances between the child and parent, child and robot, and child and experimenter (smoothed by averaging over every second). Blue vertical bars indicate the beginning of a press for social interaction, or a looking / imitating / dancing game.

We are first motivated to normalize the absolute distances between actors during an interaction; a child that is two feet from the robot and ten feet from their parent clearly has a different comfort level than a child that sits on their parent two feet away from the robot. There, we need a normalization method that reduces the absolute distances to something that evokes the different comfort levels between child and robot or child and parent. To that end, we created a new metric, "Distance Ratio," that considers the distance to the robot, N, and the distance to the caregiver, CG, in Eq. 1.

$$N/(N + CG) \tag{1}$$

We use this to reduce the variable distance to NAO and distance to caregiver time series to a single, unit-less number scaled $[0, 1]$ for any particular moment in time during an interaction, where a 1 value means closest to caregiver and a 0 value means closest to the robot. This allows us to more easily see clear trends in the data, regardless of the absolute distance. Fig. 3 shows all participants; some participants show clear trends over time in their average Distance Ratio per press. The data shows four clear subsets, along with a small group with no obvious movement pattern. These subsets include children that stay near their caregiver the entire time (the lines near 1 at the top of the graph, 12 / 59 children), those that stay near the NAO robot more of the time (the lines near 0 at the bottom of the graph, 22 / 59), those that began the interaction close to their parent but moved closer to the robot over the duration of the interaction (6 / 59), those that move from closer to the robot to closer to their parent (3 / 59), as well as children with no reliable pattern of Distance Ratio (16 / 59).

Note that each interaction varies in length due to potential buffer time between presses in a single interaction. The buffer

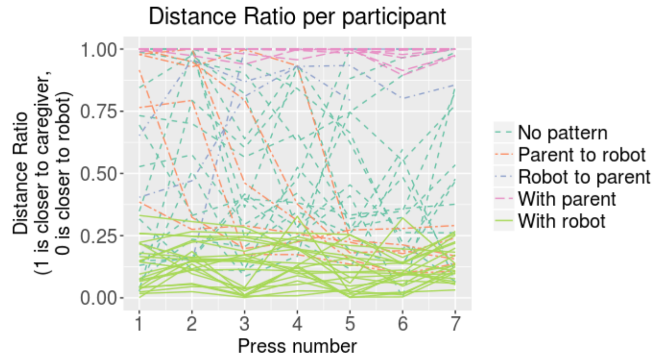

Figure 3: Some visible patterns: children that stay near the robot (22), stay near the caregiver (12), move from caregiver to robot over time (6), and move from robot to caregiver over time (3). Some have no discernible pattern (16). (Best viewed in color.)

time, lasting up to one minute between presses as needed, is included in the last press that occurred. For example, if Child A needed a 40 second break after a one minute press, but Child B didn't need a break and only used two seconds after the same one minute press, the press lasted for 100 seconds for Child A but only 62 seconds for Child B.

This flexibility in interaction time naturally raises the question of how to compare these variable length data. The time difference between presses is capped at 60 seconds, and any time between presses is used to draw the participant's attention back to the robot. By and large, the excess time was spent by the participant getting a snack or toy, playing with other items in the room, or talking to their caregiver; none of the buffer time was spent interacting with the robot while the robot was not moving autonomously. Therefore, we consider the time between presses to be noise, and we analyze only the times the robot and child spent interacting together.

To remove the noise, we choose a simple method of aligning all presses in all interactions, and we truncate each press to the length of the shortest occurrence of that press over all participants. For example, say Child A took 60 seconds during Press 1 with a 30 second break, then 180 seconds during Press 2 with a 10 second break. Say Child B took 60 seconds during Press 1 with a 2 second break, then 180 seconds during Press 2 with a 2 second break. It should be noted that in most cases the experimenter manually starts the next press for attention with the robot, so short breaks of 1-3 seconds are simply the time taken to reach over and push buttons on the robot (or occasionally, to first re-orient the robot towards the child if they shifted position). Our first data exploration only compares participant reactions to the robot while the robot is actively moving or speaking; thus, Press 1 is truncated to 62 seconds for both participants and Press 2 is truncated to 182 seconds for both participants.

The effect of such data loss, i.e. 28 seconds after Press 1 and 8 seconds after Press 2 in the above example, admittedly contains some distance data. Either the participant didn't move between presses or approached the robot again from somewhere else in the room. In theory, a participant might have outlasted the one minute buffer time and started the next

Table 1: Data spread of used scores from Vineland, Mullen, and vrRSB or SRS assessments.

	Min	Max	Mean	Std Dev
Restricted behaviors	0	8	2.36	2.06
Ritual behaviors	0	9	2.79	2.01
Composite R/R Items	0	26	9.64	6.79
Socialization	82	136	108.46	11.90
Composite adaptive behaviors	86	133	107.81	12.03
Communication	81	135	109.79	11.21
Early learning composite (ELC)	83	148	115.09	17.15
vrRSB or SRS	0.03	0.58	0.13	0.09

press further from the robot than they were at the (truncated) end of the previous press. However, in practice, this did not happen; only the first two cases occurred. Thus, minimal interaction information was lost due to this truncation.

In addition to the social distances and metrics above, we consider several scales from the Vineland and Mullen assessments described above. Specifically, we consider the number of restricted behaviors endorsed, number of ritual items endorsed, composite restricted or ritualized behavior items endorsed, socialization score, adaptive behavior composite scores, communication scores, early learning composite score, and the Social Responsiveness Scale or Video-Referenced Ratings of Reciprocal Social Behavior (whichever was appropriate based on the participants' age). The minimum, maximum, mean, and standard deviation of these scores across all participants are given in Table 1.

We performed several analyses on our data. First, to see if the Distance Ratio was strongly correlated with any of the other score values of interest, we generated a correlation matrix. No strong redundant relationships were found, as the highest value between the Distance Ratio and any other score was only 0.15. Next, we check if our participants varied enough to be grouped by the Distance Ratio. We hypothesized that the Distance Ratio (using mean Distance Ratio for an interaction) with a physical dimension and a social dimension could result in distinct clusters of participants. A reliable, good clustering would be stable (participants would be clustered in the same way for a large number of iterations) and have internal consistency (statistically significant differences in the scores used to cluster). If the clusters also had external consistency, or statistically significant differences in scores that were not used to cluster, then we will have validated our novel distance metric varies between children and that can be used in conjunction with existing measures to show differences in behaviors in a child-robot interaction experiment.

We performed KMeans clustering across the participants, from $K = 2$ to $K = 5$, using Distance Ratio, ritual behavior sub-score, and socialization sub-score. The most stable clustering was $K = 2$; participants were grouped the same way every time out of $1,000$ iterations. We then performed a t-test over the participant groups' scores used for clustering. All three scores between the two groups were statistically significant, at $p < 0.002$, meaning these clusters are very stable and contain participants with distinctly different ritual and socialization behaviors. Children who spent time closer to NAO were also the children with higher ritualized behavior scores, but also higher socialization scores.

We then sought to validate the clusters with different measures, not used for clustering, along another physical dimension (restricted behavior sub-score), another social dimension, (communication behavior sub-score), reciprocal behaviors score (vrRSB or SRS, whichever was appropriate based on age), and the early composite learning score. Performing a two-tailed t-test on each set of scores per group resulted in statistically significant differences in both the restricted behaviors sub-score and the communication sub-score ($p < 0.05$ for each), but not the others. A graph of group assignment color coded in blue and red, with the three values as axes, is shown in Fig. 4. To see if these clusters differ in statistically significant ways, we performed a two-tailed t-test on all test scores listed in Table 1. The two groups' test score averages and standard deviations are reported in Table 2. The second group, with a much smaller Distance Ratio (spends more time closer to NAO), has higher restricted behaviors but also higher communication skills (an unexpected finding, but statistically significant nonetheless).

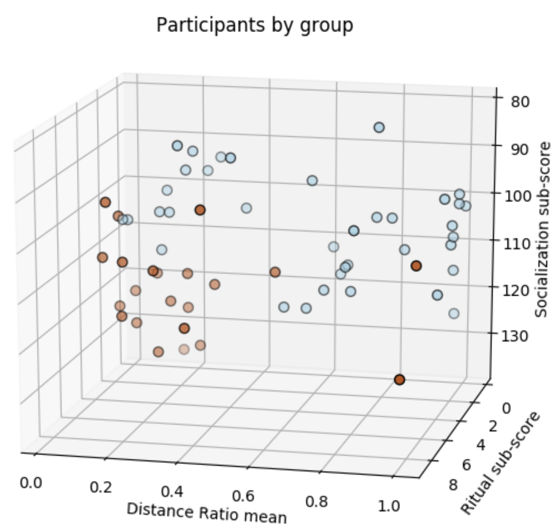

Figure 4: Distance Ratio versus ritual behavior sub-score versus socialization sub-score, by group.

5 Conclusions and Future Work

In this paper we have shown our new metric, Distance Ratio, that scores a child's social distance when interacting with a novel robot, elicits meaningful variability in proxemics. We showed that children who spent more time in close proximity to the robot also show more ritualized and restricted behaviors. The distance ratio is generated from applying computer vision techniques to overhead video footage of participants interacting with a robot, and normalizing the absolute distance between the child and robot and child and caregiver.

Lastly, while our initial method of equalizing participants' time series is truncation of excess noise, it could be that the

Table 2: Group differences by score, where M is mean and SD is standard deviation. * indicates statistical significance of $p < 0.05$, and ** indicates statistical significance of $p < 0.002$.

	G_0 M	G_0 SD	G_1 M	G_1 SD
Distance Ratio**	0.61	0.33	0.27	0.26
Ritual**	1.95	1.22	4.29	2.18
Socialization**	104.19	10.03	115.8	11.00
Restricted*	1.95	1.90	3.10	2.07
Comm.*	107.5	11.12	113.6	10.01
vrRSB or SRS	0.13	0.10	0.14	0.06
ELC	113.1	17.69	119.4	14.49

time lapse between presses holds key information– for example, longer time between presses potentially indicates a less interested child or a more active child that repeatedly needs their attention drawn back to the robot. Thus, other methods that include the entire interaction, such as dynamic time warping, might show useful differences.

Acknowledgments

The authors thank the National Science Foundation "I/UCRC: Robots and Sensors for the Human Well-being," [NSF IIP-1439728] and the MnDRIVE RSAM initiative at the University of Minnesota for partial support for this work.

References

[Asada *et al.*, 2016] Kosuke Asada, Yoshikuni Tojo, Hiroo Osanai, Atsuko Saito, Toshikazu Hasegawa, and Shinichiro Kumagaya. Reduced personal space in individuals with autism spectrum disorder. *PLoS One*, 11(1), 2016.

[Association, 2013] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental disorders*. American Psychiatric Publishing, Arglinton, VA, 5 edition, 2013.

[Constantino and Gruber, 2002] John N. Constantino and C Gruber. *The Social Responsiveness Scale*. Western Psychological Services, Los Angeles, 1 edition, 2002.

[Dautenhahn and Werry, 2004] Kerstin Dautenhahn and Iain Werry. Towards interactive robots in autism therapy: background, motivation and challenges. *Pragmat. Cogn.*, 12(1):1–35, 2004.

[Diehl *et al.*, 2012] Joshua J Diehl, Lauren M Schmitt, Michael Villano, and Charles R Crowell. The clinical use of robots for individuals with autism spectrum disorder: a critical review. *Res. Autism Spectr. Disord.*, 6(1):249–262, 2012.

[Fasching *et al.*, 2015] Joshua Fasching, Nicholas Walczak, Vassilios Morellas, and Nikolaos Papanikolopoulos. Classification of motor stereotypies in video. *IEEE Int. Conf. Intell. Robot. Syst.*, pages 4894–4900, 2015.

[Feil-Seifer and Matarić, 2011] David J Feil-Seifer and Maja J Matarić. Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. *Proc. 6th Int. Conf. Human-robot Interact. (HRI)*, pages 323–330, 2011.

[Freund and Schapire, 1999] Yoav Freund and Robert E. Schapire. A short introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5):771–780, 1999.

[Hashemi *et al.*, 2012] Jordan Hashemi, Thiago Vallin Spina, Mariano Tepper, Amy Esler, Vassilios Morellas, Nikolaos Papanikolopoulos, and Guillermo Sapiro. Computer vision tools for the non-invasive assessment of autism-related behavioral markers. *2012 IEEE Int. Conf. Dev. Learn. Epigenetic Robot. ICDL 2012*, pages 1–33, 2012.

[Hashemi *et al.*, 2014] Jordan Hashemi, Mariano Tepper, Thiago Vallin Spina, Amy Esler, Vassilios Morellas, Nikolaos Papanikolopoulos, Helen Egger, Geraldine Dawson, and Guillermo Sapiro. Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism Res. Treat.*, 2014.

[Marrus *et al.*, 2015] Natasha Marrus, Anne L Glowinski, Theodore Jacob, Ami Klin, Warren Jones, Caroline E Drain, Kieran E Holzhauer, Vaishnavi Hariprasad, Rob T Fitzgerald, Erika L Mortenson, Sayli M Sant, Lyndsey Cole, Satchel A Siegel, Yi Zhang, Arpana Agrawal, Andrew Heath, and John N Constantino. Rapid video-referenced ratings of reciprocal social behavior in toddlers: A twin study. 56(12):1338–1346, 2015.

[Mead *et al.*, 2013] Ross Mead, Amin Atrash, and Maja J Matarić. Automated proxemic feature extraction and behavior recognition: applications in human-robot interaction. *Int. J. Soc. Robot.*, 5(3):367–378, 2013.

[Mullen, 1995] Eileen M Mullen. *Mullen scales of early learning*. AGS Circle Pines, MN, 1995.

[Pennisi *et al.*, 2016] Paola Pennisi, Alessandro Tonacci, Gennaro Tartarisco, Lucia Billeci, Liliana Ruta, Sebastiano Gangemi, and Giovanni Pioggia. Autism and social robotics: A systematic review. *Autism Res.*, 9(2):165–183, 2016.

[Scassellati *et al.*, 2012] Brian Scassellati, Maja J Matarić, and Henny Admoni. Robots for use in autism research. *Annu. Rev. Biomed. Eng.*, 14(1):275–294, 2012.

[Scassellati, 2005] Brian Scassellati. Quantitative metrics of social response for autism diagnosis. *Proc. IEEE Int. Work. Robot Hum. Interact. Commun.*, 2005:585–590, 2005.

[Scassellati, 2007] Brian Scassellati. How social robots will help us to diagnose, treat, and understand autism. *Robot. Res.*, pages 552–563, 2007.

[Sparrow *et al.*, 1984] Sara S Sparrow, David A Balla, Domenic V Cicchetti, Patti L Harrison, and Edgar A Doll. Vineland adaptive behavior scales. 1984.

[Wolff *et al.*, 2016] Jason J Wolff, Brian A Boyd, and Jed T Elison. A quantitative measure of restricted and repetitive behaviors for early childhood. *Journal of neurodevelopmental disorders*, 8(1):27, 2016.