Imperial College London

Department of Mechanical Engineering

Socially Assistive Robotics for the Development of Communication and Interpersonal Skills in Children with Autism Spectrum Disorder

ME3 Literature Review Project

Author: Xue Xian Lim (CID 01766000)
Supervisor: Dr Ravi Vaidyanathan

Abstract

There has been a recent surge in interest in the employment of socially assistive robotics (SAR) for the development of social skills in children with autism spectrum disorder (ASD). This literature review critically examines studies on social SAR interventions for children with ASD published within the last five years. Overall, research in this field remains nascent while studies still face methodological limitations, such as small sample sizes, short intervention durations, and inappropriate outcome measurement indicators. The studies reported largely mixed outcomes from the SAR interventions. The relatively fixed and unchanging overt behaviours – namely eye contact, gestural recognition and production, facial expression recognition and production, and social routines – tended to show more promising results and were hence, recommended as an achievable goal within the next decade once research was carried out to verify the generalisability and sustainability of the improvements from the SAR interventions. The more complex social skills – imitation, joint attention, verbal communication – either showed highly inconsistent results or were simply not covered in depth. This review advocates for closer collaboration with the clinical community given the disproportionate focus on the technicalities of the SARs and insufficient attention on the therapy itself and the needs of the user – children with ASD.

Statement of Project Objectives

The objectives of this literature review are to:

1 Assess the efficacy of utilising socially assistive robotics in helping children with autism spectrum disorder develop social skills
2 Identify the challenges and limitations of current research
3 Outline potential future research directions

Table of Contents

1 Introduction

Autism spectrum disorder (ASD) is a complex, persistent, and heterogeneous neurodevelopmental condition. Individuals diagnosed with ASD face difficulties in social communication and interaction while also exhibiting restricted interests and repetitive behaviours (American Psychiatric Association, 2013). Recent studies have found that there is a growing prevalence of ASD globally, with the latest surveys suggesting that the prevalence rates may be as high as 1 in 44 children in the US (Centers for Disease Control and Prevention, 2021) and 1 in 57 children in the UK (Roman-Urrestarazu et al., 2021). However, this may be in part attributable to updates to the reporting practices and diagnostic criteria for ASD (Hansen, Schendel & Parner, 2015).

There is currently no definitive causal factor for ASD although growing evidence points towards a combination of genetic variations and multiple environmental factors which may act either independently or via synergistic mechanisms (Posar & Visconti, 2017). The variable contributions of a multitude of causal factors give rise to a highly heterogeneous autism spectrum which manifests itself as a unique set of observable traits in each affected individual. ASD is generally not curable (Bölte, 2014) although early diagnosis and intervention can lead to sustained, improved outcomes in intellectual ability, language, and social communication for children with ASD (Estes et al., 2015; Fuller & Kaiser, 2020). Any advances in the treatment of ASD in young children are therefore considerably valuable in helping them lead better lives as part of a highly gregarious society.

1.1 Socially Assistive Robotics

Recently, there has been a surge in interest in the employment of socially assistive robotics (SAR) to enhance outcomes in ASD interventions. However, research into its efficacy is still relatively exploratory and nascent. Consequently, the clinical use of SARs as an evidence-based practice to treat ASD remains virtually non-existent (Kim, Elizabeth et al., 2012).

SAR has been defined by Feil-Seifer and Mataric (2005) as an amalgamation of assistive robotics and socially interactive robotics, whereby “close and effective (social) interaction with a human user” is utilised as a means to achieve the end goal of “measurable progress in convalescence, rehabilitation, learning” and other assistive domains. In a seminal study by Emanuel and Weir (1976), a seven-year-old child with ASD was allowed to play with a programmable turtle device and was reported to have displayed communicative behaviours that were previously seldom observed. Since then, close to 30 different robots of varying forms and complexities have been explored for use with children with ASD (Kostrubiec & Kruck, 2020).

There are several potential advantages to utilising SARs in the treatment of children with ASD and these have been succinctly summarised in a review by Cabibihan et al. (2013) and corroborated by recent research. Firstly, SARs are not as complex as humans. SARs can be programmed to respond in a deterministic and predictable manner which is congruent with the desire for repetition and predictability in children with ASD. A recent study by Schadenburg et al. (2021) found a clear link between robot predictability and child engagement. Secondly, SARs are not as intimidating as humans. Children with ASD implicitly perceive SARs as toys and machines (Peca et al., 2014) and have positive beliefs about SARs being nice and doing good things (Costa et al., 2018). Thirdly, SARs enable tangible and embodied interactions. Children can interact with SARs via both verbal communications and physical interactions in the form of gestures and touch. SARs can also incorporate displays, lights, music, sounds, movements, and facial expressions to provide for a more engaging experience. The first two characteristics confer SARs with an advantage over conventional human therapists while the third characteristic provides SARs with an edge over computerised interactions and fully virtual agents.

SARs can be designed to operate either autonomously or via teleoperation by a human supervisor. The latter approach is often termed as Wizard of Oz (WoZ) control.

1.2 Social Skills

Social skills is a broad umbrella term comprising a range of integrative, complementary components (Jurevičienė, Kaffemaniene & Ruškus, 2018) that enable individuals to effectively function in social contexts. In order to provide a framework for future developments of robotic interventions for children with ASD, Huijnen et al. (2016) invited medical practitioners and educators to attend focus group sessions. They defined 74 therapeutical and educational objectives in nine main domains. The objectives from the communication domain and the social / interpersonal interactions and relations domain are shown verbatim in Table 1. These objectives will form the basis upon which the findings from SAR trials are analysed. Listed below are further details about some of the objectives.

1.2.1 Gestures and Non-Verbal Communication

Gestures are a subset of non-verbal communication, involving the movements of the hands (manual gestures) and body (non-manual gestures) to communicate meaning. Gestural research has primarily centred on manual, co-speech gestures (Abner, Cooperrider & Goldin-Meadow, 2015) and this focus is reflected in the reviewed literature as well. These gestures can be deictic (pointing to an object or location), iconic (depicting properties of objects, actions, and scenes; e.g. flapping arms to depict a bird or flying), emblematic / symbolic (conveying a specific meaning within a culture; e.g.

thumbs-up to signify good), or beat gestures (carrying no semantic meaning but accompanying the rhythm of speech).

Other forms of non-verbal communication include eye contact, facial expressions, vocalics / paralanguage (e.g. intonation, volume, rhythm), proxemics (interpersonal distance and spatial behaviour), and haptics (interaction via touch).

Table 1 Therapeutical and Educational Objectives for SARs (Huijnen, C. A. G. J. et al., 2016)

Communication		Social / Interpersonal Interactions and Relations	
1	Orientation to listen	10	Imitation
2	Making contact	11	Attention
3	Learn a new form of communication	12	Appropriately cope with own anger / sadness / …
4	Understand intention of gesture	13	Awareness of feelings, wishes, behaviour, thoughts of others
5	Understand intention of image / symbol	14	Appropriately react to behaviour of others
6	Understand intention of word	15	Social routines (greet, say goodbye, introduce)
7	Use gesture	16	Turn taking (behaviour)
8	Use non-verbal abilities	17	Respect / value others (or things)
9	Talk – use verbal abilities	18	Appropriate behaviour with respect to physical proximity / contact or personal space
		19	Collaboration / joint attention
		20	Ask for help
		21	Conflict management

1.2.2 Verbal Communication

Verbal communication involves the usage of words to convey meaning via receptive (comprehension) and productive (expression) abilities. Strictly speaking, it encompasses written and sign language as well although colloquial usage often equates it synonymously with oral, verbal communication (i.e. speech) (Chandler & Munday, 2011).

1.2.3 Imitation

Imitation plays a crucial role in learning while also serving a social function. It enables the acquisition of new skills and knowledge in children and promotes affiliative bonding in imitative social exchanges (Vivanti & Hamilton, 2014). A review by Ingersoll (2008) highlighted that imitation was closely associated with the development of other social communication aspects, such as language, play skills and joint attention.

1.2.4 Joint Attention

Joint attention (JA) refers to the shared focus by two or more individuals on a common point of reference, be it an object, event, person, or action. JA behaviours can be classified under responding to JA (RJA) or initiating JA (IJA). JA is crucial for early learning and development in children as it facilitates social learning (Mundy & Newell, 2007). Furthermore, JA is closely associated with the development of cooperative skills (Wu et al., 2013) and successful collaborative problem solving (Poysa-Tarhonen et al., 2021).

1.3 Review Outline

This review aims to assess the efficacy of utilising SARs to help children with ASD develop social skills, with particular attention being paid to the individual components of social skills, given that it is such a wide-ranging concept. The next section outlines the review protocol and the inclusion criteria. Following that, the findings of extant literature are presented. Subsequently, the validity and implications of these findings are discussed and recommendations for future research directions in the short term and long term are suggested.

2 Methodology

A comprehensive search was performed on the following academic databases: Engineering Village, IEEE Xplore, PubMed, and Web of Science. Truncation and Boolean operators were applied to the following keywords for the search: (social robot* OR socially assistive robot* OR human robot interaction) AND child* AND (autis* OR ASD) AND (social* OR interpersonal OR communicat*).

The inclusion criteria are as follows:

1. A physically embodied SAR was used. The SAR must interact with the child via speech, facial expressions, gestures, or any combination thereof.
2. The intervention group comprised children who were 12 years old or younger and had been clinically diagnosed with ASD.
3. Trials were carried out on children with published results consisting of quantitative metrics or qualitative observations that were relevant to the development of social skills. These trials should be done with the objective of improving any aspect of social skills and the methodology should reflect so. Studies that involved human-robot interaction (HRI) without any form of social skills intervention / therapy / treatment (e.g. a study exploring the perceptions of children with ASD towards SARs) were excluded.
4. The papers were published within the last 5 years (i.e. from 2016 onwards).

3 Findings

Table 2 presents a summary of the reviewed studies where SARs were employed in interventions for children with ASD.

Table 2 Summary of Reviewed Studies of SAR Interventions for Children With ASD

Title, Author	Robot	Participants, Duration	Methodology	Results	Assessed Objectives
A low-cost socially assistive robot and robot-assisted intervention for children with autism spectrum disorder: field trials and lessons learned **(Boccanfuso et al., 2017)**	CHARLIE	n = 11 (3 – 6 y/o). 12 sessions over 6 weeks.	Intervention grp w/ robot & control grp w/ speech therapy. Game-based therapy w/a focus on creative, child-directed play. Mean length of spontaneous utterances measured. VABS-II, MIS, UIA (Ingersoll & Lalonde, 2010) & EVT-2 scores evaluated.	Improvements in spontaneous utterances, social interaction, JA & requesting behaviours. No statistically significant increases in communication scores, vocabulary, or motor imitation. Significant differences in VABS-II scores between intervention & control grp for 2 subscales – Receptive Language ($p < 0.0421$) as well as Play & Leisure ($p < 0.0469$).	9, 10, 15, 19, 20
A Randomized Controlled Trial of an Intelligent Robotic Response to Joint Attention Intervention System **(Zheng et al., 2020)**	NAO	n = 23 (1.64 – 3.14 y/o, avg. age = 2.54 y/o). 3 children showed significant distress in initial session. Final n = 20. Intervention over 3 – 9 weeks.	Randomised waitlist control design w/ a waitlist control grp & a grp receiving immediate robotic intervention. Robot attempted to direct child's attention to 2 computer monitors. Child rewarded if they accurately responded. Otherwise, more instructive levels of prompts used. Avg. prompt level & target hit rate measured. STAT score also evaluated.	Non-statistically significant improvement in STAT scores overall. Prompt level & hit rate had no significant diff. before and after intervention. Plausible link between those whose STAT scores improved and their lower ADOS-2 scores ($p = 0.03$) and higher IQ scores ($p = 0.06$). Plausible link between younger participants ($p = 0.07$) and decreased prompt levels & increased hit rates.	19
An Adaptive Multi-Robot Therapy for Improving Joint Attention and Imitation of ASD Children **(Ali et al., 2019)**	2 × NAO	n = 12 (11 M, 1 F; 3.7 – 10.4 y/o). 16 sessions over 6 months.	2 experiments: HRI w/ & w/o inter-robot interaction. Both experiments had a JA module & an imitation module. Least-to-most prompting in JA module. Child's eye contact & delay in gaze shifting measured. Accuracy of imitation measured in imitation module. EEG recordings taken to estimate cognitive brain state. CARS score evaluated as well.	Increase in eye contact duration and reduction in delay in making eye contact after stimulus. Improvement in CARS scores.	10, 19
Analysis of the use of a robot to improve social skills in children with autism spectrum disorder **(Valadão et al., 2016)**	MARIA	n = 10 (5 ASD, 5 TD; 4 M 1 F; 7 – 8 y/o). Single 40-min session.	Phase 1: Robot self-presentation, robot was revealed & moved in a pre-defined set of movements. Freq. of child looking away counted. Phase 2: Interaction invitation, mediators invite child to play an imitation game w/ robot. Freq. of touching robot & imitating mediators counted. GAS used to convert data to an overall score. Likert questionnaire conducted.	4 out of 5 children with ASD had GAS scores over 50, interacting very well or moderately well with the robot and the mediator. Some examples of social behaviours observed, such as verbal communication and gestures.	2, 10, 11
Collaborative Research Project: Developing and Testing a Robot-Assisted Intervention for Children With Autism **(Kostrubiec & Kruck, 2020)**	White, spherical prototype	n = 20 (17 M, 3 F; 5 – 10.2 y/o, low-functioning autism). 2 sets of 7 lessons over 24 weeks.	ABA-based intervention. Special educators taught 2 sets of lessons on requesting and turn-taking together with the robot. Each lesson administered in 2 conditions – w/ robot & w/ ordinary ball. Video recordings analysed by blinded observers using observation grid w/ 16 response categories.	More prosocial behaviours observed w/ ball than w/ robot ($p = 0.026$). No statistically reliable effects on reducing undesirable behaviours. The greater the severity of ASD (i.e. higher SCQ scores), the lower the proportion of prosocial behaviours w/ robot as compared to w/ ball ($p = 0.026$).	Taught but not assessed: 16, 20. Assessed: 1, 7, 9, 11
Effects of a Robot-Enhanced Intervention for Children With ASD on Teaching Turn-Taking Skills **(David et al., 2020)**	NAO	n = 5 (3 M, 2 F; 3 – 5 y/o). 16 sessions over 16 days.	Child exposed to either robot-enhanced intervention or standard intervention each session. Sessions delivered in randomised order. Touchscreen device-based turn-taking task for each session. Turn-taking performance measured along w/ several secondary behavioural outcomes identified by 2 clinical psychologists from video recordings.	Most children benefited to a similar extent for turn-taking skills w/ & w/o robot except for 1 child who had poorer performance w/ robot that was even below his baseline measurements before the interventions. Eye contact improved. Robot-enhanced intervention led to increase in both stereotyped / maladaptive behaviours & adaptive behaviours.	9, 11, 12, 16, 20
Follow the white robot: Efficacy of robot-assistive training for children with autism spectrum disorder **(Ghiglino et al., 2021)**	Cozmo	10 children unable to understand instructions of game or uncomfortable / uninterested in robot. Final n = 24 (19 M, 5 F; avg. age = 5.79 y/o; avg. IQ = 58.08; DSM-5 autism levels 1 – 3). 10 sessions over 5 weeks.	2-period crossover design: standard therapy w/ & w/o robot-assisted training. Robot turned to look at 1 of 2 lit cubes and then looked back at child. Child asked by healthcare professional to identify which cube robot looked at. Robot reacted w/ feedback. ESCS scores evaluated.	Standard therapy + robot-assisted training more effective in improving tendency to initiate social interaction ($p = 0.026$) & behavioural requests ($p = 0.007$) than standard therapy alone. No significant / marginal diff for responding to ($p > 0.05$) & maintaining social interaction ($p = 0.051$), responding to behavioural requests ($p = 0.050$) & initiating, responding to & maintaining JA behaviours ($p > 0.05$).	2, 8, 14, 19, 20

Fostering Emotion Recognition in Children with Autism Spectrum Disorder **(Silva et al., 2021)**	Zeno R50	n = 6 (4 M, 2 F; 8 – 9 y/o). Seven 2 – 3-min sessions.	Imitation activity: robot displays a facial expression, child prompted to exhibit the same expression. Emotions activity: robot asks the child to display a facial expression. No. of right & wrong responses + response time measured.	In general, increase in no. of right responses & decrease in no. of wrong responses w/ fluctuations across the sessions. Response time appeared to increase however.	8, 10
Getting Engaged: Assisted Play with a Humanoid Robot Kaspar for Children with Severe Autism **(Zorcec, Robins & Dautenhahn, 2018)**	Kaspar	n = 2 (1 M, 1 F). 1 year of intervention w/ sessions once / twice a week.	Imitation games, teaching greetings, teaching emotions w/ co-present therapist & parents.	Parents reported that children started using greetings in everyday life after 10 sessions w/o anxiety, resistance or discomfort. They also knew the diff between happiness & sadness w/ use of expressions & gestures in real life. Initial forceful behaviours towards robot (e.g. eye-poking) drastically declined after behaviour was pointed out to be hurtful & unacceptable.	7, 8, 15, 17
Human–Robot Interaction in Autism Treatment: A Case Study on Three Pairs of Autistic Children as Twins, Siblings, and Classmates **(Taheri et al., 2018)**	NAO + Alice-R50	n = 6 (6 M; 7 y/o twins: 1 high-functioning, 1 low-functioning; 10 & 15 y/o siblings: high-functioning; 6 & 7 y/o classmates: high-functioning). 12 sessions over 3 months.	ABA-based intervention. Imitation & JA games. 4 types of measuring instruments: quantitative content analysis of video records, GARS + ASSP questionnaires, blind psychologist assessments & interviews w/ parents.	Autism severity decreased based on GARS scores. ASSP results for the 5 high-functioning children showed that their social skills improved. However, no change was found in the ASSP result for the low-functioning child.	9, 10
Improving social skills in children with ASD using a long-term, in-home social robot **(Scassellati et al., 2018)**	Jibo	n = 12 (7 M, 5 F; 6 – 12 y/o, avg. age = 9.02 y/o; avg. non-verbal IQ = 94.17). Daily sessions for 1 month.	Single-subject withdrawal design. 6 interactive games: 1 story game to target social & emotional understanding, 2 barrier games to target perspective-taking, 3 ordering & sequencing games. System adapted difficulty of games w/ child's performance. Likert survey for caregivers & JA assessment (Bean & Eigsti, 2012) carried out.	Caregivers reported children's increased eye contact ($p = 0.03$), initiation of communication ($p = 0.014$) & response to communication (0.012). Increase in JA scores following robot intervention ($p < 0.05$).	2, 9, 19
Pivotal Response Treatment with and without robot-assistance for children with autism: a randomized controlled trial **(van den Berk-Smeekens et al., 2021)**	NAO	n = 73 (61 M, 12 F; 3 – 8 y/o). 20 weekly sessions.	3-armed RCT: PRT, PRT + robot, TAU. Three-term contingency learning trials for social communicative skills implemented. Percentage of spontaneous appropriate initiations by child calculated. SRS scores evaluated by parents & teachers. CGI-I & ADOS-2 scores evaluated by blinded clinicians.	All 3 grps displayed positive treatment outcomes. No significant diff between PRT, PRT + robot & TAU for SRS scores at end of treatment. However, at 3-month follow-up, SRS scores for PRT + robot higher than other 2 grps ($p = 0.008$). No significant diff for CGI-I between the 3 grps. More children from PRT + robot displayed a decrease in ADOS-2 severity category than other 2 grps ($p = 0.008$).	9, 20
Playful Interaction with Humanoid Robots for Social Development in Autistic Children: a Pilot Study **(Cervera, del Pobil & Cabezudo, 2019)**	NAO	n = 14 (3 – 5 y/o). Weekly sessions over 10 months.	2 games played w/ either caregiver or robot. Expressive activity: child is shown a card w/ an animal / vehicle / colour / clothing item & asked to name the item. Understanding activity: child shown 2 cards with items from same class & asked to point at requested item. Instances of watching, pointing & utterances counted. BDI scores evaluated.	In communicative & social domains, no variations in BDI scores pre & post-intervention. In fact, a few children in control grp showed improvements in BDI scores. Control grp also exhibited more pointing & uttering than grp w/ robot.	7, 9, 11
Robot-based intervention may reduce delay in the production of intransitive gestures in Chinese-speaking preschoolers with autism spectrum disorder **(So, Wong, Lam et al., 2018b)**	NAO	n = 45 (30 ASD, 15 TD; 4 – 6 y/o). Four 30-min sessions.	Intervention grp, waitlist control grp & TD grp. Robot narrated 5 stories & gestured. Children told to imitates gestures. Accuracy & appropriateness of gestures assessed in pre & post-tests w/ novel, non-training stories.	Intervention grp produced more accurate & appropriate gestures in training & novel stories w/ more concomitant emblematic gestures compared to the waitlist control grp. Positive outcomes maintained even in delayed post-tests. Accuracy of gestural production in the intervention grp was comparable to TD grp in delayed post-tests w/ novel stories, suggesting that children w/ ASD might be able to catch up to TD counterparts.	4, 7
Robot-based play-drama intervention may improve the narrative abilities of Chinese-speaking preschoolers with autism spectrum disorder **(So et al., 2019)**	2 × NAO	n = 26 (23 M, 3 F; 4 – 6 y/o). Nine 45-min sessions.	Intervention grp & waitlist control grp. Robots performed a drama twice. Child was invited to roleplay in each role w/ robots & then w/ human researcher. Narrative elicitation pre & post-tests administered.	Improvements in number of narrative clauses ($p < 0.001$), proportion of complex clauses ($p < 0.001$), goal-based stories ($p < 0.001$) & stories w/ cognitive ($p < 0.004$) inferences & no. of deictic gestures per clause ($p < 0.01$). Non-significant effect on proportion of stories w/ affective ($p < 0.07$) inferences, no. of iconic ($p < 0.18$), emblematic ($p < 0.03$) & speech beat gestures ($p < 0.07$) per clause.	7, 9

Robot-Enhanced CBT for dysfunctional emotions in social situations for children with ASD **(Pop, Vanderborght & David, 2017)**	Keepon	n =27 (20 M, 9 F; 6 – 12 y/o, avg. age = 8.7 y/o). Six 2-hour weekly sessions.	CBT-based intervention. Intervention grp & TAU control grp. Social scenario was presented to child & questions were posed to child. Robot provided feedback on their social answers & strategies. Frequency of appropriate chosen strategies, rational & irrational beliefs, emotional intensity as adaptive & maladaptive behaviours measured.	Improvement in rational beliefs ($p = 0.007$) & emotional intensity ($p = 0.001$) compared to control grp. No significant differences for social knowledge ($p = 0.567$), irrational beliefs (0.532) & adaptive behaviours ($p = 0.406$).	12, 14
Robot-Mediated Imitation Skill Training for Children with Autism **(Zheng et al., 2016)**	NAO	6 children withdrew due to distress / refusal to sit in chair. Final n = 16 (8 ASD, 8 TD; avg. age = 3.84 y/o). Duration not stated.	2 sessions each w/ human & robot. 2 gestures per session w/ 2 trials for each gesture. Trial I: Robot demonstrated gesture twice & prompted child to imitate. If correct, praise given. Otherwise, feedback given. Trial II: If correct imitation in trial I, trial II is simply a repeat of trial I. Otherwise, robot entered mirroring mode & followed the child's movements after demonstrating gesture. If correct, session ended. Otherwise, final 2 demonstrations & last imitation attempt for child. Imitation accuracy graded by computer algorithm.	Children w/ ASD spent 11% more time paying attention to the robot than human therapist although this is not statistically significant ($p = 0.0663$). Children displayed better imitation skills more quickly in robot sessions than human sessions. Trial I scores w/ robot better than w/ human ($p = 0.0494$).	10
Robot-Mediated Social Skill Intervention Programme for Children with Autism Spectrum Disorder: An ABA Time-Series Study **(Chung, 2021)**	NAO	n = 15 (5 – 11 y/o). 12 weekly sessions.	3 phases: baseline, intervention, generalisation. Social game, story-based activity + singing & movement to music in each session. Freq. & duration of eye contact & freq. of verbal initiation measured.	Improvements in eye contact freq. ($p < 0.01$), duration ($p < 0.01$) & verbal initiation freq. ($p < 0.01$). during intervention phase. Decline in all variables after robot was withdrawn in generalisation phase although these were still higher than the baseline phase.	9, 11
Self-initiations in young children with autism during Pivotal Response Treatment with and without robot assistance **(De Korte et al., 2020)**	NAO	n = 44 (37 M, 7 F; 3 – 8 y/o). 20 weekly sessions.	2 grps: PRT w/ or w/o robot. 9 game scenarios w/ 7 levels of complexity for HRI based on the motivational techniques of PRT. Functional & social self-initiations identified from video coding. SRS scores evaluated by parents & teachers.	Growth in total self-initiations for both grps. PRT + robot grp showed larger growth in functional self-initiations than PRT grp. No grp diffs in growth in social self-initiations. Change in social awareness subscale of parent-rated SRS linked to change in total self-initiations ($p = 0.011$) & functional self-initiations ($p = 0.016$). No significant correlations between self-initiations & the other SRS scales for social-communicative skills.	9, 20
Social Skills Training for Children with Autism Spectrum Disorder Using a Robotic Behavioral Intervention System **(Yun et al., 2017)**	iRobiQ + CARO	n = 15 (15 M; 4 – 7 y/o; verbal IQ ≥ 60). Eight 30-min sessions.	Treatment grp w/ robot & control grp w/ human therapist. DTT protocol. Robot asked child to identify its facial emotional expression. Eye contact + correct answers rewarded according to reinforcement schedule. Otherwise, encouragement given. 10 trials each session. Freq. of eye contact & facial expression recognition measured. ADOS, EWHA-VABS, SCQ, SRS & K-CBCL scores evaluated.	Symptomatology in the ADOS social & communication domains unchanged ($p > 0.05$). SCQ & SRS showed non-statistically significant improvements. ADOS play domain improved significantly ($p = 0.03$). Eye contact dramatically increased in both grps but to a larger extent in treatment grp ($p < 0.05$). However, eye contact gradually declined in treatment grp but was generally maintained in control grp by the 8th session. Eye contact post-treatment decreased to baseline level. Accuracy of facial expression recognition significantly increased in both grps by 4th session ($p = 0.046$) but no significant increase from 5th – 8th session ($p = 0.104$).	8, 11
The effectiveness of socially assistive robotics in children with autism spectrum disorder **(Fachantidis, Syriopoulou-Delli & Zygopoulou, 2018)**	Daisy	n = 4 (7 – 12 y/o; IQ ≥ 70). Eight 30-min sessions.	Single-case design intervention for the 4 children. Robot & teacher carried out 4 sessions each. Tasks designed to be appropriate for the needs and capacities of each child. Quality of interaction measured via freq. of eye contact, proximity (child getting up from seat to approach partner) & spontaneous verbal interaction. Involvement in teaching process measured via degree of difficulty in paying attention (not answering questions / interrupting partner), inability to sit still & degree of difficulty following instructions.	All parameters associated w/ quality of interaction & involvement in teaching process higher when child interacted w/ robot compared to teacher.	9, 11

The Effects of Long-Term Child–Robot Interaction on the Attention and the Engagement of Children with Autism **(van Otterdijk, M. T. H. et al., 2020)**	NAO	n = 6 (5 M, 1 F; 3 – 8 y/o, avg. age = 5.17 y/o; avg. TIQ = 94.20). 20 sessions over 6 months.	Child could select different games: card games, building w/ LEGO or forming a jigsaw puzzle. Each game had 3 variations & 7 levels. Robot movements & speech used to elicit targeted behaviours, reward child, keep child focused & prompt child. Attention measured via coding scheme comprising gaze behaviour & arm / hand behaviour. Engagement measured via accompanying positive facial expressions & gestures.	Children remained attentive & engaged to robot & game throughout the 20 sessions. Increase in attention & engagement toward other humans in the therapy room, particularly w/ the parent.	11
The impact of robotic intervention on joint attention in children with autism spectrum disorders **(Kumazaki et al., 2018)**	CommU	n = 68 (30 ASD, 38 TD; 20 M, 10 F; 5 – 6 y/o). Single 15-min session.	Intervention grp w/ robot & control grp w/ human. 3-part interaction sequence: children interacted w/ human A, then robot / human B, then human A. Robot / human called out "Ne!", gazed towards child for 1 s, then towards an image on child's left / right for 3 s. Achievement of JA defined as child looking at the correct image within the 3-s window.	On average, children with ASD demonstrated better JA response w/ robot than human B ($p < 0.01$). The improvement was maintained w/ human A ($p < 0.01$) after robot was replaced. However, only 8 of the 16 children w/ ASD in intervention grp actually had improved JA responses although no children in intervention grp worsened after interacting w/ robot.	19
The Socially Assistive Robot Daisy Promoting Social Inclusion of Children with ASD **(Pliasa, Velentza & Fachantidis, 2021)**	Daisy	n = 18 (6 ASD, 12 TD; 5 – 9 y/o). Duration not stated.	Step 1: Robot initiated & guided interactions between teacher, TD child and child w/ ASD. Step 2: Teacher no longer present. Robot continues w/ supportive role. Step 3: Robot declared it was "tired" & stayed as an observer. Child w/ ASD interacted w/ a new TD child. TEACCH information & discussion skills scales used to assess children.	Children w/ ASD participated more willingly in dialogue, providing information about themselves & even initiating conversation. Improvement in assessed TEACCH scores from step 1 to step 2 w/ a slight decline from step 2 to step 3 although this was still higher than the baseline level at step 1.	2, 9
Training Autistic Children on Joint Attention Skills with a Robot **(Carlson et al., 2018)**	CuDDler	n = 20 (20 M; 4 – 6 y/o, avg. age = 5.3). Eight 10-min sessions.	Training grp w/ robot & control grp w/ teddy bear. Robot turned its head to look at 1 of 2 screens. Both screens displayed 2 similar objects that only varied by colour. Robot asked, "What colour is this?" Robot praised the child if they were correct. ESCS scores evaluated pre & post-training.	The training grp's ESCS scores improved for IJA ($p = 0.013$) & RJA ($p = 0.023$). The control grp had no significant improvements.	19
Using a social robot to teach gestural recognition and production in children with autism spectrum disorders **(So, Wong, Lam et al., 2018a)**	NAO	n = 13 (10 M, 3 F; 6 – 12 y/o). Eight 30-min sessions.	Intervention grp & waitlist control grp. Phase I: gesture recognition. Phase II: gesture production. 4 sets of narrations: S1 – S4 (S1 + S2: robot-narrated training scenarios, S3: robot-narrated non-training scenarios, S4: human-narrated non-training scenarios). During trainings, robot narrated scenarios from S1 & S2 produced corresponding gestures. Additionally, child requested to imitate gestures in phase II. Accuracy of gestural recognition & production assessed in pre & post-tests w/ S1 – S4. Visual-motor coordination skills evaluated via Beery VMI test & VP subtest.	Improvement in gestural recognition from pre-test to immediate post-test for both S1 + S2 ($p < 0.001$) & S3 + S4 ($p < 0.001$). Positive outcomes maintained in delayed post-tests. Improvement in gestural production from pre-test to immediate post-test for S1 + S2 ($p < 0.007$) & S3 ($p = 0.05$). Non-statistically significant change for S4 ($p < 0.16$). Positive outcomes for S1 – S3 maintained in delayed post-tests.	4, 7
Using the humanoid robot Kaspar in a Greek school environment to support children with Autism Spectrum Condition **(Karakosta et al., 2019)**	Kaspar	n = 7 (5 M, 2 F; 7 – 11 y/o). 2 – 3 sessions per week for 10 weeks.	Play-based sessions. Cause & effect games, imitation games & turn-taking games. Evaluation questionnaire, video recordings & interviews w/ school's psychologist + teachers used to measure outcomes.	For the evaluation questionnaire, non-significant change in sensory development ($p = 0.58$), cognitive development ($p = 0.93$) & social + emotional ($p = 0.55$) domains. Improvement in communication domain ($p = 0.08$) approaching significance. Significant improvement in psychomotor domain ($p = 0.05$). Varying results for coded behaviours from video recordings.	9, 10, 11, 16

3.1 Gestural Recognition and Production

So et al. (2018) conducted a substantial amount of research in this domain. They utilised a NAO robot to narrate stories and teach children with ASD to recognise and subsequently produce eight commonly used gestures. Comprehensive post-tests were carried out, one immediately after the intervention and another two weeks after. These post-tests comprised four sets of narrations, two of which had been used in training and the other two of which were novel, non-training narrations that the child had not seen before so as to test for generalisation of the skills. Of the two sets of non-training narrations, one was presented by the robot while the other was presented by a human to test for generalisation in human-human interaction (HHI). The authors found that in the immediate post-tests with training narrations, the intervention group improved in gestural recognition ($p < 0.001$) and production ($p < 0.007$) compared to the control group. For non-training narrations, the intervention group improved in gestural recognition ($p < 0.001$) and production but only with the robot ($p = 0.05$). The improvement in gestural production with the human was not statistically significant ($p < 0.16$). The aforementioned positive outcomes were maintained in the delayed post-tests two weeks later. Overall, this suggests that gestural recognition and production can be effectively taught by SARs. Moreover, these gains are sustained and can be generalised to some degree in novel contexts.

In a separate study, So et al. (2018) engaged a larger sample size of children with ASD and also introduced a control group with typically developing (TD) children. They found similar improvements in gestural recognition and production for the intervention group. Notably, the children with ASD had comparable performances to the TD children in the delayed post-tests, suggesting that they could catch up to their TD counterparts. Nevertheless, the authors acknowledged that there was no evidence that the children applied their enhanced gestural production skills to their everyday lives.

Finally, So et al. (2019) conducted a play-drama intervention using NAO robots with the primary goal of improving narrative abilities in children with ASD although they also counted the frequency of co-speech gestures as an auxiliary measure. They found an improvement ($p < 0.001$) in the gesture frequency for the intervention group in the immediate and delayed post-tests. However, this increase was predominantly contributed by the increase in deictic gestures ($p < 0.001$) rather than iconic ($p < 0.18$), emblematic ($p < 0.03$), or speech beat gestures ($p < 0.07$).

3.2 Non-Verbal Abilities

Several studies examined the recognition and production of emotional or facial expressions. Silva et al. (2021) conducted an exploratory study on facial expression imitation and production with the

Zeno R50 robot. Three of the six children had positive trends in the accuracy of their facial expression imitation and production while two had fluctuating performances. The final child was more interested in the robotic components than the activity or displayed signs of tiredness or irritation. Consequently, the child ended most sessions unsuccessfully with inconclusive results.

Yun et al. (2017) used the iRobiQ and CARO robots in an intervention based on the discrete trial training (DTT) method. Both the intervention group with the robots and the control group with a human therapist showed similar improvement trends in their facial emotion recognition accuracy.

Zorcec, Robins and Dautenhahn (2018) carried out a year-long intervention with two children using a KASPAR robot. The children's parents observed that the children knew how to distinguish between happiness and sadness and express it in their daily lives. The authors also noted that the children learnt expressions associated with happiness much faster and more easily than those associated with negative emotions.

3.3 Verbal Abilities

There was a surprising dearth of studies which focused on improving the verbal abilities of children with ASD despite communication having always been conventionally and closely associated with the verbal aspect of it. Most studies that did assess verbal abilities in one way or another did not target it as a primary objective, examined it generically without much depth, or evaluated it as part of a more general combined set of coded behaviours.

The exception was the robot-based play-drama intervention by So et al. (2019) although the authors investigated this from a highly specific and niche angle – namely the narrative abilities of children with ASD. They found their intervention to be effective in improving narrative length ($p < 0.001$), syntactic complexity ($p < 0.001$), and narrative structure ($p < 0.001$). The proportion of narrated stories with cognitive inferences (references to a character's purposes, knowledge, and thoughts) also increased ($p < 0.004$) although that for affective inferences (references to a character's emotions) showed a non-statistically significant increase ($p < 0.07$).

A few studies found improvements in the spontaneous initiation of verbal communication or utterances (Boccanfuso et al., 2017; Chung, 2021; De Korte et al., 2020; Scassellati et al., 2018; van den Berk-Smeekens et al., 2021) and response to speech prompts or communication bids (Karakosta et al., 2019; Scassellati et al., 2018). However, De Korte et al. recognised that these initiations were mostly functional rather than social in nature. A study (Fachantidis, Syriopoulou-Delli & Zygopoulou, 2018) also found more spontaneous verbal interaction from children in their intervention with the Daisy robot than with a human teacher. Another study used the Daisy robot to mediate interaction between pairs of children – one TD child and another child with ASD. The children with ASD displayed heightened interest in initiating conversations and providing information about themselves. This diminished slightly when the robot was relegated to an observer role but nevertheless, still remained higher than the baseline level.

In the intervention by David et al. (2020) that aimed to teach turn-taking skills, the authors measured the frequency of contextually appropriate verbal utterances as a secondary outcome. Unfortunately, they could not draw any conclusions as zero instances of this behaviour was recorded from three of the five children throughout the intervention. It was not stated whether these children already had non-verbal ASD although their ADOS scores indicated they had low-functioning ASD. No significant trends were present for the other children.

There was a study (Cervera, del Pobil & Cabezudo, 2019) that found counterproductive effects in their game-based robot interactions with children. The children who interacted with a human therapist showed an increase in their frequency of utterances while no progress was seen for the children that interacted with the robot. The authors stated that it was almost as if the robot impeded the development of these verbal skills. It is worth noting that apart from the delivery of pre-programmed speech lines and queries, there were very little inherent social elements to the games, which had its core premise solely based on identifying physical objects from various categories.

3.4 Imitation

Zheng et al. (2016) designed a NAO robot-mediated imitation skill training architecture which could operate autonomously to train, provide feedback and assess the child in their ability to imitate four basic gestures. They found that the children had better imitation scores with the robot than with a human therapist, particularly for the first trial where the gestures were demonstrated for the first time ($p = 0.0494$). Nevertheless, by the second trial, the children's imitation scores with the human therapist had improved ($p = 0.0006$) to catch up to the imitation scores with the robot. There were no significant improvements between the two trials with the robot. The authors pointed out the unresolved question of whether the differences in performance between the robot and human therapist conditions were merely an artefact of novelty.

Taheri et al. (2018) used a NAO and an Alice robot for imitation games with varying difficulties, ranging from easy gross motor skills to the more difficult playing of a xylophone. Nevertheless, four out of six children encountered the ceiling effect in the easier games. One child also had difficulty in one-legged balance and this adversely affected his imitation scores. This highlights how confounding factors can complicate the assessment of other abilities. Overall, there were no statistically significant improvement trends although the psychologist's assessment did note that two children did show some improvement in their fine imitation abilities.

While Ali et al. (2019) did initially target imitation as an aim of their multi-robot therapy, there was little meaningful analysis of the actual imitation outcomes in their paper. However, the authors did publish the graphs of the imitation accuracy for each child. A cursory inspection reveals no clear trends and large variations in the imitation accuracy across the intervention period and amongst the children.

Two studies included imitation games using a Kaspar (Karakosta et al., 2019) and a MARIA robot (Valadão et al., 2016) as part of a more broad-based intervention. Karakosta et al. did not find any statistically significant changes to unprompted ($p = 0.312$) and prompted imitations ($p = 0.812$) overall. However, they noted that five of the seven children did show some positive trends in unprompted initiation while two did not. Valadão et al. found that the median frequency of imitations in children with ASD was comparable to that of the TD control group. However, their imitation game comprised two extremely simple movements – going towards and away from the robot. The claim that a child has sufficiently demonstrated an act of imitation every time they moved towards or away from a robot might be contentious for some.

3.5 Social Routines

There was only one study (Zorcec, Robins & Dautenhahn, 2018) that reported on social routines as an outcome. In the year-long intervention with a KASPAR robot, the practising of social routines was included as a part of every session. The children's parents reported that after ten sessions, the children started using greetings in their daily lives without anxiety, resistance or discomfort.

3.6 Turn Taking

David et al. (2020) set out to teach turn taking skills using discrete trial training with a NAO robot and compared the results to training delivered by a human therapist. The authors found that the children generally benefited to similar extents in their turn taking performance for both the robot and human conditions. However, there were variations in the results amongst the children. Specifically, one of the five children benefited more from training with the human therapist than with the robot. The authors highlighted that this child started out with higher baseline turn taking performance and posited that children with prior turn taking skills would benefit less from training with a robot. In contrast, there was also another child that benefited more from training with the robot than with the human therapist. The small sample size precluded further conclusions.

3.7 Joint Attention

There were multiple studies that explicitly targeted JA as a primary objective of their SAR intervention. Ali et al. (2019) utilised two NAO robots that operated concurrently, autonomously and adaptively via a least-to-most stimulus protocol – that is, the robots started with subtler stimulus cues, such as blinking, and progressively moved on to more prominent speech and movement cues if the measured level of RJA from the child did not pass the desired threshold. The authors concluded that RJA in the children had improved based on positive changes to the frequency of eye contact and the delay in gaze shifting. There was also a reduction in the CARS score of the children although it was not stated whether this change was statistically significant.

Zheng et al. (2020) utilised a NAO robot to autonomously provide JA prompts to direct the child to focus on one of two target monitor displays. Similar to the previous study, the level of prompting was adjusted based on the child's performance. The study found no statistically significant improvements in the prompt level required, the target hit rate, and the STAT scores of the children on average. Further examination revealed a plausible link between the children whose STAT scores did improve and their significantly lower baseline ADOS-2 scores ($p = 0.03$) and higher IQ scores ($p = 0.06$). Additionally, the effect of a younger age on decreased prompt level required or increased target hit rate approached significance ($p = 0.07$).

Carlson et al. (2018) expanded on a prior study (Kajopoulos et al., 2015) and utilised the CuDDler robot to target RJA skills with a larger sample size, more training sessions and the addition of a control group. RJA and IJA were assessed via ESCS scores. Interestingly, in the intervention group, improvements were present in both RJA ($p = 0.023$) and IJA ($p = 0.013$) despite the training only targeting RJA. This was in contrast to the results of the prior study which only recorded improvements in RJA. The authors stopped short of concluding that RJA training had somehow had collateral effects on IJA and instead postulated that this was the result of a larger sample size. They also acknowledged potential ambiguity in the results since the control group started off with better baseline RJA and IJA scores and so, despite them not having improvements in their scores, the post-training scores of the intervention group did not surpass that of the control group.

Kumazaki et al. (2018) utilised the CommU robot via WoZ control. CommU was unique since it could turn its eyes, thereby accurately replicating gaze shifting, whereas the previous robots could only simulate it by turning their head. The authors defined achievement of JA as the child looking at the image that the robot was looking at within a 3-second time window. Despite the intervention group only interacting with the robot for a relatively short span of 5 minutes, they had better overall JA achievements than the control group during interaction with the robot ($p < 0.01$) and even when the robot was subsequently replaced by a human ($p < 0.01$). However, it is worth pointing out that only half of the children with ASD in the intervention group actually improved although none of them worsened after interacting with the robot, indicating large variances amongst the children's results.

Two studies did not specifically focus on JA as an objective although they did assess JA following the intervention. Ghiglino et al. (2021) found no significant effects on the tendency to initiate, maintain or respond to JA behaviours as measured by the ESCS following their intervention with a small and simple Cozmo robot. The authors justified that this was consistent with their experimental design which did not focus on this aspect. However, their methodology was remarkably similar to those in the aforementioned studies by Zheng, Carlson, and Kumazaki et al. These studies used a robotic stimulus to direct a child's attention to one of two displays or images. Ghiglino et al. had Cozmo turn to face one of two lit, coloured cubes and asked the child which cube Cozmo looked at.

Scassellati et al. (2018) utilised the Jibo robot in an extended, home-based intervention that involved social skills games on a separate touchscreen device with a caregiver. The robot displayed context-dependent social gaze behaviours, making eye contact or looking at the screen together with the child. JA was gauged via a naturalistic assessment developed by Bean and Eigsti (2012). JA scores on the last day of intervention was significantly higher than the baseline scores ($p < 0.05$) although they returned to baseline levels 30 days after the intervention. The authors also noted a relationship between greater JA gains and lower non-verbal reasoning abilities as measured by the DAS ($p = 0.020$). They found a strong link between low baseline non-verbal reasoning abilities and low baseline JA scores ($p = 0.005$) and therefore suggested that these children had more capacity to develop in their JA skills.

4 Discussion

4.1 Current State of Knowledge

The findings of the above reviewed studies were largely mixed.

The most promising area for SAR interventions appears to be the development of gestural abilities. Not only did the studies find improvements in gestural appropriateness and accuracy, the outcomes were also sustained after approximately two weeks and were applicable to novel testing scenarios. This might be because gestures are generally fixed and unchanging in their meanings, which enables children with ASD to rapidly acquire them through repetition and familiarisation.

The aspects of non-verbal communication that the reviewed studies did look at also showed mostly positive results. Notably, all studies that measured eye contact found improvements in the frequency and duration of eye contact. The accuracy of facial expression recognition was also improved in most children with ASD through SAR interventions. However, there were no studies to investigate the other aspects of non-verbal communication, such as proxemics and haptics. It might be a beneficial area for future studies to investigate given that children with ASD have atypically reduced personal space (Asada et al., 2016) and hyper-reactivity to touch (Quinde-Zlibut et al., 2020).

The research into verbal communication was found to be wanting. Besides the increase in the frequency and lengths of spontaneous utterances within the experimental context noted by a few studies and the improvements in narrative abilities observed by a single study, there were no studies that had speech and language therapy as their central focus. This is rather surprising considering how researchers have been keen to apply SAR technologies for first and second language learning, covering a whole gamut of skills such as grammar, vocabulary, speaking, and sign language, albeit only for TD children and adults (Kanero et al., 2018; van den Berghe et al., 2019).

SAR interventions for JA yielded highly inconsistent results. The studies that explicitly targeted JA all had methodologies that elicited RJA rather than IJA behaviours. Some found improvements in RJA, one found no effect at all, and another found improvements in IJA despite not training it. As for the studies that did not focus on JA but still assessed it as an outcome, one observed no effects while the other found an improvement which eventually declined to baseline levels after 30 days.

Imitative abilities were largely unaffected by SAR interventions. Only one study reported some sort of improvement in the children who received the SAR intervention and even so, this was swiftly caught up to by children who had a conventional human therapist. The remaining studies found no significant improvements. This seems to suggest that either the use of SARs is ineffective in developing imitation or that the methodologies of these studies are flawed.

Additionally, several general observations can be drawn from the reviewed literature, regardless of which particular aspect of social skills the study focused on.

4.1.1 Heterogeneity of Participants and Outcomes

Firstly, there is an acute awareness in a majority of the reviewed studies of the inhomogeneity of the ASD children population and its potential impact on the variability of treatment outcomes amongst the children. ASD presents itself as a diverse spectrum of symptoms that differs from individual to individual. Hence, different children with ASD respond differently to the same treatment. The distress experienced by some children in at least three studies (Ghiglino et al., 2021; Zheng et al., 2016; Zheng et al., 2020) suggests that the mere prospect of SAR interventions might not even be viable for these children. Recent studies have recognised the possibility of children exhibiting challenging behaviours with SARs that might result in harm to themselves (Alhaddad, Cabibihan & Bonarini, 2019) and have therefore proposed methods to mitigate this (Alban et al., 2021). Notwithstanding the adverse reactions shown by some children towards SARs, there is still the outstanding question of whether the effectiveness of SARs in delivering their purported treatment benefits is contingent on any traits of the child with ASD and if so, which ones it is dependent on. Interestingly, there appears to be two contrasting lines of thought put forth by different studies. Some researchers postulate that children with higher baseline cognitive or social competencies have less room for improvement and thus, benefit less from SAR interventions. In contrast, other researchers posit that these children with higher baseline competencies have a greater capacity to learn and develop their social skills. Unfortunately, none of the studies could further stratify their participants to conclusively support their hypotheses due to the insufficiently large sample sizes and the differing tools used to diagnose the children with ASD which precluded a fair comparison. A review by Pennisi et al. (2016) attempted to investigate whether gender, IQ, and age affected therapeutical outcomes with SARs but encountered similar difficulties due to the paucity of studies that made explicit comparisons between different demographic groups. It would be ideal for subsequent studies to include pertinent statistical data about the participants to facilitate the conduct of future meta-analyses.

4.1.2 Generalisability and Sustainability of Outcomes

Currently, there is a lack of evidence for whether SAR intervention outcomes are generalisable (transferable and applicable to everyday social interactions beyond the controlled experimental environment) and sustained (maintained for a period of time or even indefinitely without substantial decline). A few studies did attempt to measure generalisability through the use of novel contexts in testing (Kumazaki et al., 2020; So et al., 2018; So et al., 2019) although these were still performed in a controlled testing environment and might not be reflective of social interactions in the children's daily lives. Other studies used improvements in scores recorded by ASD assessment tools to justify that actual social gains had been made. However, Taheri et al. (2018) warned of the risk of overgeneralisation, whereby a behaviour is employed in irrelevant or inappropriate situations and in the worst case, repetitively exhibited to the point of becoming a new stereotyped behaviour. A few studies also carried out post-tests with delay periods ranging from a week to a month after the SAR intervention. The results were mixed with studies observing a maintenance in the children's social abilities (So et al., 2018; So et al., 2019), a mild decline (Chung, 2021), or a regression to baseline levels (Scassellati et al., 2018; Yun et al., 2017).

4.1.3 Better Outcomes With SARs Compared to Human Interventions

On a more promising note, many studies found the improvements in social skills from SAR interventions to be greater than (Boccanfuso et al., 2017; De Korte et al., 2020; Fachantidis, Syriopoulou-Delli & Zygopoulou, 2018; Ghiglino et al., 2021; Kumazaki et al., 2018; Pop, Vanderborght & David, 2017; van den Berk-Smeekens et al., 2021; Yun et al., 2017; Zheng et al., 2016) or at least on par (David et al., 2020) with those from equivalent human interventions. This trend concurs with a previous review by Pennisi et al. (2016).

4.2 Methodological Limitations in Extant Research

Several methodological limitations have been observed in the review literature.

4.2.1 Sample Size and Selection

Firstly, at least eight studies had small sample sizes with 10 children or less while only two had more than 50 children. Studies with small sample sizes and concomitantly, a low statistical power are more likely to encounter false positives although the results might be statistically significant. Even if the effect is true, the effect size tends to be inflated. Brysbaert (2019) underscored that to detect a moderate effect (Cohen's d = 0.4) which would be of relevance to most psychological studies (Open Science Collaboration, 2015), a minimum of 52 participants would already be required for a basic comparison of two within-subjects conditions in a study with a statistical power of 80%.

Notably, researchers must also account for participant dropout rates in these studies as a result of children withdrawing due to distress or the inability to comply with instructions. These rates ranged from 13% (3 out of 23 children) (Zheng et al., 2020) to as high as 29% (10 out of 34 children) (Ghiglino et al., 2021) of the initially recruited children. Given these non-trivial numbers, there is the possibility that other studies omitted this detail in their papers and included only the children that completed the intervention programmes in their entirety which might result in attrition bias in the reported findings.

A number of studies included as part of their inclusion criteria for the children with ASD a minimum IQ and / or the ability to produce verbal utterances comprising a minimum number of words (Chung, 2021; De Korte et al., 2020; Fachantidis, Syriopoulou-Delli & Zygopoulou, 2018; Kim, Elizabeth S. et al., 2013; Kumazaki et al., 2020; van den Berk-Smeekens et al., 2021; van Otterdijk, M. T. H. et al., 2020; Yun et al., 2017). While this is not a methodological flaw in itself if there are reasonable justifications for excluding certain subsections of the ASD population for a study, it simply means that the researchers should be cognisant of the fact that their findings are skewed by selection bias and they should therefore be wary of generalising their findings for all ASD children. Unfortunately, this also means that there is less available research and data to back up the efficacy of SAR interventions for children with lower cognitive abilities or non-verbal ASD even though they are precisely the ones who require the most support and intervention.

4.2.2 Intervention Intensity and Duration

At least four studies had interventions that were extremely short, with a total treatment dosage of less than two hours. This concern was highlighted in a review by Begum, Serna and Yanco (2016) as well and they expressed their doubts that the behaviours of children with ASD could be meaningfully modified within such a short time span. Moreover, any improvements from such a short intervention could possibly be attributed to the novelty factor. Indeed, research has expectedly found that both treatment intensity (i.e. hours per week) and duration (i.e. the number of weeks / months the intervention lasts for) were predictors of the mastery of intervention objectives in children with ASD (Linstead et al., 2017) with duration having a stronger impact than intensity.

4.2.3 Measurement of Outcomes

It was observed that there was a myriad of methods that researchers used to gauge the efficacy of their interventions. These generally fall into the broad categories of screening, diagnostic and assessment tools, performance evaluation, behavioural coding, and interviews or questionnaires with parties close to the child with ASD, such as caregivers and teachers (Table 3).

Nearly half of the reviewed studies used existing screening, diagnostic or assessment tools, including but not limited to the ADOS-2, ASSP, CARS, CBCL, CGI-I, ESCS, EVT-2, GARS, MIS, SCQ, SRS, STAT, and VABS-II. This review makes no attempt to rank the superiority of the different tools given the lack of an agreed-upon gold standard (Fletcher-Watson & McConachie, 2017) but recommends that researchers use their judgement and discretion to select a sensible tool that is most applicable to their treatment goals. For example, the ESCS is frequently utilised as a structured measure of JA (Roos et al., 2008) and would thus be apt for studies examining interventions for JA.

Researchers also evaluated the performance of children with task-based metrics, such as the accuracy of their answers or the frequency of prompts required. These were relatively objective and straightforward to measure. It was even feasible for the SAR or peripheral devices to evaluate performance automatically and independently during the intervention (Silva et al., 2021; Zheng et al., 2016; Zheng et al., 2020). However, these metrics are useful only within the contextual constraints of the task and do not demonstrate whether acquired social skills, if any, are applied to social interactions beyond the experimental setting.

Behavioural coding was used to identify and label overt behavioural responses which were more subjective or ambiguous in nature. For example, prosocial behaviours might be characterised by smiling and desirable vocalisations amongst other behaviours. The definition of a smile and a desirable vocalisation would be determined by the researchers and recorded by observers. Ideally, these observers should not be the researchers themselves but rather, distinct blinded assessors to prevent observer bias (Hróbjartsson et al., 2013).

Finally, interviews or questionnaires with caregivers or teachers were carried out. These range from open-ended questions to more structured Likert surveys and can often highlight social behaviours that were performed in the children's everyday lives outside of the experimental setting. However, given that consent is required for the interventions with the children with ASD, it is practically impossible for caregivers to be blinded. Hence, they are highly susceptible to observer bias.

Table 3 Advantages and Disadvantages of Different Outcome Measurement Methods

Measurement Method	Advantages	Disadvantages
Screening / diagnostic / assessment tools	- Large range of established, readily available tools - Latest revisions have been updated and refined from years of usage - Provides a general summary of the child's cognitive / social competencies, often with breakdown scores for different domains - Enables inter-study comparisons if the same tool is used	- Some tools need to be administered by trained professionals - Risk of misuse if tool is not utilised for its intended purpose or under its intended conditions - Potential for data dredging given the large amounts of quantitative data returned
Performance evaluation	- Easily measured and can be automatically done *in situ* by the SAR during the intervention - Provides a decent measure of child's engagement with the intervention	- Specific to the task with no information provided about generalisability of the social skills - Improvements in performance might be due to task repetition and familiarisation rather than actual gains in social skills
Behavioural coding	- Systematic way of drawing meaningful patterns and insights from a gamut of observed behaviours - Inter-observer reliability can be ascertained - Provides a good measure of child's engagement with the intervention	- Resource-intensive - Outcomes are highly dependent on the coding scheme employed - Behaviours are usually only observed during the intervention itself and give no information about generalisability and sustainability beyond the intervention
Interviews / questionnaires with caregivers / teachers	- Highlights social behaviours demonstrated outside of the intervention programme - Open-ended questions can reveal collateral effects and unintended consequences of intervention	- Highly susceptible to observer bias - Might be difficult to draw trends from disparate observations by caregivers / teachers

Ultimately, researchers have the prerogative to select the measurement methods which they deem to be the most suitable for analysing outcomes. However, there were several instances where questionable measurement metrics were used. One study averaged two incompatible data types – JA / imitation scores generated from behavioural observations and a score generated from electroencephalography (EEG) results to create an arbitrary combined metric. Another study used goal attainment scaling (GAS) to calculate a T-score. GAS is a rather unique metric since the goals are completely specified by the user and is therefore commonly used to provide a highly individualised measure of therapeutical outcomes (Shankar, Marshall & Zumbo, 2020). However, since the goals and the scales are defined entirely by the user, special care must be taken in the goal selection and scaling process (Shankar, Marshall & Zumbo, 2020) to achieve the proper mean and standard deviation as intended by the developers of GAS. There was no evidence of this in the study. The goals were haphazard while the same GAS metric was applied uniformly to all participants which defeated the purpose of GAS. To this end, this review advocates for future studies to carefully consider the suitability of different outcome measurement methods and subsequently, to avoid data dredging and obfuscation in the presentation of data and results.

4.3 Potential Future Implementation and Research Directions

4.3.1 Short Term

Given the findings from extant literature, this review proposes that the use of SAR interventions to improve eye contact, gestural abilities, facial expression recognition and production, and social routines is a feasible goal within the next decade. The reviewed studies have found promising outcomes in these areas. These objectives are all overt, recognisable behaviours that are relatively fixed and unchanging, lending themselves to be taught in a structured and predictable manner by SARs, such as via the proven DTT approach.

The barriers impeding the above are not technological but rather, methodological and evidential in nature. There are sufficient preliminary, proof-of-concept studies in these areas. The primary focus now should be to narrow the research to implementation gap with randomised, controlled trials. Future studies need to be much more rigorous in their methodologies, particularly with respect to sample sizes, intervention duration, and outcome measurement indicators. The studies must incorporate delayed post-tests involving HHI to verify that the acquired social skills can be generalised to novel social contexts while also being maintained after HRI is ended. For studies investigating facial expression recognition, it might be worth examining whether the fidelity of the facial expressions displayed by the SAR has significant effects on the child's ability to generalise facial expression recognition for human faces. For example, the iRobiQ and CARO robots used by Yun et al. (2017) display facial expressions via LED lights and LCD displays respectively. It might be easier for children with ASD to recognise facial expressions in these simplified models but this might be detrimental when it comes to actual human faces.

The studies should also have a control group undergoing an equivalent intervention but with a human therapist. It is insufficient to merely show that SAR interventions can improve social skills in children with ASD. Studies must prove that these improvements are comparable to or exceeding those achieved from conventional human interventions for clinical professionals to consider it as a viable option. After all, the increased logistical demands of introducing SARs into clinics, schools, and homes must be justified by a commensurate improvement in outcomes beyond the status quo. Once the clinical utility of SARs is proven, the rest is a matter of implementation details.

There are two viable approaches to the employment of SARs. They can be used complementarily as an adjunctive tool by therapists since children with ASDs demonstrate increased engagement with SARs. Alternatively, they can be used supplementarily at home for further training and practice as per the study by Scassellati et al. (2018). Either way, the current technologies are assessed to be mature enough for the above objectives. The reviewed studies have shown that SARs and additional peripheral devices can autonomously identify the child's gaze (Yun et al., 2017), gestures (Zheng et al., 2016), and facial expressions (Silva et al., 2021) and provide feedback in real-time to the child. Additionally, some studies incorporated some form of autonomous adaptability in the SAR, gradually increasing the stimulus or prompting to elicit a behaviour (Ali et al., 2019; Zheng et al., 2020) or adjusting the difficulty of the tasks or games (Scassellati et al., 2018) to promote sustained engagement. The latter is also useful in preventing ceiling or floor effects. While adaptability is well-documented, there is little research about the effectiveness of personalisation for SAR interventions, whereby modules are chosen to match the child's needs and interests. This would be a beneficial area to examine given the heterogeneity of ASD.

4.3.2 Long Term

This review asserts that one of the most crucial steps for progress to be made in the long term in the field of SAR interventions for children with ASD is to ensure tight-knit collaboration between the robotics community and the clinical community. This sentiment is echoed by Kim et al. (2012) who identified a cultural divide between the robotics community and the clinical community in three key aspects: research approach, study design, and publication and dissemination of results. Kim et al. noted that SAR research is often driven by the technological innovations and capabilities of the SAR rather than the needs of the target population. Indeed, in the reviewed literature, many studies meticulously detailed the algorithms and protocols used by the SAR while glossing over the design of the actual therapy. Ultimately, SAR interventions are fundamentally still an intervention for children with ASD and this is an area in which the expertise of the clinical community is unequivocally required. Several of the reviewed studies have started diverging into rather eclectic approaches towards their intervention and while that is not necessarily misguided per se, the intervention approach should still be grounded in psychological theories and special education pedagogies. Collaboration with the clinical community is therefore essential to iteratively improve the design of SARs and SAR interventions for optimal outcomes.

Ideally, SAR interventions should not just be able to target individual, overt social behaviours through repetitive, structured trials. They should be able to target pivotal skills (e.g. motivation, responsiveness to multiple cues, self-management, and self-initiations) that have widespread, collateral effects on various domains of a child's social development (Lei & Ventola, 2017). This proven naturalistic approach is known as pivotal response training (PRT). However, PRT occurs in a natural, open-ended play environment and crucially, is directed by the child. This introduces an added layer of dynamicity into the decision loop of the SARs which makes it more challenging to deal with. Currently, this issue is circumvented via WoZ control. Future research would need to look to advances in affective computing and natural language processing to take in "noisy" speech and human affect signals (Clabaugh & Matarić, 2019) from the children with ASD to eventually wean SARs off the resource-intensive WoZ control.

5 Conclusion

In summary, the field of SAR has been receiving a substantial amount of attention recently. Naturally, researchers have been keen to apply SAR interventions for the development of social skills in children with ASD. Despite the proliferation of a multitude of studies investigating the effects of SAR interventions on various aspects of social skills, the results remain largely mixed and even underwhelming for several aspects. As inconclusive as it may be, the best conclusion one can draw is that SAR interventions are effective for *some* children with ASD in the development of *some* social skills, namely the relatively fixed and unchanging overt behaviours. There is evidence that the heterogeneity of the ASD population affects the treatment outcomes although the exact dependencies have not been elucidated. Extant preliminary studies face methodological issues with their sample sizes, intervention duration, and measurement of outcomes. There is a lack of clear-cut evidence that gains from SAR interventions are generalisable to social interactions in the daily lives of the children and that these gains are sustained. Hence, there remains a stark research to implementation gap. In the long term, integration between the robotics community and the clinical community will be crucial to making advances in this field while affective computing and natural language processing capabilities for the SARs have to be examined in order for the SARs to truly be autonomous.

References

[1] Abner, N., Cooperrider, K. & Goldin-Meadow, S. (2015) Gesture for Linguists: A Handy Primer. *Language and Linguistics Compass.* 9 (11), 437-451. 10.1111/lnc3.12168.

[2] Alban, A. Q., Ayesh, M., Alhaddad, A. Y., Khalid Al-Ali, A., So, W. C., Connor, O. & Cabibihan, J. (2021) Detection of Challenging Behaviours of Children with Autism Using Wearable Sensors during Interactions with Social Robots. *2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN).* pp.852-857.

[3] Alhaddad, A., Cabibihan, J. & Bonarini, A. (2019) Head Impact Severity Measures for Small Social Robots Thrown During Meltdown in Autism. *International Journal of Social Robotics.* 11 10.1007/s12369-018-0494-3.

[4] Ali, S., Mehmood, F., Dancey, D., Ayaz, Y., Khan, M. J., Naseer, N., Amadeu, R. D. C., Sadia, H. & Nawaz, R. (2019) An Adaptive Multi-Robot Therapy for Improving Joint Attention and Imitation of ASD Children. *IEEE Access.* 7 81808-81825. 10.1109/ACCESS.2019.2923678.

[5] American Psychiatric Association. (2013) *Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed.* Arlington, VA, US, American Psychiatric Publishing, Inc.

[6] Asada, K., Tojo, Y., Osanai, H., Saito, A., Hasegawa, T. & Kumagaya, S. (2016) Reduced Personal Space in Individuals with Autism Spectrum Disorder. *PLoS ONE.* 11 (1), 10.1371/journal.pone.0146306.

[7] Bean, J. L. & Eigsti, I. (2012) Assessment of joint attention in school-age children and adolescents. *Research in Autism Spectrum Disorders.* 6 (4), 1304-1310. 10.1016/j.rasd.2012.04.003.

[8] Begum, M., Serna, R. W. & Yanco, H. A. (2016) Are robots ready to deliver autism interventions? A comprehensive review. *International Journal of Social Robotics.* 8 (2), 157-181. 10.1007/s12369-016-0346-y.

[9] Boccanfuso, L., Scarborough, S., Abramson, R., Hall, A., Wright, H. & O'Kane, J. (2017) A low-cost socially assistive robot and robot-assisted intervention for children with autism spectrum disorder: field trials and lessons learned. *Autonomous Robots.* 41 10.1007/s10514-016-9554-4.

[10] Bölte, S. (2014) Is autism curable? *Developmental Medicine & Child Neurology.* 56 (10), 927-931. 10.1111/dmcn.12495.

[11] Brysbaert, M. (2019) How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition.* 2 (1), 16. 10.5334/joc.72.

[12] Cabibihan, J., Javed, H., Jr, M. & Aljunied, S. (2013) Why Robots? A Survey on the Roles and Benefits of Social Robots in the Therapy of Children with Autism. *International Journal of Social Robotics.* 5 10.1007/s12369-013-0202-2.

[13] Carlson, K., Wong, A. H. Y., Dung, T. A., Wong, A. C. Y., Tan, Y. K. & Wykowska, A. (2018) Training Autistic Children on Joint Attention Skills with a Robot. *International Conference on Social Robotics. Social Robotics. November 28 - 30, 2018,* Springer International Publishing. pp.86-92.

[14] Centers for Disease Control and Prevention. (2021) Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. *MMWR Surveillance Summaries.* 70 (11), 1-16. 10.15585/mmwr.ss7011a1.

[15] Cervera, E., del Pobil, A. P. & Cabezudo, M. (2019) Playful Interaction with Humanoid Robots for Social Development in Autistic Children: a Pilot Study. pp.1-6.

[16] Chandler, D. & Munday, R. (2011) *A Dictionary of Media and Communication.* Oxford University Press.

[17] Chung, E. (2021) Robot-Mediated Social Skill Intervention Programme for Children with Autism Spectrum Disorder: An ABA Time-Series Study. *International Journal of Social Robotics.* 13 10.1007/s12369-020-00699-w.

[18] Clabaugh, C. & Matarić, M. (2019) Escaping Oz: Autonomy in Socially Assistive Robotics. *Annual Review of Control, Robotics, and Autonomous Systems.* 2 33-61. 10.1146/annurev-control-060117-104911.

[19] Costa, A. P., Schweich, T., Charpiot, L. & Steffgen, G. (2018) Attitudes of Children with Autism towards Robots: An Exploratory Study. *ArXiv.*

[20] David, D. O., Costescu, C. A., Matu, S., Szentagotai, A. & Dobrean, A. (2020) Effects of a Robot-Enhanced Intervention for Children With ASD on Teaching Turn-Taking Skills. *Journal of Educational Computing Research.* 58 (1), 29-62. 10.1177/0735633119830344.

[21] De Korte, M. W., van den Berk-Smeekens, I., van Dongen-Boomsma, M., Oosterling, I. J., Den Boer, J. C., Barakova, E. I., Lourens, T., Buitelaar, J. K., Glennon, J. C. & Staal, W. G. (2020) Self-initiations in young children with autism during Pivotal Response Treatment with and without robot assistance. *Autism: The International Journal of Research and Practice.* 24 (8), 2117-2128. 10.1177/1362361320935006.

[22] Emanuel, R. & Weir, S. (1976) Catalysing communication in an autistic child in a LOGO-like learning environment. *Edinburgh UK. AISB'76: Proceedings of the 2nd Summer Conference on Artificial Intelligence and Simulation of Behaviour. Jul 12-14, 1976,* Amsterdam, Netherlands, IOS Press. pp.118-129.

[23] Estes, A., Munson, J., Rogers, S. J., Greenson, J., Winter, J. & Dawson, G. (2015) Long-Term Outcomes of Early Intervention in 6-Year-Old Children With Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry.* 54 (7), 580-587. 10.1016/j.jaac.2015.04.005.

[24] Fachantidis, N., Syriopoulou-Delli, C. K. & Zygopoulou, M. (2018) The effectiveness of socially assistive robotics in children with autism spectrum disorder. *International Journal of Developmental Disabilities.* 66 (2), 113-121. 10.1080/20473869.2018.1495391.

[25] Feil-Seifer, D. & Mataric, M. J. (2005) Defining socially assistive robotics. - *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.* pp.465-468.

[26] Fletcher-Watson, S. & McConachie, H. (2017) The Search for an Early Intervention Outcome Measurement Tool in Autism. *Focus on Autism and Other Developmental Disabilities.* 32 (1), 71-80. 10.1177/1088357615583468.

[27] Fuller, E. A. & Kaiser, A. P. (2020) The Effects of Early Intervention on Social Communication Outcomes for Children with Autism Spectrum Disorder: A Meta-analysis. *Journal of Autism and Developmental Disorders.* 50 (5), 1683-1700. 10.1007/s10803-019-03927-z.

[28] Ghiglino, D., Chevalier, P., Floris, F., Priolo, T. & Wykowska, A. (2021) Follow the white robot: Efficacy of robot-assistive training for children with autism spectrum disorder. *Research in Autism Spectrum Disorders.* 86 101822. 10.1016/j.rasd.2021.101822.

[29] Hansen, S. N., Schendel, D. E. & Parner, E. T. (2015) Explaining the increase in the prevalence of autism spectrum disorders: the proportion attributable to changes in reporting practices. *JAMA Pediatrics.* 169 (1), 56-62. 10.1001/jamapediatrics.2014.1893.

[30] Hróbjartsson, A., Thomsen, A. S. S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., Ravaud, P. & Brorson, S. (2013) Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ : Canadian Medical Association Journal.* 185 (4), E201-E211. 10.1503/cmaj.120744.

[31] Huijnen, C. A. G. J., Lexis, M. A. S., Jansens, R. & de Witte, L. P. (2016) Mapping Robots to Therapy and Educational Objectives for Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders.* 46 (6), 2100-2114. 10.1007/s10803-016-2740-6.

[32] Ingersoll, B. (2008) The Social Role of Imitation in Autism: Implications for the Treatment of Imitation Deficits. *Infants & Young Children.* 21 (2), 107–119. 10.1097/01.IYC.0000314482.24087.14.

[33] Ingersoll, B. & Lalonde, K. (2010) The impact of object and gesture imitation training on language use in children with autism. *Journal of Speech, Language, and Hearing Research : JSLHR.* 53 (4), 1040-1051. 10.1044/1092-4388(2009/09-0043).

[34] Jurevičienė, M., Kaffemaniene, I. & Ruškus, J. (2018) Concept and Structural Components of Social Skills. *Baltic Journal of Sport and Health Sciences.* 3 10.33607/bjshs.v3i86.266.

[35] Kajopoulos, J., Wong, A., Yuen, A., Dung, T., Kee, T. & Wykowska, A. (2015) Robot-Assisted Training of Joint Attention Skills in Children Diagnosed with Autism. pp.296-305.

[36] Kanero, J., Geckin, V., Oranç, C., Mamus, E., Küntay, A. & Goksun, T. (2018) Social Robots for Early Language Learning: Current Evidence and Future Directions. *Child Development Perspectives.* 12 146-151. 10.1111/cdep.12277.

[37] Karakosta, E., Dautenhahn, K., Syrdal, D. S., Wood, L. J. & Robins, B. (2019) Using the humanoid robot Kaspar in a Greek school environment to support children with Autism Spectrum Condition. *Paladyn, Journal of Behavioral Robotics.* 10 (1), 298-317. 10.1515/pjbr-2019-0021.

[38] Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R. & Scassellati, B. (2013) Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders.* 43 (5), 1038-1049. 10.1007/s10803-012-1645-2.

[39] Kim, E., Paul, R., Shic, F. & Scassellati, B. (2012) Bridging the Research Gap: Making HRI Useful to Individuals with Autism. *Journal of Human-Robot Interaction.* 1 10.5898/jhri.v1i1.25.

[40] Kostrubiec, V. & Kruck, J. (2020) Collaborative Research Project: Developing and Testing a Robot-Assisted Intervention for Children With Autism. *Frontiers in Robotics and AI.* 0 10.3389/frobt.2020.00037.

[41] Kumazaki, H., Muramatsu, T., Yoshikawa, Y., Matsumoto, Y., Ishiguro, H., Kikuchi, M., Sumiyoshi, T. & Mimura, M. (2020) Optimal robot for intervention for individuals with autism spectrum disorders. *Psychiatry and Clinical Neurosciences.* 74 (11), 581-586. 10.1111/pcn.13132.

[42] Kumazaki, H., Yoshikawa, Y., Yoshimura, Y., Ikeda, T., Hasegawa, C., Saito, D. N., Tomiyama, S., An, K., Shimaya, J., Ishiguro, H., Matsumoto, Y., Minabe, Y. & Kikuchi, M. (2018) The impact of robotic intervention on joint attention in children with autism spectrum disorders. *Molecular Autism.* 9 46. 10.1186/s13229-018-0230-8.

[43] Lei, J. & Ventola, P. (2017) Pivotal response treatment for autism spectrum disorder: current perspectives. *Neuropsychiatric Disease and Treatment.* 13 1613-1626. 10.2147/NDT.S120710.

[44] Linstead, E., Dixon, D. R., Hong, E., Burns, C. O., French, R., Novack, M. N. & Granpeesheh, D. (2017) An evaluation of the effects of intensity and duration on outcomes across treatment domains for children with autism spectrum disorder. *Translational Psychiatry.* 7 (9), e1234. 10.1038/tp.2017.207.

[45] Mundy, P. & Newell, L. (2007) Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science.* 16 (5), 269-274. 10.1111/j.1467-8721.2007.00518.x.

[46] Open Science Collaboration. (2015) Estimating the reproducibility of psychological science. *Science.* 10.1126/science.aac4716.

[47] Peca, A., Simut, R., Pintea, S., Costescu, C. & Vanderborght, B. (2014) How do typically developing children and children with autism perceive different social robots? *Computers in Human Behavior.* 41 268-277. 10.1016/j.chb.2014.09.035.

[48] Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S. & Pioggia, G. (2016) Autism and social robotics: A systematic review. *Autism Research: Official Journal of the International Society for Autism Research.* 9 (2), 165-183. 10.1002/aur.1527.

[49] Pliasa, S., Velentza, A. -. & Fachantidis, N. (2021) The Socially Assistive Robot Daisy Promoting Social Inclusion of Children with ASD. *Studies in Computational Intelligence.* 982 89-102. 10.1007/978-3-030-77022-8_8.

[50] Pop, C., Vanderborght, B. & David, D. (2017) Robot-Enhanced CBT for dysfunctional emotions in social situations for children with ASD. *Journal of Evidence-Based Psychotherapies.* 17 119-132. 10.24193/jebp.2017.2.7.

[51] Posar, A. & Visconti, P. (2017) Autism in 2016: the need for answers. *Jornal De Pediatria.* 93 (2), 111-119. 10.1016/j.jped.2016.09.002.

[52] Poysa-Tarhonen, J., Awwal, N., Hakkinen, P. & Otieno, S. (2021) Joint attention behaviour in remote collaborative problem solving: exploring different attentional levels in dyadic interaction. 10.1186/s41039-021-00160-0.

[53] Quinde-Zlibut, J. M., Okitondo, C. D., Williams, Z. J., Weitlauf, A., Mash, L. E., Heflin, B. H., Woodward, N. D. & Cascio, C. J. (2020) Elevated Thresholds for Light Touch in Children With Autism Reflect More Conservative Perceptual Decision-Making Rather Than a Sensory Deficit. *Frontiers in Human Neuroscience.* 0 10.3389/fnhum.2020.00122.

[54] Roman-Urrestarazu, A., van Kessel, R., Allison, C., Matthews, F. E., Brayne, C. & Baron-Cohen, S. (2021) Association of Race/Ethnicity and Social Disadvantage With Autism Prevalence in 7 Million School Children in England. *JAMA Pediatrics.* 175 (6), e210054. 10.1001/jamapediatrics.2021.0054.

[55] Roos, E. M., McDuffie, A. S., Weismer, S. E. & Ann Gernsbacher, M. (2008) A comparison of contexts for assessing joint attention in toddlers on the autism spectrum. *Autism : The International Journal of Research and Practice.* 12 (3), 275-291. 10.1177/1362361307089521.

[56] Scassellati, B., Boccanfuso, L., Huang, C., Mademtzi, M., Qin, M., Salomons, N., Ventola, P. & Shic, F. (2018) Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics.* 3 (21), 10.1126/scirobotics.aat7544.

[57] Schadenburg, B. R., Reidsma, D., Evers, V., Davison, D. P., Li, J. J., Heylen, D. K. J., Neves, C., Alvito, P., Shen, J., Pantić, M., Schuller, B. W., Cummins, N., Olaru, V., Sminchisescu, C., Dimitrijević, S. B., Petrović, S., Baranger, A., Williams, A., Alcorn, A. M. & Pellicano, E. (2021) Predictable Robots for Autistic Children — Variance in Robot Behaviour, Idiosyncrasies in Autistic Children's Characteristics, and Child–Robot Engagement. *ACM Transactions on Computer-Human Interaction.* 28 (5), 1-42. 10.1145/3468849.

[58] Shankar, S., Marshall, S. K. & Zumbo, B. D. (2020) A Systematic Review of Validation Practices for the Goal Attainment Scaling Measure. *Journal of Psychoeducational Assessment.* 38 (2), 236-255. 10.1177/0734282919840948.

[59] Silva, V., Soares, F., Esteves, J. S., Santos, C. P. & Pereira, A. P. (2021) Fostering emotion recognition in children with autism spectrum disorder. *Multimodal Technologies and Interaction.* 5 (10), 10.3390/mti5100057.

[60] So, W., Cheng, C., Lam, W., Wong, T., Law, W., Huang, Y., Ng, K., Tung, H. & Wong, W. (2019) Robot-based play-drama intervention may improve the narrative abilities of Chinese-speaking preschoolers with autism spectrum disorder. *Research in Developmental Disabilities.* 95 103515. 10.1016/j.ridd.2019.103515.

[61] So, W., Wong, M. K., Lam, C. K., Lam, W., Chui, A. T., Lee, T., Ng, H., Chan, C. & Fok, D. C. (2018a) Using a social robot to teach gestural recognition and production in children with autism spectrum disorders. *Disability and Rehabilitation. Assistive Technology.* 13 (6), 527-539. 10.1080/17483107.2017.1344886.

[62] So, W., Wong, M. K., Lam, W., Cheng, C., Yang, J., Huang, Y., Ng, P., Wong, W., Ho, C., Yeung, K. & Lee, C. (2018b) Robot-based intervention may reduce delay in the production of intransitive gestures in Chinese-speaking preschoolers with autism spectrum disorder. *Molecular Autism.* 9 34. 10.1186/s13229-018-0217-5.

[63] Taheri, A., Meghdari, A., Alemi, M. & Pouretemad, H. (2018) Human–Robot Interaction in Autism Treatment: A Case Study on Three Pairs of Autistic Children as Twins, Siblings, and Classmates. *International Journal of Social Robotics.* Online http://rdcu.be/wQuF. 10.1007/s12369-017-0433-8.

[64] Valadão, C., Goulart, C., Rivera, H., Caldeira, E., Filho, T., Frizera-Neto, A. & Carelli, R. (2016) Analysis of the use of a robot to improve social skills in children with autism spectrum disorder. 10.1590/2446-4740.01316.

[65] van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., van der Ven, S. & Leseman, P. (2019) Social Robots for Language Learning: A Review. *Review of Educational Research.* 89 (2), 259-295. 10.3102/0034654318821286.

[66] van den Berk-Smeekens, I., de Korte, Manon W. P., van Dongen-Boomsma, M., Oosterling, I. J., den Boer, J. C., Barakova, E. I., Lourens, T., Glennon, J. C., Staal, W. G. & Buitelaar, J. K. (2021) Pivotal Response Treatment with and without robot-assistance for children with autism: a randomized controlled trial. *European Child & Adolescent Psychiatry.* 10.1007/s00787-021-01804-8.

[67] van Otterdijk, M. T. H., de Korte, M. W. P., van den Berk-Smeekens, I., Hendrix, J., van Dongen-Boomsma, M., Den Boer, J. C., Buitelaar, J. K., Lourens, T., Glennon, J. C., Staal, W. G. & Barakova, E. I. (2020) The effects of long-term child–robot interaction on the attention and the engagement of children with autism. *Robotics.* 9 (4), 1-15. 10.3390/robotics9040079.

[68] Vivanti, G. & Hamilton, A. (2014) Imitation in Autism Spectrum Disorders. In: Volkmar, F. R., Paul, R., Rogers, S. J. & Pelphrey, K. A. (eds.). *Handbook of Autism and Pervasive Developmental Disorders, Vol 1: Diagnosis, Development, and Brain Mechanisms.* , Wiley. pp. 278-301.

[69] Wu, Z., Pan, J., Su, Y. & Gros-Louis, J. (2013) How joint attention relates to cooperation in 1- and 2-year-olds. *International Journal of Behavioral Development.* 37 (6), 542-548. 10.1177/0165025413505264.

[70] Yun, S., Choi, J., Park, S., Bong, G. & Yoo, H. (2017) Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system. *Autism Research.* 10 (7), 1306-1323. 10.1002/aur.1778.

[71] Zheng, Z., Nie, G., Swanson, A., Weitlauf, A., Warren, Z. & Sarkar, N. (2020) A Randomized Controlled Trial of an Intelligent Robotic Response to Joint Attention Intervention System. *Journal of Autism and Developmental Disorders.* 50 (8), 2819-2831. 10.1007/s10803-020-04388-5.

[72] Zheng, Z., Young, E. M., Swanson, A. R., Weitlauf, A. S., Warren, Z. E. & Sarkar, N. (2016) Robot-Mediated Imitation Skill Training for Children With Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* 24 (6), 682-691. 10.1109/TNSRE.2015.2475724.

[73] Zorcec, T., Robins, B. & Dautenhahn, K. (2018) Getting Engaged: Assisted Play with a Humanoid Robot Kaspar for Children with Severe Autism. *10th International ICT Innovations Conference, ICT Innovations 2018, September 17, 2018 - September 19, 2018.* Ohrid, Macedonia, Springer Verlag. pp.198-207.