

Gaze contingent joint attention with an avatar in children with and without ASD

G.E. Little¹, L. Bonnar¹, S.W. Kelly¹, K. S. Lohan² and G. Rajendran²

¹School of Psychological Sciences & Health, University of Strathclyde, Glasgow

²Heriot-Watt University, Edinburgh

Abstract—Problems with joint attention (JA) are core features of Autism Spectrum Disorders (ASD). Here, we investigate how typically developing (TD) children and children with ASD respond to joint attention (RJA) and initiate joint attention (IJA) with a gaze contingent avatar. Thirty-one participants with ASD and 33 TD matched controls followed and directed the avatar’s gaze to a series of referent images. Viewing times and recognition memory for the referent images were measured and compared between RJA and IJA conditions. Analysis of correctly identified target images suggests comparable target recognition between IJA and RJA conditions for both groups, but poorer overall recognition memory in the ASD group. Eye tracking data suggests different viewing strategies between the ASD and TD groups. These findings demonstrate the importance of considering processing time and saliency of referent objects when creating interactive social technology for children with ASD and further highlights the potential of interactive, gaze contingent social characters for enhancing our knowledge of joint attention.

Keywords– Autism Spectrum Disorders, ASD, Joint Attention, Gaze contingency, Gaze

I. INTRODUCTION

At its most basic level, joint attention (JA) is where two people coordinate their attention to the same object or event. This highly social process is achieved via a combination of social cues including: eye gaze, vocalisation, pointing and gesture. JA behaviours emerge early in infancy and develop throughout childhood to facilitate social learning and information processing [1]. This places JA as a vital building block for the development of language, theory of mind and social cognition [2]–[4]. Recent research suggests two distinct but interacting aspects of JA: responding to joint attention (RJA) and initiating joint attention (IJA). RJA refers to the ability to follow another’s gaze, gesture or head turn to a referent object. IJA refers to the ability to spontaneously indicate a referent object to another person. RJA and IJA activate distinct and overlapping visual attention and social cognitive mechanisms associated with interpreting communicative intentions, reward and motivation [5]–[7]. Reward-related brain areas [6] and goal-directed attention mechanisms [7] have been shown as being more actively engaged during IJA, highlighting the inherent disposition for humans to share social experience. The eyes are important cues for social interaction and the direction of someone’s gaze is a strong indication of their focus of attention. Subsequently, much of the experimental work on JA has been in the context of responding to eye gaze. A vast body of research has adapted the Posner style cueing paradigm to measure the effects of gaze cueing on visual attention. This has demonstrated that the presentation of

gaze correctly cueing a target elicits faster orienting, location and discrimination of the target [8]. As well as modulating spatial attention, observing a gaze shift can affect evaluative judgments about the face producing the gaze shift and the referent object itself. For example, faces gazing to cue targets correctly are judged to be more trustworthy and objects under JA are rated more favourably than those not viewed under JA or indicated by non-social, arrow cues [9], [10]. Viewing objects with another person under JA, compared to the absence of JA, has also been shown to enhance object processing in 9-month-old infants [11]. This is demonstrated by reduced subsequent viewing time to the object viewed under JA than no JA when both are viewed together, indicative of increased recognition and familiarity for objects viewed under JA.

Until recently, the focus of these experimental studies has been on how individuals respond to gaze shifts in paradigms eliciting only RJA. This focus on the ‘responder’ has resulted in a comparative dearth of knowledge about how individuals initiate joint attention, the effects IJA may have on cognitive processing and any differences between RJA and IJA. However, modern advances in eye tracking and virtual reality technology have heralded more interactive experimental paradigms, which are better able to capture the reciprocity of JA interactions (e.g. [5]–[7], [12], [13]). In a recent, novel virtual reality paradigm, Kim and Mundy [12] developed a virtual avatar that could emulate both RJA and IJA. Clinically normal adults either directed or followed the avatar’s gaze to a series of target images of faces, houses and abstract patterns – and subsequently their memory for the target images was tested. Their findings suggested enhanced recognition memory for house and abstract images in the IJA condition, i.e. when participants’ gaze was followed to the image by the avatar. This suggests that having a gaze shift reciprocated in IJA enhanced depth of encoding relative to RJA. This pioneering study highlights the importance of JA for information processing and supports the idea that RJA and IJA may involve distinct yet overlapping mechanisms. Further investigation of any differences between RJA and IJA are therefore likely to improve our understanding of JA and social information processing in typical development – and may also be illuminating for understanding these processes in atypical development such as in autism spectrum disorders (ASD). Individuals with ASD are known to experience significant deficits in JA, which may have a negative effect on social cognition and learning throughout the lifespan. So, in this reported study we recruited individuals with ASD to provide a unique window into RJA and IJA processes.

ASD is a pervasive neurodevelopmental disorder, characterised by difficulties with social interaction and communication and restricted and repetitive behaviours. Reduced eye contact and impairments in JA are some of the earliest diagnostic indicators of ASD [14], discriminating infants with ASD from their peers with typical development and intellectual developmental disorders [15]. Gaze cueing studies suggest that the ability to follow gaze is intact in individuals with ASD [16], however, evidence that individuals with ASD distinguish less between gaze and non social cues than their TD peers suggests that this skill could be achieved through a non-social mechanism (e.g. [17]). Furthermore, evidence from eye-tracking studies also suggests that individuals with ASD spend less time looking at objects gazed at by others in social scenes (e.g. [18]–[20]). This may reflect difficulties in understanding of the socially mediated referential nature of gaze in ASD. It has also been argued that deficits in IJA may be more severe and longer lasting than deficits in RJA. However, less is known about gaze behaviour in IJA in individuals with ASD or any differences in information processing between IJA and RJA.

Psychologists are increasingly using interactive virtual environments and socially assistive robots to understand and improve communication in ASD [21]. Children with ASD show enhanced engagement with novel virtual and robotic social partners, illustrating the potential value of these tools for intervention and therapy [21]. Recent studies have shown increased spontaneous JA from children with ASD when engaging with social virtual characters and robots (e.g. [22], [23]). A better knowledge of how children with ASD engage in JA with both human and virtual agents will be informative for future development of technological interventions using social environments and socially assistive robots for children with ASD. Following the JA paradigm by Kim and Mundy [12], we investigated how children with and without ASD initiate and respond to JA with a gaze contingent avatar. The avatar made a head turn and gaze shift towards a series of referent images under RJA and IJA conditions. Similarly to Kim and Mundy’s [12] study with adults, children’s recognition memory for the target images was recorded. Recognition memory and viewing times were compared to investigate gaze behaviour and social cognition between the groups and RJA and IJA conditions – this allows us to look at the process as well as the outcome (see [24] for discussion about process vs. outcome). Here we expected that the TD children would demonstrate significantly better recognition memory for target images in the IJA condition than the RJA condition. This effect was not expected in the ASD group or was expected to be considerably weaker.

II. METHOD

A. Participants

Following permission from relevant education authorities, participants were recruited from schools with autism units, schools for children with Special Educational Needs and mainstream primary schools in central Scotland. Children with ASD were identified by head teachers and recruited on the basis of their diagnosis, which was further confirmed with an ADOS [25] classification. Thirty-one children with ASD and 33 typically developing (TD) children (four female participants

per group) participated in the study. One female participant from the ASD group did not meet the ADOS classification for ASD, however, due to their community diagnosis of ASD and their score on the Social Communication Questionnaire (SCQ) in the high range [26], they were still included in the study. Parents/guardians gave written consent for their child’s participation and children gave verbal consent at testing. The study protocol received ethical approval from the University of Strathclyde.

Two participants from the ASD group did not complete the task, one due to technical problems with the hardware and the other due to limited task engagement. Only participants who completed the task at above chance level performance for both RJA and IJA conditions were included in the reported analysis. This led to the removal of a further 5 participants from the ASD group. The 24 remaining participants with ASD were matched to a participant from the TD group on the basis of age and IQ. IQ was measured by overall performance on four subscales (FSIQ-4) of the Weschler Abbreviated Scale of Intelligence (WASI-II) [27]. Table I shows the characteristics of the 48 participants included in the final analysis.

TABLE I: Participant characteristics

	ASD (N= 24)	TD (N=24)	Total (N=48)
Age (months)	128.63 (21.23)	127.33 (14.54)	127.98 (18.01)
FSIQ-4	90.54 (12.21)	91.04 (10.82)	90.79 (11.41)
SCQ	25.12(6.83) ^{***a}	6.10(4.34) ^{***b}	14.84(11.09) ^c

^{***} $p < .001$

^a 17 parents of participants with ASD completed the SCQ

^b 20 parents of TD participants completed the SCQ

^c Total SCQ returns, N=37

B. Hardware and software

The experiment was created and run with E-prime extensions for Tobii (E-prime) and displayed on a Dell latitude E6510 laptop with a 15.5” screen, 1366/768 resolution and a 60Hz refresh rate. Eye movements were tracked using a Tobii X260 eye tracker (Tobii, Stockholm, Sweden) which sat below the laptop screen, recording eye movements binocularly at 60Hz with a spatial resolution of 0.4. Participants used a chin-rest, positioned 62 cm from the screen.

C. Stimuli

The avatar was rendered using ‘Blender’ 2.6 (blender.org). The character, ‘‘Danny’’ was male, childlike and cartoon-like to avoid any uncanny valley effects (see [28] about how uncanny valley may be different for people with ASD), portraying a neutral facial expression. The avatar measured 6.9° by 8.7° and was centrally presented on a black background. The avatar was flanked by two placeholders measuring 5.6° by 6.9°, displayed 7.8° from the centre of the screen to the left and right. The avatar was animated to make a head turn and gaze towards the placeholders. This animation was presented at 25 fps.

Colour images of 48 faces, houses and abstract patterns were presented for the memory task. Abstract and house images from Kim and Mundy’s [12] study were used here alongside supplementary face, abstract and house images. Twenty-two house and 15 abstract images were from Kim and Mundy’s [12] original stimuli set. A further 24 house images and 33 images of abstract designs were sourced from the Internet.

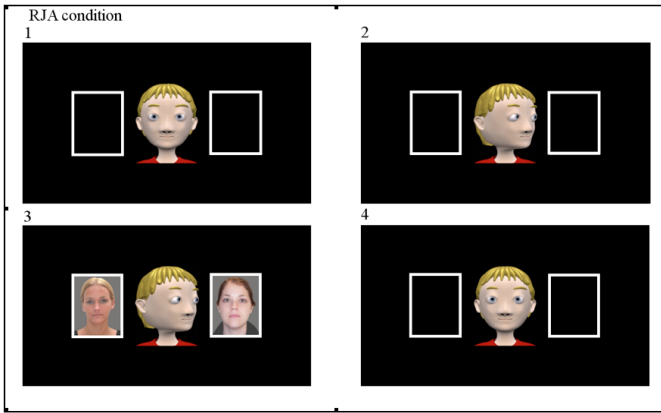


Fig. 1: Presentation of an RJA trial. 1. Initiation look, the avatar gazes forward for 300ms. 2. Reference look- the avatar turns to look at placeholder, head turn duration: 640ms. 3. When participant makes fixation (≥ 200 ms) on the correct placeholder images appear for 1000ms, when images disappear the avatar remains fixated on empty placeholder for 400ms. 4. Sharing look, the avatar returns his gaze to the participant.

Face images were taken from normalised versions of the CAL/PAL faces database [29]. These comprised 24 male and 24 female faces with neutral expressions shown on a grey background.

1) *Joint attention task:* RJA and IJA conditions differed in how participants chose, or were guided to the target images. Both conditions comprised 2 learning phases, each followed by test phases.

2) *RJA Learning phase:* Fig. 1 shows the time course and presentation of an RJA trial. Participants were instructed to look at "Danny", follow his gaze to one of the placeholders and remember the image that appeared in his chosen placeholder. The avatar's gaze direction was randomised, with an equal number of looks to the left and right. Images the participant was directed to view were target images. Images in the opposite placeholder were non-target images; these were presented but not designated for viewing and memorising.

3) *IJA learning phase:* Fig. 2 shows the time course and presentation of an IJA trial. Participants were instructed to look at "Danny" and then choose one placeholder to look at. Eye tracker feedback allowed the avatar to follow participants' gaze to the same placeholder. Images subsequently appeared in both placeholders. Here, the target image was the image in the placeholder the participant chose to look at and the non-target image was the image presented in the opposite placeholder. Participants completed two blocks of 12 learning trials for each joint attention (JA) condition. Presentation order for RJA and IJA blocks was counterbalanced across participants. Each block comprised 4 trials of face, house and abstract stimuli. Presentation of images as targets, non-targets, and novel images was also counterbalanced across participants with the image presentation being randomised. Target and non-target images presented in a single trial were always unique and from the same image category. Participants were asked to study the target images and remember as many as they could for later. They were asked to look back at the avatar to complete each trial. Each block of trials was followed by a test phase.

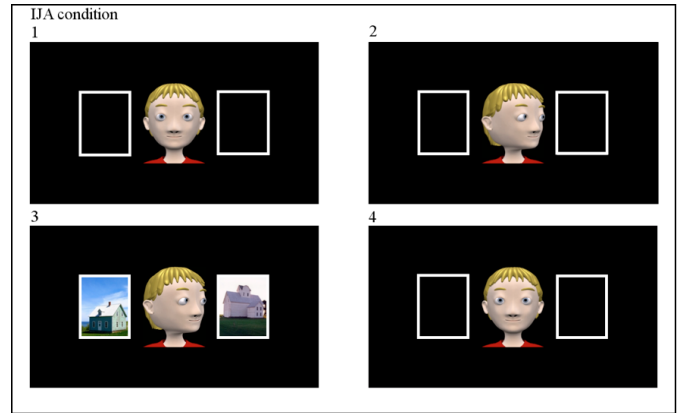


Fig. 2: Presentation of an IJA trial. 1. Initiation look, the avatar gazes forward waiting for participant's reference look. 2. Participant chooses placeholder by making reference look to one placeholder. After participant fixation of ≥ 200 ms on one placeholder, the avatar makes head turn to the same placeholder. Duration of head turn: 640ms 3. Images appear for 1000ms, when images disappear the avatar remains fixated on empty placeholder for 400ms. 4. Sharing look, the avatar returns his gaze to the participant.

4) *Test phase:* The test phase was completed after each learning phase. This comprised random presentation of 24 familiar images (12 target and 12 non-target images from the previous learning phase) and 12 novel images not previously presented. These were sequentially presented and participants indicated with a button press whether they recognised the image or not. The response buttons were clearly marked with large stickers showing a 'Y' for remember and 'N' for do not remember.

D. Procedure

A standard calibration procedure was repeated until a good calibration was achieved and the task could begin. Before the first RJA and IJA blocks 4 'learning phase' practice trials were presented, showing images of common fruit and vegetables. These could be repeated as many times as necessary. Participants then completed the learning and test phases for each block. A short break was offered between blocks until participants were ready to continue with the task.

III. RESULTS

For brevity, analyses comparing the image types (abstract, face and house images) are not reported here, and these are collapsed into overall RJA and IJA conditions. The variable of interest for target image memory was d' . d' is a discriminability index where participants correct target identifications (hits) and false positive identifications (false alarms) were calculated using the loglinear approach [30] and transformed according to signal detection theory for a measure of response sensitivity. To calculate hit rates, the total number of hits (+ 0.5) was divided by the total number of target images (+ 1). False alarm rates were calculated by dividing the total number of false alarms (+ 0.5) by the total number of foils (+ 1). Z-scores of the false alarm rates were then subtracted from the z-scores of the hit rates to find the d' value [30]. Table II shows the d' for each group for the RJA and IJA conditions.

TABLE II: Group d' target recognition for RJA and IJA condition.

Group	RJA	IJA
ASD (N=24)	1.01 (0.57)	0.92 (0.56)
TD (N=24)	1.83 (0.77)	1.74 (0.87)
Total (N=48)	1.42 (0.79)	1.33 (0.84)

1) *Effect of JA condition:* ¹A 2 (JA condition, RJA/IJA) x 2 (group, ASD/TD) mixed ANOVA on the d' scores, with 'group' as the between groups factor revealed no significant main effect of JA condition ($F(1,46) = 1.48, p = .229$), suggesting no significant differences in d' between RJA and IJA conditions. There was a significant main effect of group $F(1,46) = 18.63, p < .001, \eta^2 = .288$ with the ASD group demonstrating significantly lower d' scores than the TD group for both the RJA and the IJA conditions. No interaction between JA condition and group was found $F(1,46) = .000, p = .993$.

2) *Gaze data analysis:* The gaze data was firstly visually inspected for anomalies. This led to the removal of the entire data set for one ASD participant. This was due to eye tracker error leading to missing trials. Gaze data was analysed for the period that the target image was presented on the screen. An 'on-screen' percentage was calculated for each participant. This was the percentage of time participants' point of gaze was detected anywhere on the screen during target presentation. Inspection of the eye tracking data led to the removal of two outliers from the ASD group, one for reduced on-screen percentage ($< 75\%$) and the other for high screen viewing time with limited looking at the task AOIs. The data was square-root transformed to eliminate negative skew. Independent samples t -tests suggested significant differences in on-screen percentage between the ASD and TD groups for the RJA condition ($t(20.29) = 5.90, p < .001, d = 2.62$), ASD mean = 93.63% (SD = 5.82), TD mean = 99.68% (SD = 0.23) and the IJA condition, ($t(20.08) = 3.24, p = .004, d = 1.45$), ASD mean = 93.23% (SD = 6.83), TD mean = 98.21% (SD = 0.19). Levene's test indicated violation of the assumption of equality of variance therefore degrees of freedom were adjusted accordingly. This analysis suggests that the ASD group spent more time looking off-screen, or the eye tracker detected fewer gaze samples for this group. Individual group analysis using paired samples t -tests suggested no significant differences for on-screen time between the RJA and IJA conditions for the ASD group ($t(20) = -.066, p = .948$), and a small but statistically significant difference for the TD group with less on-screen time in the IJA condition ($t(23) = -19.65, p < .001, d = 4.01$).

The areas of interest (AOIs) for the gaze data analysis were the target and the non-target image placeholders and the avatar, these AOIs included a margin of 1° around their borders. Total viewing times to these AOIs and elsewhere on the screen were calculated for the show target period. Viewing time was defined as the total amount of time the eye tracker detected the

¹Initial screening of the data revealed skew in some experimental conditions that was not eliminated by transformations. Precautionary nonparametric tests (which do not assume a normal distribution) were conducted and the results were consistent with those of parametric analyses. Parametric tests are considered to be robust to violations of normal distribution [31] so parametric analyses are reported here to allow interaction effects to be investigated.

participant's point of gaze within the respective AOIs. Table III shows the means and standard deviations of total AOI viewing times for each group and JA condition. To analyse viewing

TABLE III: Total AOI viewing times

JA	AOI	ASD (N=21)	TD (N=24)	Tot. (N=45)
RJA	Target	16032 (3526)	20040 (2799)	18169 (3735)
	Non-target	2743 (1968)	1413 (1414)	2034 (1804)
	the avatar	1935 (1190)	500 (465)	1170 (1131)
	Elsewhere	590 (788)	379 (51)	477 (542)
IJA	Target	11330 (3396)	15427 (3750)	13515 (4107)
	Non-target	4986 (2612)	3494 (2779)	4190 (2776)
	the avatar	4201 (2487)	2576 (1428)	3334 (2133)
	Elsewhere	592 (903)	444 (59)	513 (615)

times. Three separate 2 (Group, ASD/TD) x 2 (JA condition, RJA/IJA) mixed ANOVAs were conducted, one for each AOI. *Target image AOI-* Analysis of viewing times to target images suggested a significant main effect of JA condition ($F(1,43) = 114.89, p < .001, \eta^2 = .728$) with longer target viewing times in the RJA condition than the IJA condition. There was also a significant main effect of group ($F(1, 43) = 19.62, p < .001, \eta^2 = .313$) suggesting significantly longer target viewing times for the TD group than the ASD group. No significant interaction between group and JA condition was found ($F(1, 43) = .011, p = .918$).

Non-target image AOI- The gaze data for non-target image viewing times was square root transformed to eliminate slight positive skew. Analysis of the transformed data suggested a significant main effect of JA condition ($F(1,43) = 52.89, p < .001, \eta^2 = .552$) with longer non-target viewing times in the IJA condition than the RJA condition. There was also a significant main effect of group ($F(1,43) = 7.41, p = .009, \eta^2 = .147$) suggesting significantly longer non-target viewing times for the ASD group than the TD group. No significant interaction between group and JA condition was found ($F(1,43) = .177, p = .676$).

The avatar AOI- The gaze data for the avatar AOI was also square root transformed to eliminate positive skew. Analysis of the transformed viewing time data suggested a significant main effect of JA condition ($F(1,43) = 126.58, p < .001, \eta^2 = .746$) with longer viewing times to the avatar in the IJA condition than the RJA condition. There was also a significant main effect of group ($F(1,43) = 22.23, p < .001, \eta^2 = .341$) suggesting significantly longer avatar viewing times for the ASD group than the TD group. An interaction between group and JA condition approached but did not reach significance ($F(1,43) = 3.83, p = .057, \eta^2 = .082$).

Elsewhere AOI- The data for viewing time elsewhere on the screen violated assumptions of homogeneity of variance and was not normally distributed. This data was not transformed as the distribution was differently shaped for the ASD and TD groups. Non-parametric Wilcoxon signed-rank tests were therefore conducted for both groups separately. This revealed no significant difference in viewing times between the RJA (Mdn = 251) and IJA (Mdn = 214) conditions ($Z = -.109, p = .913$) for the ASD group. Analysis for the TD group suggested significantly less looking elsewhere on the screen for the RJA condition (Mdn = 378) than the IJA (Mdn = 460) condition ($Z = -3.73, p < .001, r = 0.76$).

Mann-Whitney tests (also non-parametric) were conducted to

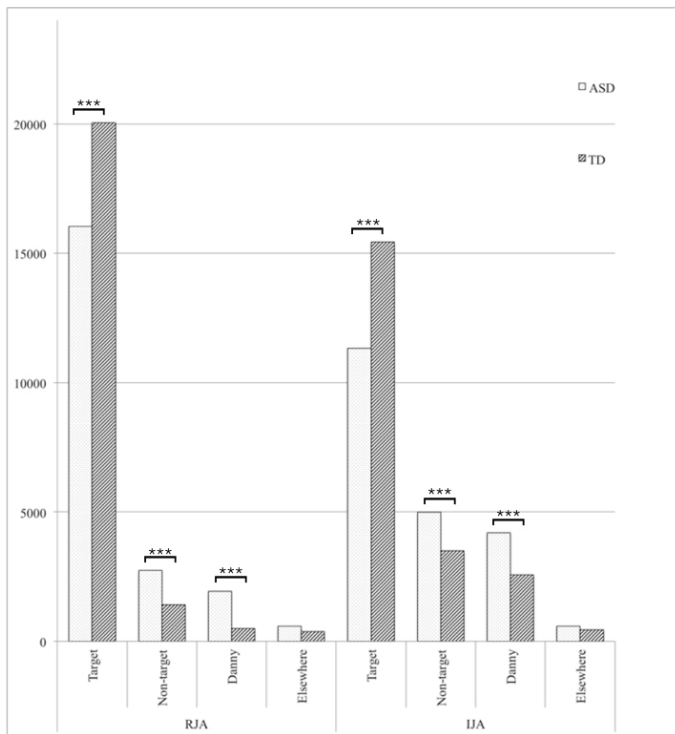


Fig. 3: Total ASD and TD viewing times for each AOI, *** $p < .001$.

investigate the differences between the groups for each JA condition separately. This revealed no significant difference between the ASD and TD groups in the RJA condition ($U = 204.00$, $Z = -1.09$, $p = .275$) or the IJA condition ($U = 194.00$, $Z = -1.32$, $p = .186$) suggesting similar overall viewing time of on screen areas irrelevant to the task.

Fig. 3 shows the groups viewing times for each area on the screen for the RJA and IJA conditions.

IV. DISCUSSION

Our study used a gaze-contingent avatar to investigate JA interactions with a group of children with ASD and their IQ and age matched TD peers. Here, participants followed the avatar's gaze to target images to emulate RJA and guided the avatars gaze to their chosen targets to emulate IJA. While sharing attention to the images with the avatar, participants were asked to memorise them for a subsequent memory test. Depth of encoding and recognition, as measured by d' , was recorded. Performance was compared between JA conditions and participant groups to uncover any differences in how the referent image is encoded and processed during RJA and IJA for ASD and TD children. To investigate both groups viewing strategies – and, therefore, investigate outcome versus process – total viewing times to the target image, non-target image, avatar and elsewhere on the screen were also calculated and compared between conditions.

Analysis of d' scores showed that for both ASD and TD groups, memory for target images did not significantly differ between the RJA and IJA conditions. This suggests similar levels of encoding for the referent image in each condition. This is contrary to findings with typically developing adults using a similar paradigm [12]. Compared to following an avatars gaze, Kim and Mundy [12] found that when adults' gaze shifts towards target images were reciprocated by corresponding gaze shifts from the avatar, recognition memory for images of houses and patterns was enhanced. In the current study, this effect was neither found in TD children, nor children with ASD.

The investigation of viewing times to the target images is

useful in trying to understand this finding. Despite a lack of significant differences in recognition memory between JA conditions, both groups showed significantly less total viewing times for target images in the IJA condition than the RJA condition. This suggests different visual strategies for both tasks. It appears here that the children may not have had to view the referent image for as long during IJA to achieve the same level of encoding as they did during RJA. This could reflect different thresholds of time required to process the referent in RJA and IJA conditions, where less time may be required in the IJA condition. Care was taken here to control viewing times to the target image. In both RJA and IJA conditions, the target only appeared when both the avatar and participant were fixated on the relevant placeholder. This however, did not guarantee that gaze was focused on the target image for the whole target presentation time. Future experiments using this paradigm, in which we can vary available target viewing time, will be useful in disentangling this finding. For their adult group, Kim and Mundy [12] found no differences in target viewing times between the two JA conditions. Their comparable target viewing times may reflect an even more task-focused approach from adult participants and less distraction from non-target images and the avatar. It must also be noted that they used a more immersive virtual reality paradigm.

Overall, the gaze contingency element of the task worked well. In the IJA condition the avatar reliably looked at the participants' chosen placeholder. However, it must be noted that there was more variability in the time taken for the image stimuli to be presented here. This may have occurred due to eye tracker flicker. However, it should be noted that if the eye tracker completely lost the participant's point of gaze, the task was automatically terminated. In some instances, this variability may also have been due to the self-paced nature of the task. Creative methodology to more strictly control the self-paced nature of IJA should be implemented in future research. Interestingly, participants from both groups spent more looking time at the avatar during target presentation in the IJA condition than the RJA condition. This may reflect the nature of the IJA process, which is more communicative and declarative than RJA. Here, the participant had to decide which placeholder to view and allow the avatar to follow their gaze. It would be logical here for the initiator to monitor whether they have captured the attention of their social partner. Indeed, instances of IJA often include "check back" looks for this purpose [13].

Overall, the ASD group performed more poorly at the memory task than the TD group. This was reflected in different viewing times between the groups, which highlight different allocation of attention to the referent image and the other AOIs. During the show target period, the ASD group demonstrated significantly reduced viewing time to the target image compared with the TD group and significantly more viewing time to the non-target image and the avatar. This relative lack of attention to the target images and increase of attention to areas of the screen irrelevant for the task may have allowed for less processing time for the target image in the ASD group. This could be indicative of a generally less focused strategy for the task. Reduced looking times to the referent image from the ASD participants in the RJA condition concurs with previous findings showing that individuals with ASD look less at a gazed at object than their TD peers ([18]–[20]).

Overall, these findings are valuable when considering how to develop interventions for triadic joint attention with children with autism through virtual or robotic socially assistive mediums. They suggest ways in which these interactions could be modulated to allow for more effective communication and learning. Because the children with ASD spent less time looking at the referent object their encoding seemed to suffer as a consequence. This suggests that platforms which allow more

time for children to engage in JA interactions and increase the saliency of social cues and the referent object may improve interactive computer-based paradigms like this one. Indeed, this mirrors guidance for teaching joint attention to children with ASD in face-to-face interventions [32]. Increasing the saliency of the referent has also been demonstrated to be effective in computer-based word learning tasks with children with ASD. For example, making a referent object bounce slightly when a schematic, dynamic face gazed towards it increased both attentional focus and subsequent learning in a gaze contingent novel object learning paradigm [33]. Investigating ASD in this context provides a unique insight into joint attention and the challenges for its study. Future gaze contingent paradigms could be improved by providing additional cues than a gaze shift and head turn to the referent object. For example, pointing and vocalising or touching the referent in both IJA and RJA interactions may further increase attention to the referent and enhance subsequent processing. Furthermore, including more interactive interfaces such as touch screens and sound cues may also enhance the interactive experience. Further studies should investigate the additive effects of all of these cues in this type of easily controllable virtual environment. Improving understanding of when to follow participant gaze shifts, repeat gaze shifts, and how long to give children to respond to gaze, will help explore optimal levels of JA interaction for children in general. It should also be noted that the efficacy of different cues may vary between children with ASD and the potential for more tailor-made systems which can incorporate an individual's personal object of interest, or individual looking strategies, may be helpful for future intervention. This will be useful particularly in research with children with ASD who demonstrate more attention and engagement difficulties than TD. So, we argue this area of research has great future potential in both basic and applied research.

ACKNOWLEDGMENT

Thank you to all of the children, parents and schools who participated in this study. Thank you to Bill Woodside and Dr. Bruce McGregor for their technical help and advice with the experimental program and animation for the study.

REFERENCES

- [1] P. Mundy, L. Sullivan, and A. M. Mastergeorge, "A parallel and distributed-processing model of joint attention, social cognition and autism," *Autism research*, vol. 2, no. 1, pp. 2–21, 2009.
- [2] P. Mundy and L. Newell, "Attention, joint attention, and social cognition," *Current directions in psychological science*, vol. 16, no. 5, pp. 269–274, 2007.
- [3] T. Charman, "Why is joint attention a pivotal skill in autism?" *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 358, no. 1430, pp. 315–324, 2003.
- [4] M. Tomasello, "Joint attention as social cognition," *Joint attention: Its origins and role in development*, pp. 103–130, 1995.
- [5] U. J. Pfeiffer, L. Schilbach, B. Timmermans, B. Kuzmanovic, A. L. Georgescu, G. Bente, and K. Vogeley, "Why we interact: on the functional role of the striatum in the subjective experience of social interaction," *Neuroimage*, vol. 101, pp. 124–137, 2014.
- [6] L. Schilbach, M. Wilms, S. B. Eickhoff, S. Romanzetti, R. Tepest, G. Bente, N. J. Shah, G. R. Fink, and K. Vogeley, "Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry," *Journal of Cognitive Neuroscience*, vol. 22, no. 12, pp. 2702–2715, 2010.
- [7] E. Redcay, M. Kleiner, and R. Saxe, "Look at this: the neural correlates of initiating and responding to bids for joint attention," *Towards a neuroscience of social interaction*, p. 491, 2012.
- [8] A. Frisken, A. P. Bayliss, and S. P. Tipper, "Gaze cueing of attention: visual attention, social cognition, and individual differences," *Psychological Bulletin*, vol. 133, no. 4, p. 694, 2007.
- [9] A. P. Bayliss, M. A. Paul, P. R. Cannon, and S. P. Tipper, "Gaze cuing and affective judgments of objects: I like what you look at," *Psychonomic bulletin & review*, vol. 13, no. 6, pp. 1061–1066, 2006.
- [10] A. P. Bayliss and S. P. Tipper, "Predictive gaze cues and personality judgments should eye trust you?" *Psychological Science*, vol. 17, no. 6, pp. 514–520, 2006.
- [11] T. Striano, X. Chen, A. Cleveland, and S. Bradshaw, "Joint attention social cues influence infant learning," *European Journal of Developmental Psychology*, vol. 3, no. 3, pp. 289–299, 2006.
- [12] K. Kim and P. Mundy, "Joint attention, social-cognition, and recognition memory in adults," *Frontiers in Human Neuroscience*, vol. 6, 2012.
- [13] A. P. Bayliss, E. Murphy, C. K. Naughtin, A. Kritikos, L. Schilbach, and S. I. Becker, "gaze leading: Initiating simulated joint attention influences eye movements and choice behavior," *Journal of Experimental Psychology: General*, vol. 142, no. 1, p. 76, 2013.
- [14] J. Osterling and G. Dawson, "Early recognition of children with autism: A study of first birthday home videotapes," *Journal of Autism and Developmental Disorders*, vol. 24, no. 3, pp. 247–257, 1994.
- [15] J. A. Osterling, G. Dawson, and J. A. Munson, "Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation," *Development and Psychopathology*, vol. 14, no. 02, pp. 239–251, 2002.
- [16] K. Nation and S. Penny, "Sensitivity to eye gaze in autism: is it normal? is it automatic? is it social?" *Development and Psychopathology*, vol. 20, no. 01, pp. 79–97, 2008.
- [17] P. H. Vlamings, J. E. Stauder, I. A. van Son, and L. Mottron, "Atypical visual orienting to gaze-and arrow-cues in adults with high functioning autism," *Journal of autism and developmental disorders*, vol. 35, no. 3, pp. 267–277, 2005.
- [18] S. Fletcher-Watson, S. R. Leekam, V. Benson, M. Frank, and J. Findlay, "Eye-movements reveal attention to social information in autism spectrum disorder," *Neuropsychologia*, vol. 47, no. 1, pp. 248–257, 2009.
- [19] M. Freeth, P. Chapman, D. Ropar, and P. Mitchell, "Do gaze cues in complex scenes capture and direct the attention of high functioning adolescents with asd? evidence from eye-tracking," *Journal of autism and developmental disorders*, vol. 40, no. 5, pp. 534–547, 2010.
- [20] D. M. Riby, P. J. Hancock, N. Jones, and M. Hanley, "Spontaneous and cued gaze-following in autism and williams syndrome," *Journal of neurodevelopmental disorders*, vol. 5, no. 1, p. 1, 2013.
- [21] B. Scassellati, H. Admoni, and M. Mataric, "Robots for use in autism research," *Annual review of biomedical engineering*, vol. 14, pp. 275–294, 2012.
- [22] K. Porayska-Pomsta, C. Frauenberger, H. Pain, G. Rajendran, T. Smith, R. Menzies, M. E. Foster, A. Alcorn, S. Wass, S. Bernadini *et al.*, "Developing technology for autism: an interdisciplinary approach," *Personal and Ubiquitous Comput.*, vol. 16, no. 2, pp. 117–127, 2012.
- [23] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A. Mirza, and M. Blow, "Kaspar—a minimally expressive humanoid robot for human–robot interaction research," *Applied Bionics and Biomechanics*, vol. 6, no. 3–4, pp. 369–397, 2009.
- [24] P. E. McKenna, A. Glass, G. Rajendran, and M. Corley, "Strange words: Autistic traits and the processing of non-literal language," *Journal of autism and developmental disorders*, vol. 45, no. 11, pp. 3606–3612, 2015.
- [25] M. Rutter, P. DiLavore, S. Risi, K. Gotham, and S. Bishop, "Autism diagnostic observation schedule: Ados-2," *Torrance, CA: Western Psychological Services*, 2012.
- [26] T. Charman, G. Baird, E. Simonoff, S. Chandler, A. Davison-Jenkins, A. Sharma, T. O'Sullivan, and A. Pickles, "Testing two screening instruments for autism spectrum disorder in uk community child health services," *Developmental Medicine & Child Neurology*, 2015.
- [27] D. Wechsler and C. Hsiao-pin, *WASI-II: Wechsler abbreviated scale of intelligence*. Pearson, 2011.
- [28] Y. Ueyama, "A bayesian model of the uncanny valley effect for explaining the effects of therapeutic robots in autism spectrum disorder," *PLoS one*, vol. 10, no. 9, p. e0138642, 2015.
- [29] N. C. Ebner, "Age of face matters: Age-group differences in ratings of young and old faces," *Behaviour Research Methods*, vol. 40(1), no. 1, pp. 130–136, 2008.
- [30] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behav. Res. Methods Instrum. Comput.*, vol. 31, no. 1, pp. 137–149, 1999.
- [31] D. Clark-Carter, *Doing quantitative psychological research: From design to report*. Psychology Press/Erlbaum (UK) Taylor & Francis, pp. 205, 1997.
- [32] E. A. Jones and E. G. Carr, "Joint attention in children with autism theory and intervention," *Focus on autism and other developmental disabilities*, vol. 19(1), pp. 13–26, 2004.
- [33] H. Akechi, A. Senju, Y. Kikuchi, Y. Tojo, H. Osanai, and T. Hasegawa, "Do children with asd use referential gaze to learn the name of an object? an eye-tracking study," *Research in Autism Spectrum Disorders*, vol. 5, no. 3, pp. 1230–1242, 2011.