



Multiple deictic cues allow ASD children to direct their visual attention

Federica Cilia¹ · Julie Brisson² · Luc Vandromme³ · Cécile Garry² · Barbara Le Driant¹

Accepted: 29 October 2022 / Published online: 19 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The aim of this study was to investigate the impact of the deictic cue used to initiate joint visual attention behaviour. This eye-tracking study involves children with Autism Spectrum Disorder (ASD) whose developmental level, specifically in terms of communication skills, was below their chronological age. They were therefore matched with two groups of typical children, in chronological and developmental age. The video stimulus depicted an actress who used different deictic cues, in combination or not, to attract children's visual attention, i.e., glances alone; glances and verbalisations; or glances together with pointing and verbalisations. The data were analysed using a post hoc visual area of interest methodology. In the various conditions, ASD children paid less attention to the actress' face, but more to the background, compared to typical children. Visual exploration differs, which may explain why ASD children follow deictic cues less than the control groups. The combination of gaze cues and verbalisations was the least relevant for ASD children followed by deictic gaze cues with eye and/or head orientation. In contrast, adding pointing to gaze direction and verbalisations resulted in more responses to the initiation of joint attention. Our results suggest the importance of using multiple cues in combination to help ASD children acquire this language prerequisite and to better understand directionality in interaction. These results provide new ideas to adjust early remediation programs for ASD children.

Keywords Autism Spectrum Disorder · Deictic behaviors · Eye-tracking · Joint attention · Respons to Joint Attention · Target Perception

Introduction

Joint attention requires the child to pay attention to an object and at the same time the adult coordinates his or her own attention with that of the child. Thus, during joint attention one is aware of the presence of the other and the fact that s/he is paying attention to the same object at the same time. Social and perceptual aspects are involved in the development of joint attention and specifically with regard to gaze tracking (Astor et al., 2021). Indeed, the other communicates

with us non-verbally (smiles or other facial expressions, looks, gestures) or verbally (vocalisations or verbalisations). Joint attention is studied in children with Autism Spectrum Disorder (ASD) because this neurodevelopmental disorder is characterised in part by an impairment in communication and social interaction (American Psychiatric Association, 2013). Joint attention can be studied by naturalistic methodologies, where the initiation and response behaviours of interaction partners are analysed (Cilia & Le Driant, 2020; Cilia et al., 2020; Schietecatte et al., 2012). Other studies have questioned visual attention (face exploration, social scenes, eye tracking and joint attention) using eye-tracking methodology (for review: Cilia et al., 2018; Guillon et al., 2014).

Face perception plays a key role in the development of social interaction. According to Chawarska et al. (2013), a visual attention deficit for faces is dependent on the context of presentation and is related to the presence of explicit cues of a face directed at the child as well as language directed at the child. Eye-tracking studies do not necessarily show face aversion in ASD (e.g. Cilia et al., 2019a, b) or show

✉ Federica Cilia
federica.cilia@u-picardie.fr

¹ Centre de Recherche en Psychologie: Cognition, Psychisme Et Organisations, CRP-CPO, UR UPJV7273, Université de Picardie Jules Verne, Amiens, France

² Centre de Recherches Sur Les Fonctionnements Et Dysfonctionnements Psychologiques CRFDP, UR URN7475, Université de Rouen Normandie, Rouen, France

³ Chirurgie Et Extrémité Céphalique Caractérisation Morphologique Et Fonctionnelle, Chimère, UR UPJV7516, Université de Picardie Jules Verne, Amiens, France

mild face aversion but intense facial expressions and head orientation in such a way as to enable them to take in more information about the face (Franchini et al., 2017b). Griffin and Scherf (2020) indicate that the difficulty for ASD individuals is not related to face exploration but rather to understanding the social communication aspects of gaze information. Nevertheless, the pattern of exploration is sometimes distinctive with marked staring at the background when presented with a face stimulus (Deschamps et al., 2014). This is also observed in more natural interaction during the use of the Mundy & colleagues (2003) Early Social Communication Scale (ESCS) for example (Cilia & Le Driant, 2020; Cilia et al., 2020; Schietecatte et al., 2012). In a joint attention scene, the target (also called the referent) of the joint attention scene is a salient feature of the interaction scene (Freeth et al., 2011). However, it should be noted that ASD children who are most attracted to social scenes are those who show a greater degree of joint attention behaviour (Franchini et al., 2017a) and that the results differ according to the population. For example, in ASD adolescents without intellectual disabilities, Griffin and Scherf (2020) found no difference in fixation time on faces or joint attention targets compared to a control group. But in 8-year-olds with a level of communicative development far below their chronological age (i.e., 2 years), Cilia (2021) show less fixation on the joint attention referent in ASD children compared to control group children matched for their developmental age.

In an episode of joint attention, several deictic cues can be used: gazes, pointing, and verbalisations. For gaze following, results are contradictory and depend on the study population and the stimuli used. At 10 months infants at high familial risk for ASD mostly follow gaze in the head orientation condition compared to the eye orientation condition, whereas there is no difference between the two conditions in children at low risk of developing autism (Thorup et al., 2016). Between 7 and 13 month ASD children are able to follow the direction of gaze but pay less prolonged attention to the object of interest compared to typically developing babies (Bedford et al., 2012). At age 7, the difference between ASD children with no developmental delays and typical children is becoming more pronounced. There is no effect of eye or head orientation on gaze following (Cilia et al., n.d.). Furthermore, if in typical children, the duration of first fixation on the object is longer when the object is congruent to the gaze, this is not the case for children with ASD who look at the object for the same length of time regardless of whether the actor is looking at it or not (Swanson & Siller, 2013). This raises questions about their understanding of gaze direction. On this subject, in 7 years old ASD children when the gaze was incongruent with the position of the target the time spent looking at eyes and gaze tracking objects is negatively correlated. Even if they are looking at the eyes, they are less likely to use gaze information. This result is

not observed in the condition where gaze is congruent to the target, the authors explain this difference by the interactive complexity that results from the incongruence between gaze and target (Qiandong Wang et al., 2020). Finally, in a situation of joint attention, another visually salient element that involves movement corresponds to the pointing gesture. Benjamin and colleagues (Benjamin et al., 2014) show the effect of adding deictic cues in the visual exploration of school-aged ASD children (4–10 years). In the pointing and head orientation condition, ASD children explore the target and the actor's face for the same amount of time as their typical developmental peers matched on nonverbal cognitive level (typical children aged 2–5 years). Pointing therefore helps low-developmental ASD children to better perceive a joint attentional proposition. Vocalisations can also be studied as deictic cues added to gazes or pointing in a joint attentional proposition. Chawarska et al. (2013) show that if there is no verbal cue and eye-to-eye gaze, then the distribution of attention between the main features of the social scene is comparable between young ASD and typical children. However, when the gaze and verbal cues are directed towards the child, ASD children aged 13 to 25 months watch the scene with less attention. These results highlight that time spent exploring a scene with verbal cues is associated with autistic symptoms as well as with lower levels of non-verbal functioning than in control children.

While experimental studies address this issue by presenting a congruent or non-congruent gaze to the target of joint attention, other studies provide more naturalistic data. Schietecatte et al. (2012) analysed the joint attention response behaviours of 36-month-old ASD children during a social preference situation. The results show that 65.2% of children with ASD followed the adult's gaze. Cilia and Le Driant (2020) and Cilia et al. (2020) show a developmental progression in the responses of children with ASD provided that several deictic cues are used together (the adult looks and points, or looks and verbalises).

The current study

We conducted this study using the eye-tracking device to assess the effect of the presence and absence of the referent in the child's visual field during a proposed joint attention using different deictic cues (gazes, pointing, vocalisations). The occurrence of joint attention episodes also depends on the characteristics of the two partners. That is, the adult's ability to perceive the specificity of the ASD's functioning and to motivate the ASD to interact with them on a target (Khalilzadeh & Khodi, 2021). Compared to previous studies that present different targets, some congruent to the proposed deictic cues, others non-congruent (Astor et al., 2021; Chawarska et al., 2013; Griffin & Scherf, 2020; Wang

et al., 2020), the originality of our study lies in a purified stimulus where the deictic cues are necessarily congruent to the presentation of the referent. This choice of stimulus also follows the analysis of preliminary results from a thesis whose abstract has been published (Cilia, 2021). These initial results highlighted the value of using a video sequence rather than photographs to represent cues to joint attention. Preliminary results also highlighted the value of using a gaze-congruent joint attention target.

Method

Participants

Twenty-six children with ASD took part in the study with an average chronological age of 7,5 (2,6) years old but a developmental age of 21,10 (8,8) month. They were recruited either through the *Centre Ressources Autisme* (CRA) of Picardy or through the medico-social institutions where the children are cared for. They received a DSM-IV-TR or DSM-5 clinical consensus diagnosis of Autism or, respectively, ASD. The diagnosis was made using standardised instruments (Autism Diagnostic Interview-Revised: ADI-R and/or Autism Diagnostic Observation Schedule-Generic: ADOS-G) and confirmed by a health professional in *Hauts-de-France*. However, we did not obtain permission to access to all the children's scores at these scales. These data are therefore not included in this study. In addition, a technical problem resulted in the loss of eye-tracking data from 10 ASD children who could not be included in the analyses.

Symptoms severity was measured with the Childhood Autism Rating Scale (CARS). Twenty-nine typically developing children (TD) matched for developmental communication age as assessed using the French version of the Early Social Communication Scale (ECSP, Guidetti & Tournette, 2009), and 27 typically developing children matched for chronological age (TC) participated in this study. Given the chronological age of ASD children, it seemed essential to us to propose two control groups, one matched in age of communicative development, and the other linked to their life experience and therefore of the same chronological age. Table 1 presents the characteristics of the final sample.

Stimulus

The children watched stimuli from different protocols from more general research on social cognition (Cilia, 2021): the results presented here are from a part of a 58-s video. The overall protocol presented to the participants lasted approximately 10 min. The order of presentation of the stimuli was counterbalanced and the presence of

Table 1 Descriptive statistics of participants

	ASD	TC	TD
N	16	27	29
Gender	♀ = 2 ♂ = 14	♀ = 17 ♂ = 10	♀ = 10 ♂ = 19
Chronological age (years, months) M(SD)	7,5 (2,6)	7,7 (2,9)	2 (0,4)
Developmental age at ESCS (months, days) M(SD)	21,10 (8,8)		24,1 (5,4)
Total ESCS score	124,21 (59,8)		143,7 (38)
CARS score	34,4 (7)		

ASD represent Autism Spectrum Disorder group, TC represent typical children matched on chronological age, TD represent typical children matched on developmental age based on ESCS total score

the referent on the right or left of the screen was also counterbalanced.

The sequence is as follows:

- The actress looks to the left of the screen without moving her head, then looks in the direction of the child (duration of 3 s; sequence called: Eyes), the actress turns her head to look at one side of the screen, then looks in the direction of the child (duration of 3 s; sequence called: Head). The Eyes and Head sequences were combined to form the sequence: Gaze (6 s)
- The actress turns her head to look at one side of the screen and points and then looks in the direction of the child and says "Did you see? Then the actress turns her head to look at one side of the screen while continuing to point and says: "A ball! What colour is it? (duration 6 s; sequence called: GPV for Gaze, Pointing, Verbalisations).
- The actress repositions her hands on the table, turns her head to look at one side of the screen and says: "That's my favourite colour", then faces the camera and says: "What is your favorite color? (6 s long; sequence called: Verbalisations).

These sequences of equal duration were constructed to take into account both visual and auditory salience. Thus, in all conditions, we proposed 2 head movements. In the conditions with verbalisation, we proposed 2 sentences addressed to the child. Finally, in the pointing condition, we made the methodological choice of proposing only one additional movement in order to avoid adding visual salience by multiplying the pointing gestures.

Device

The study was carried out with an SMI-RED 250 Hz eye-tracker (SensoMotoric Instruments, SMI) calibrated at 60 Hz. Data and stimuli were recorded in SMI's *I View X* and *Experiment Center* software respectively. They were presented on a 17" screen in 4:3 format, the dimensions of which were 34.7 × 25.9 cm.

In autism studies using eye-tracking, researchers mostly use data according to predefined areas of interest called AOI (Areas Of Interest) (for review: Guillon et al., 2014). AOIs correspond to well-defined areas of the screen, in which visual information will be extracted. However, Wang and colleagues (Wang et al., 2018) remind us how difficult it is to define typical or atypical visual attention. This is even more complex when the stimulus used is a video. Indeed, in this context, contextual changes involve moving AOIs that can change frame by frame. Based on this observation, some authors have created new techniques that allow analysis of the results using a bottom-up approach. In our study, we coupled two bottom-up methods: mean-shift clustering (Santella & DeCarlo, 2004) and Voronoi-Tessellation (Over et al., 2006). The main advantage of coupling these methods is that they are dependent on the nature of the distribution and the location of the fixations and that one allows the automation of the other. Furthermore, unlike the creation of AOIs by the researcher, which can be laborious and time-consuming, these methods can be automated and do not depend on humans, but on an algorithm. To use these methods, we created a script on the free software R. This script was created using several packages that we adapted to our data.

For details of the method for creating a posteriori AOI (see Cilia et al., 2019a, b).

Procedure

During the eye-tracking procedure, the child was seated comfortably at a distance of about 60 cm from the screen, either on a chair, on a parent's lap, or on a highchair for the younger children. The experimental space was uncluttered to minimise distracting elements. To this end, the child faced the computer screen, surrounded by two white curtains. Behind each child, the parent or experimenter could help maintain the optimal position for data recording. If the child was distracted, we asked them to look at the screen.

Data analysis

After clustering, we defined the AOIs "Background", "Body", "Face", "Pointing", "Object/Referent", which depend on the deictic index used. We compared the data for the head orientation index without verbalisations (Gaze) with the pointing index with verbalisations (GPV) and the head orientation index with verbalisations (Verbalisations). Figure 1 shows a visual representation of the clustering of the data for the 3 indices.

The actress was informed and gave her written consent for scientific publication of her image.

According to Aslin (2007), the interest of eye-tracking lies in the use of composite data from the micro-structure of the gaze (i.e., fixations) which it is necessary to link to visual behaviours (i.e., displacements and durations of fixations). Thus, after this a posteriori AOI creation work, we chose

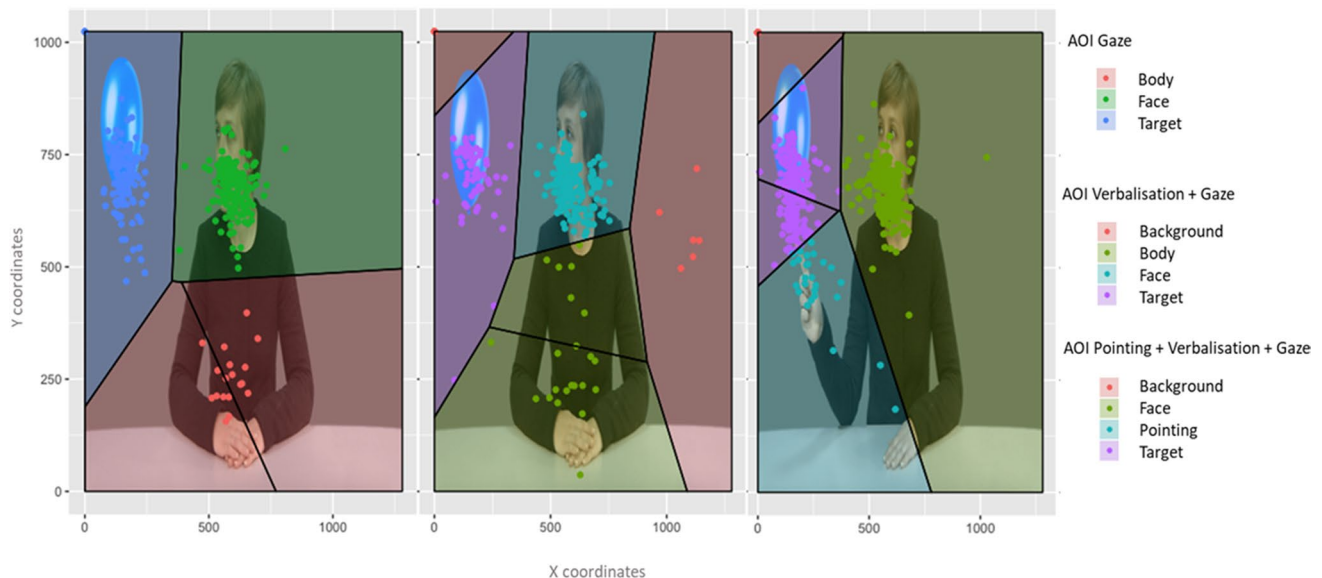


Fig. 1 A posteriori AOI representation of the GPV index

to focus on two dependent variables found in the literature (Cilia et al., 2019a, b; Charrier et al., 2017). The variable relative fixation number (RFN), i.e., the number of fixations on an AOI as a function of the total number of fixations on the screen for each child, as well as the variable relative fixation duration (RFD), which takes into account the duration of these fixations as a function of the total duration of each child's fixations on the screen. The RFD therefore shows the time spent fixating on an area in relation to the rest of the screen. In terms of joint attention, we analyse children's fixations on the referent only if they were preceded by glances at the actress's face or hand when pointing. It is therefore not a simple object present on the scene but a referent, a target of attention in a situation of joint attention. In this perspective, we represent this data by proposing a percentage of the number of children looking at the target as a proportion of the total number of children in the group.

Hypothesis

In this study we made several hypotheses. We thought that visual attention would differ according to the group of children. We thought that compared to typical children, ASD children would fixate less on the face, less on the referent and more on the background. We expected that the deictic cue used would have an impact on fixations. Compared to a scene using only gaze to make a joint attention proposal, we thought the addition of a deictic cue such as pointing accompanied by verbalisations would have a positive impact on social AOI fixations and the joint attention response (e.g., Benjamin et al., 2014; Franchini et al., 2017a, b). In line with the studies of Chawarska and colleagues (Chawarska et al., 2012, 2013), we believed that verbalisations would have a negative impact on the visual joint attention response in ASD children. To test these hypotheses, after checking the normality and homogeneity of the data, we used analyses of variance (ANOVA) with Bonferroni correction and comparisons of means (Student's *t* test).

Results

There was a group effect on gaze tracking, pointing and/or verbalisations to the referent for RFN ($F(2,208) = 7.57$, $p < 0.001$; $\eta^2 p = 0.068$). The RFN is lower in ASD children compared to TC children ($t(208) = 2.70$, $p = 0.02$) and TD children ($t(208) = 3.84$, $p < 0.001$). But the effect was not significant for RFD ($F(2,208) = 2.84$, $p = 0.060$; $\eta^2 p = 0.027$). There is also an effect of the type of cue used (Staring, Pointing, Verbalisations) towards the present referent for the RFN ($F(2,208) = 21.4$, $p < 0.001$; $\eta^2 p = 0.170$) and the RFD ($F(2,208) = 23.1$, $p < 0.001$; $\eta^2 p = 0.182$).

Response to joint attention

Gazes index When the cue used was gaze alone (Gazes), 12% of ASD children ($N = 2$) followed the orientation of the eyes and/or head towards the referent, compared to 74% of TC children ($N = 20$) and 62% of TD children ($N = 18$) (see Fig. 2). The number and duration of fixations on the referent are lower in ASD children compared to TD children and number to TC children.

Verbalisations index When the actress looks and verbalises towards the target (Verbalisations), only one ASD child looks in the right direction, which represents 0.6% of the sample, compared to 37% of TC children ($N = 10$) and 52% of TD children ($N = 15$) (see Fig. 2). The RFN on the target is lower for ASD than TC and TD.

GPV index When the actress looks, points and verbalises (GPV) to the target, 31% of ASD children ($N = 4$) follow the cues to the target, 92% of TC children ($N = 25$) and 93% of TD children ($N = 27$) (see Fig. 2). The RFN on the target are lower in ASD children compared to TCs and compared to TDs but RFD is higher for ASD than TC. For descriptive and *t* test statistics please see Table 2 for RFN statistics and Table 3 for RFD statistics.

We conducted a within-group analysis in ASD to test the effect of stimulus type. The comparison between the indices shows that fixations to the target during gaze tracking (Gaze index) are higher than those during verbalization tracking (Verbalisations index) for the RFN: $t(9) = 2.49$, $p = 0.034$. The RFN and the RFD on the target are higher in the pointing plus verbalisation condition (GPV index) than in the verbalisation condition (Verbalisation index) (RFN: $t(12) = 2.26$, $p = 0.043$, RFD: $t(12) = 3.39$, $p = 0.005$). For the comparison between the Gaze index and the combination of the three indices (GPV), there was no difference between the two conditions (RFN: $t(208) = 0.559$, $p = 0.842$; RFD: $t(208) = 0.798$, $p = 0.798$). For descriptive statistics please see Table 2 for RFN and Table 3 for RFD.

Visual exploration by AOI

Between the Gaze, Verbalisations and GPV conditions, we find an effect of AOI (RFN: $F(4,1252) = 269.9$, $p < 0.001$; $\eta^2 p = 0.463$; RFD: $F(4,1252) = 182.71$, $p < 0.001$; $\eta^2 p = 0.369$), an interaction effect between AOI and group (RFN: $F(8,1252) = 6.66$, $p < 0.001$; $\eta^2 p = 0.041$; RFD: $F(8,1252) = 4.37$, $p < 0.001$; $\eta^2 p = 0.027$), an interaction effect between the AOI and the stimulus (RFN: $F(20,1252) = 7.16$, $p < 0.001$; $\eta^2 p = 0.103$; RFD: $F(20,1252) = 10.85$, $p < 0.001$; $\eta^2 p = 0.148$), and finally,

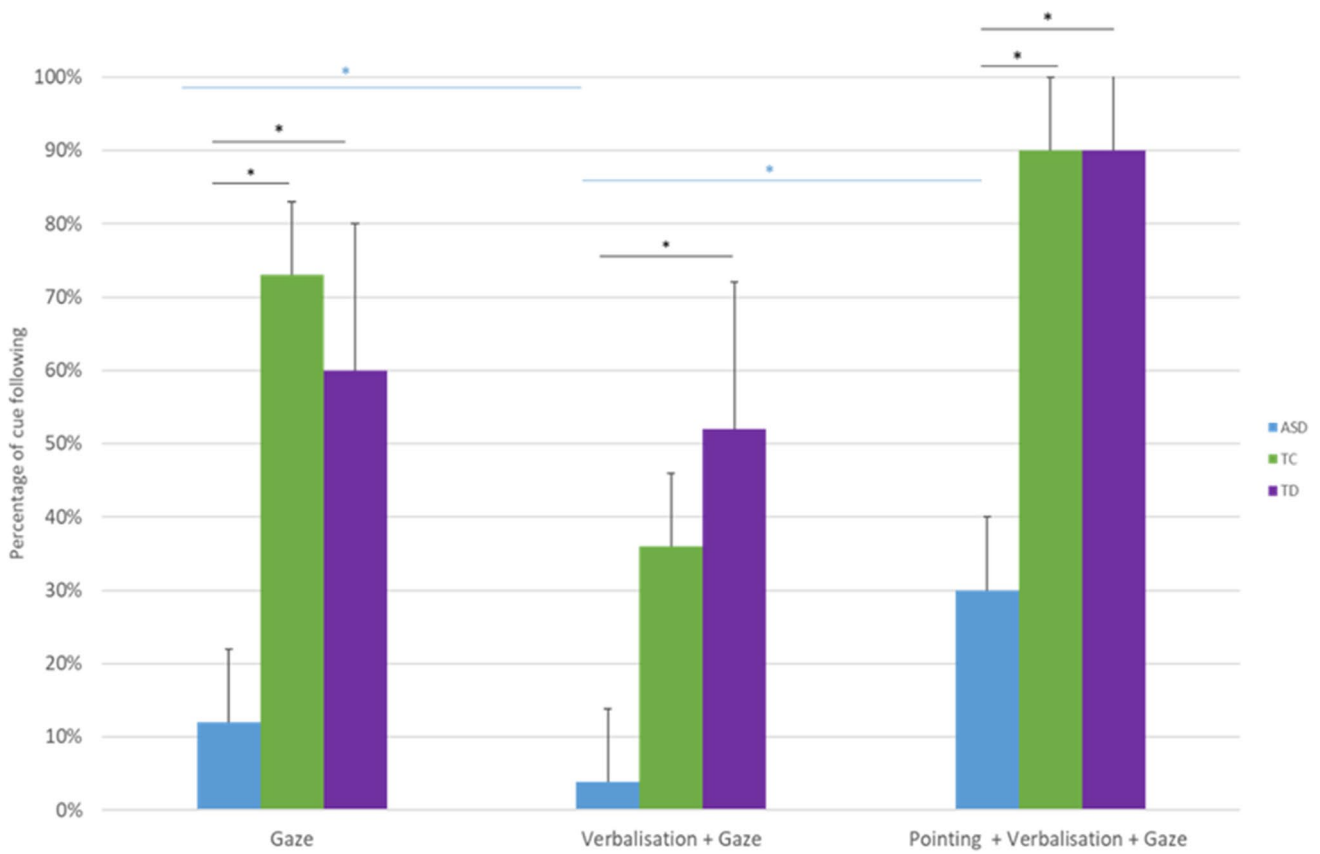


Fig. 2 Percentage of children succeeding in following the cue depending on deictic cue by group. Notes: * $p < 0.05$ Error bars represent standard deviation of the mean value. ASD represent Autism

Spectrum Disorder group; TC represent typical children matched on chronological age; TD represent typical children matched on developmental age based on ESCS total score

we observe an interaction effect between AOI, group and stimulus (RFN: $F(40,1252) = 1.54, p = 0.017; \eta^2 p = 0.047$; RFD: $F(40,1252) = 1.54, p = 0.017; \eta^2 p = 0.047$). We therefore conducted inter-group mean comparisons.

Gazes index RFN and RFD on the face were lower in ASD children compared to TCs and RFD compared to TDs. In this condition the creation of a posteriori AOI did not allow the creation of background AOI.

Table 2 RFN statistics

RFN	Group	ASD	TC	TD	ASD vs TC	ASD vs TD
AOI	Stimulus	M (SD)	M (SD)	M (SD)	T test	T test
background	Gaze	/	/	/	/	/
	Verbalisation	0.0222 (0.0497)	0.0588 (0.0944)	0.0176 (0.0532)	$t(131) = -4.12, p < .001$	$t(971) = -17.879, p < .001$
	GPV	0.0833 (0.167)	0.0121 (0.0402)	/	NS	/
face	Gaze	0.462 (0.227)	0.603 (0.199)	0.504 (0.224)	$t(131) = 2.424, p = .016$	NS
	Verbalisation	0.645 (0.393)	0.712 (0.181)	0.594 (0.274)	NS	NS
	GPV	0.435 (0.298)	0.591 (0.220)	0.481 (0.163)	$t(131) = 2.40, p = .018$	NS
pointing target	GPV	0.250 (0.500)	0.00568 (0.0188)	0.0728 (0.124)	$t(131) = -6.11, p < .001$	$t(971) = -4.675, p < .001$
	Gaze	0.326 (0.250)	0.426 (0.166)	0.434 (0.177)	$t(596) = 2.316, p = .021$	$t(1269) = 2.09, p = .036$
	Verbalisation	0.125 (0.00)	0.418 (0.046)	0.360 (0.135)	$t(524) = 3.538, p = .020$	$t(912) = 3.357, p < .001$
	GPV	0.259 (0.143)	0.346 (0.240)	0.443 (0.178)	$t(131) = 2.54, p = .012$	$t(971) = 4.009, p < .001$

AOI mean Areas of Interest, M means Mean, SD mean Standard Deviation, NS mean Not Significant

Table 3 RFD statistics

RFD	Group	ASD	TC	TD	ASD vs TC	ASD vs TD
AOI	Stimulus	M (SD)	M (SD)	M (SD)	T test	T test
background	Gaze	/	/	/	/	/
	Verbalisation	0.00704 (0.0157)	0.0337 (0.0540)	0.00840 (0.0326)	NS	NS
	GPV	0.0201 (0.0402)	0.00535 (0.0177)	/	/	/
face	Gaze	0.493 (0.371)	0.737 (0.173)	0.623 (0.236)	$t(131)=4.40, p<.001$	$t(912)=2.13, p=.034$
	Verbalisation	0.674 (0.431)	0.841 (0.120)	0.736 (0.280)	$t(134)=3.612, p<.001$	NS
	GPV	0.370 (0.266)	0.755 (0.152)	0.599 (0.215)	$t(131)=8.15, p<.001$	$t(912)=2.13, p=.034$
pointing	GPV	0.250 (0.500)	0.00145 (0.00482)	0.0585 (0.0967)	$t(131)=-6.26, p<.001$	$t(971)=-6.208, p<.001$
target	Gaze	0.312 (0.375)	0.283 (0.148)	0.384 (0.177)	NS	$t(1269)=2.17, p=.03$
	Verbalisation	0.022 (0.00)	0.226 (0.087)	0.184 (0.181)	NS	NS
	GPV	0.342 (0.237)	0.258 (0.143)	0.360 (0.226)	$t(131)=-2.54, p=.010$	NS

AOI mean Areas of Interest, M means Mean, SD mean Standard Deviation, NS mean Not Significant

Verbalisations index RFD on the face was lower in ASD children than in TCs. Fixations number to the background are higher in ASD children compared to TCs and TDs.

GPV index Face fixations are lower in ASD children compared to TCs, the same is true for RFD compared to TDs. The background data is sparse and does not allow for robust parametric statistical analyses. Finally, in this condition the creation of the AOI a posteriori allowed the creation of the AOI pointing, fixations on this AOI are higher in ASD children compared to TCs and TD. For descriptive and t test statistics please see Table 2 for RFD statistics and Table 3 for RFD statistics.

Discussion

The aim of this research was to investigate the impact of different deictic cues on the visual response to joint attention. We compared data from the 'Gaze' cue with eye orientation and head orientation, gazes associated with 'Verbalisations' and, finally, gazes associated with both verbalisations and pointing for the 'GPV' condition.

In all conditions, it was generally observed that ASDs followed the joint attention proposal less often than control children. When the actress looks towards a referent, only 12% of ASD children follow the gaze towards the target, whereas most typical children in the TD and TC groups followed the gaze towards the referent cue. This initial finding highlights that this cue and the body movement that follows it is relevant for typically developing children (Astor et al., 2021) but not for most ASD children (Swanson & Siller, 2013). Verbalisations associated with gazing at a target even decrease the possibility of a joint attention response for all children. Chawarska and colleagues (Chawarska et al., 2012, 2013) observe that verbalisations increased the difficulties

for ASD children in gaze following. This may be related to difficulties in disengaging from a stimulus (Bedford et al., 2012) i.e., the more the child disengages their attention from the object, the more they look at the adult and alternate their gaze in order to share their attention and engage in referent interaction. Furthermore, Benjamin and colleagues' (Benjamin et al., 2014) study shows that naming an object while looking at it does not help to draw ASD and typical children's attention to that object compared to a situation where it was not named. The TD children in our group range in age from 9 to 30 months, and studies of typical children's joint attention response show that at 6 months verbalisations associated with gazes help children understand that these non-verbal and vocal cues are directed at the same object (Senju & Csibra, 2008), whereas from 14 months onwards, children no longer need verbalisations to follow gaze direction (Deák et al., 2000). The age of our TD sample and the stimulus used may explain these differences in results. In our study, the coupling of the 3 cues, i.e., the verbalisations associated with pointing and gazing at the referent, resulted in a higher percentage of a joint attention response for all groups of children. While 31% of the ASD children followed the cues to the referent, for children in the control groups the follow-up of the gaze associated with pointing and verbalisations was up to 90%. The plurality of deictic cues used therefore allows ASD children to respond to a joint attention proposal whether the study is experimental (Benjamin et al., 2014; Franchini et al., 2017a, b) or more naturalistic (Cilia & Le Driant, 2020; Cilia et al., 2020). One may think that the results would have been even more significant if the stimulus did not have a defined duration. Indeed, in our video stimulus, if children stare at the pointing finger, they cannot stare at the referent at the same time. In this respect, ASD children stare at the pointing finger more and for longer than the control groups. Though we thought that the pointing cue would allow ASD children to explore the joint attention

scene in a similar way to typical children, we didn't take into consideration the temporal limit. It would have been interesting to use triggers that allowed the stimulus used to be adaptive and thus to offer the next cue only when the children had had the opportunity to follow the deictic cue to the referent. But this was not a problem for children in the control groups, regardless of age. It seems, therefore, that Gepner's (2001) movement perception impairment hypothesis explains this result in ASD. Indeed, it is likely that the result would have been different if the stimulus had been slowed down so that the ASD children had the necessary time to explore the face (Charrier et al., 2017; Gepner et al., 2020) and all the cues proposed in the scene. These visual peculiarities also raise questions about the possibility of categorising eye-tracking data to define a particular pattern in ASD compared to typically developing children (Cilia, 2021).

Our results on the tracking of deictic cues to the referent of joint attention could be related to a particular visual exploration of the stimulus. We therefore conducted analyses on the visual exploration of the different AOIs present on the stimulus. The results of the inter-group analyses showed that for each stimulus (i.e., 'Gaze', 'Verbalisations', 'GPV') ASD children looked at the face for a shorter period of time than children in the control groups (Benjamin et al., 2014; Cilia & Le Driant, 2020; Cilia et al., 2020). Background fixations are more numerous in the Verbalisation condition, whereas they are non-existent in the Gaze condition and few in the GPV condition. ASD children focus more on the background than children in control groups (Deschamps et al., 2014). As in Chawarska et al. (2013) study, under these particular conditions, verbalisations addressed to the child with direct gaze seem to have a negative impact on the exploration of a social scene and this seems to be related to the developmental age of ASD children (Cilia & Le Driant, 2020; Cilia et al., 2020).

It can therefore be concluded that it is practically useful to employ several cues to initiate joint attention towards a referent that is sufficiently salient and interesting for the child (Freeth et al., 2011). This is also widely used in specific developmental care. These results are found in studies with naturalistic methodology and in eye-tracking studies. If the more experimental tasks that use a precise methodology such as eye-tracking have the advantage of being very refined and make it possible to limit confounding variables as much as possible, the more naturalistic studies have the advantage of providing more clinical knowledge. But interaction is complex to analyse because it involves facial micro-expressions or wider body movements which, although they help the child to respond to a proposal for joint attention, are not controlled and therefore cannot always be finely analysed in the above-mentioned studies. Obviously, our study has other limitations. The sample size, due to data loss, was even smaller than originally expected for ASD children. Nevertheless, in view of the innovative and currently little used method of creating AOI a posteriori, we did not wish to

introduce noise into our data by averaging them for example, so we did not replace these data.

Finally, another variable that remains to be tested is the impact of the slowing down of movement (Gepner, 2001). Indeed, in practice, clinicians are asked to take more time to teach this type of social behaviour to ASD children. Furthermore, to grasp the meaning of pointing and develop these behaviours, the Early Start Denver Model intervention program (ESDM, Rogers & Dawson, 2013) suggests using several strategies. The object-based joint activities in ESDM prepare the child to develop joint attention skills. Indeed, children can point to the material, taking their gaze off the object to focus on the adult until the gaze alternates between the object and the adult. In addition, during 'social sensory routines', each partner focuses on the other to encourage social orientation and communication in ASD children. A longitudinal study highlights the value of this method for maintaining joint attention (Cilia et al., n.d.). Finally, the 'one more word' method is also proposed in the ESDM. The statements produced by the adult have one more word than those produced by the child. It would have been interesting to create a video scenario adapted to the child's verbal level in order to take into account this variable which may explain our results.

Conclusion

In this study, we highlighted the value of certain deictic cues in getting a response following an initiation of joint attention. Our results suggest that eye or head orientation are not sufficiently salient to get referent fixation in ASD children. We were able to show that the addition of verbalisations to gaze direction implies less visual response to joint attention. This may be related to the social or perceptual difficulties of ASD children. For example, compared to gaze alone, verbalisations increase ASD children's difficulties in responding to joint attention (Chawarska et al., 2013). In contrast, more responses are observed when the actress looks, points and verbalises at the same time. The lack of face information intake compared to control groups is also highlighted and would reflect the lack of understanding of the social communication and directionality aspects provided by gaze (Griffin & Scherf, 2020). Overall, our results suggest the importance of increasing the number of visual cues (pointing or other gestures) combined with auditory stimuli (addressed verbalisations and description of the scene), particularly in a classical clinical context.

Acknowledgements We would like to thank the participants in this study, their families, the institutions that allowed us to work with them, and the health care teams for the invaluable help they gave us while we were conducting the research.

Funding Financial support for translation of the article from French to English was received from CRP-CPO research laboratory and Université de Picardie Jules Verne to which the corresponding author belong.

Data availability The research data are available on Federica Cilia's Open Science Framework page: <https://osf.io/jxa6z/>

Declarations

Ethical statement Ethical approval was given by a local ethics committee and the CNIL (*Commission nationale de l'informatique et des libertés*: National Commission on Informatics and Liberty). Declaration number of research conformity is: 2208663 v 0. This study respects the terms of the 1964 Helsinki declaration. Before starting the study, approval was obtained from the heads of the regional and district education authorities, as well as the head and the teachers of the school. Parents also gave written informed consent and all the children were volunteers.

Competing interests All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders DSM-5*. American Psychiatric Publishing.
- Aslin, R. N. (2007). What's in a look? *Developmental Science*, 1(10), 48–53. <https://doi.org/10.1111/j.1467-7687.2007.00563.x>
- Astor, K., Thiele, M., & Gredebäck, G. (2021). Gaze following emergence relies on both perceptual cues and social awareness. *Cognitive Development*, 60(September), 1–8. <https://doi.org/10.1016/j.cogdev.2021.101121>
- Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T., & Johnson, M. H. (2012). Precursors to Social and Communication Difficulties in Infants At-Risk for Autism: Gaze Following and Attentional Engagement. *Journal of Autism and Developmental Disorders*, 42(10), 2208–2218. <https://doi.org/10.1007/s10803-012-1450-y>
- Benjamin, D. P., Mastergeorge, A. M., McDuffie, A. S., Kover, S. T., Hagerman, R. J., & Abbeduto, L. (2014). Effects of labeling and pointing on object gaze in boys with fragile X syndrome: An eye-tracking study. *Research in Developmental Disabilities*, 35(11), 2658–2672. <https://doi.org/10.1016/j.ridd.2014.06.021>
- Cilia, F. (2021). Gazes, pointing gestures, and vocalizations: Perception of joint attention bids in children with autism spectrum disorder. *Bulletin De Psychologie, Numéro*, 571(1), 49. <https://doi.org/10.3917/bupsy.571.0049>
- Cilia, F., & Le Driant, B. (2020). Les enjeux de l'attention conjointe dans les processus de communication : particularités des enfants sourds ou atteints d'un trouble du spectre de l'autisme. *A.N.A.E.*, 166, 1–8.
- Cilia, F., Le Driant, B., Garry, C., Vandromme, L., & Brisson, J. (n.d.). Evolution of joint attention in ASD children regarding ABA and ESDM programs. *PsyArXiv*, 1–23.
- Cilia, F., Garry, C., Brisson, J., & Vandromme, L. (2018). Attention conjointe et exploration visuelle des enfants au développement typique et avec TSA : Synthèse des études en oculométrie. *Neuropsychiatrie De L'enfance Et De L'adolescence*, 66(5), 304–314. <https://doi.org/10.1016/j.neurenf.2018.06.002>
- Cilia, F., Aubry, A., Bourdin, B., & Vandromme, L. (2019a). How to determine visual areas of interest without any preconceptions? Eye-tracking fixation analysis in the field of autism. *Revue De Neuropsychologie*, 11(2), 144–150. <https://doi.org/10.1684/nrp.2019.0487>
- Cilia, F., Aubry, A., Bourdin, B., Le Driant, B., & Vandromme, L. (2019b). Visual exploration of dynamic or static joint attention bids in children with autism syndrome disorder. *Frontiers in Psychology*, 10, 1–11. <https://doi.org/10.3389/fpsyg.2019.02187>
- Cilia, F., Touchet, C., Vandromme, L., & Le Driant, B. (2020). Initiation and response of joint attention bids in autism spectrum disorder children depend on the visibility of the target. *Autism and Developmental Language Impairments*, 5, 1–11. <https://doi.org/10.1177/2396941520950979>
- Cilia, F., Carette, R., Elbattah, M., Dequen, G., Guérin, J.L., Bosche, J., Vandromme, L., Le Driant, B. (2021). Computer-aided screening of Autism Spectrum disorder: Eye-tracking study using data visualization and deep learning. *JMIR Human Factors*, 8(4), 1–11. <https://doi.org/10.2196/27706>
- Charrier, A., Tardif, C., & Gepner, B. (2017). Amélioration de l'exploration visuelle d'un visage par des enfants avec autisme grâce au ralentissement de la dynamique faciale : Une étude préliminaire en oculométrie. *L'encephale*, 43(1), 32–40. <https://doi.org/10.1016/j.encep.2016.02.005>
- Chawarska, K., Macari, S., & Shic, F. (2012). Context modulates attention to social scenes in toddlers with autism. *Journal of Child Psychology and Psychiatry*, 8(53), 903–913. <https://doi.org/10.1111/j.1469-7610.2012.02538.x>
- Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, 74(3), 195–203. <https://doi.org/10.1016/j.biopsych.2012.11.022>
- Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, 36(4), 511–523. <https://doi.org/10.1037/0012-1649.36.4.511>
- Deschamps, L., Leplain, S., & Vandromme, L. (2014). Exploration visuelle de visages chez les enfants de bas niveau cognitif avec trouble du spectre de l'autisme : une étude de eye-tracking. *Enfance*, 399–425. <https://doi.org/10.4074/S0013754514004017>
- Franchini, M., Glaser, B., De Wilde, H. W., Gentaz, E., Eliez, S., & Schaer, M. (2017a). Social orienting and joint attention in preschoolers with autism spectrum disorders. *PLoS ONE*, 12(6), 1–14. <https://doi.org/10.1371/journal.pone.0178859>
- Franchini, M., Glaser, B., Gentaz, E., Wood, H., Eliez, S., & Schaer, M. (2017b). The effect of emotional intensity on responses to joint attention in preschoolers with an autism spectrum disorder. *Research in Autism Spectrum Disorders*, 35, 13–24. <https://doi.org/10.1016/j.rasd.2016.11.010>
- Freeth, M., Foulsham, T., & Chapman, P. (2011). The influence of visual saliency on fixation patterns in individuals with Autism Spectrum Disorders. *Neuropsychologia*, 49(1), 156–160. <https://doi.org/10.1016/j.neuropsychologia.2010.11.012>
- Gepner, B. (2001). Malvoyance du mouvement dans l'autisme infantile? *La Psychiatrie De L'enfant*, 44(1), 77. <https://doi.org/10.3917/psy.441.0077>
- Gepner, B., Godde, A., Charrier, A., Carvalho, N., & Tardif, C. (2020). Reducing facial dynamics' speed during speech enhances attention to mouth in children with autism spectrum disorder: An eye-tracking study. *Development and Psychopathology*, 33(3), 1006–1015. <https://doi.org/10.1017/S0954579420000292>
- Griffin, J. W., & Scherf, K. S. (2020). Does decreased visual attention to faces underlie difficulties interpreting eye gaze cues in autism? *Molecular Autism*, 11(1), 1–14. <https://doi.org/10.1186/s13229-020-00361-2>

- Guidetti, M., & Turrette, C. (2009). *Evaluation de la communication sociale précoce (ECSP)* (Eurotests). Eurotests Editions.
- Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience and Biobehavioral Reviews*, *42*, 279–297. <https://doi.org/10.1016/j.neubiorev.2014.03.013>
- Khalilzadeh, S., & Khodi, A. (2021). Teachers' personality traits and students' motivation: A structural equation modeling analysis. *Current Psychology*, *40*(4), 1635–1650.
- Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *Manual for early social communication scales (ESCS)*. Coral Gables, FL: http://www.ucdmc.ucdavis.edu/mindinstitute/ourteam/faculty_staff/escs.pdf
- Over, E. A. B., Hooge, I. T. C., & Erkelens, C. J. (2006). A quantitative measure for the uniformity of fixation density: The Voronoi method. *Behavior Research Methods*, *38*(2), 251–261. <https://doi.org/10.3758/BF03192777>
- Rogers, S. J., & Dawson, G. (2013). *L'intervention précoce en autisme. Le modèle de Denver pour jeunes enfants*. (Dunod (ed.)). Dunod.
- Santella, A., & DeCarlo, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the Eye Tracking Research & Applications Symposium on Eye Tracking Research & Applications - ETRA'2004*, 27–34. <https://doi.org/10.1145/968363.968368>
- Schietecatte, I., Roeyers, H., & Warreyn, P. (2012). Exploring the nature of joint attention impairments in young children with autism spectrum disorder: Associated social and cognitive skills. *Journal of Autism and Developmental Disorders*, *42*(1), 1–12. <https://doi.org/10.1007/s10803-011-1209-x>
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671. <https://doi.org/10.1016/j.cub.2008.03.059>
- Swanson, M. R., & Siller, M. (2013). Patterns of gaze behavior during an eye-tracking measure of joint attention in typically developing children and children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *7*(9), 1087–1096. <https://doi.org/10.1016/j.rasd.2013.05.007>
- Thorup, E., Nyström, P., Gredebäck, G., Bölte, S., & Falck-Ytter, T. (2016). Altered gaze following during live interaction in infants at risk for autism: An eye tracking study. *Molecular Autism*, *7*(1), 12. <https://doi.org/10.1186/s13229-016-0069-9>
- Wang, Q., Campbell, D. J., Macari, S. L., Chawarska, K., & Shic, F. (2018). Operationalizing atypical gaze in toddlers with autism spectrum disorders : A cohesion-based approach. *Molecular Autism*, *9*(25), 1–9.
- Wang, Q., Hoi, S. P., Wang, Y., Lam, C. M., Fang, F., & Yi, L. (2020). Gaze response to others' gaze following in children with and without autism. *Journal of Abnormal Psychology*, *129*(3), 320–329. <https://doi.org/10.1037/abn0000498>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.