Original Article

Autistic young people adaptively use gaze to facilitate joint attention during multi-gestural dyadic interactions

Nathan Caruana[1] , Patrick Nalepka[1], Glicyr A Perez[1], Christine Inkley[1],
Courtney Munro[1], Hannah Rapaport[2], Simon Brett[1], David M Kaplan[1],
Michael J Richardson[1] and Elizabeth Pellicano[1,3]

Abstract

Autistic people often experience difficulties navigating face-to-face social interactions. Historically, the empirical literature has characterised these difficulties as cognitive 'deficits' in social information processing. However, the empirical basis for such claims is lacking, with most studies failing to capture the complexity of social interactions, often distilling them into singular communicative modalities (e.g. gaze-based communication) that are rarely used in isolation in daily interactions. The current study examined how gaze was used in concert with communicative hand gestures during joint attention interactions. We employed an immersive virtual reality paradigm, where autistic ($n = 22$) and non-autistic ($n = 22$) young people completed a collaborative task with a non-autistic confederate. Integrated eye-, head- and hand-motion-tracking enabled dyads to communicate naturally with each other while offering objective measures of attention and behaviour. Autistic people in our sample were similarly, if not more, effective in responding to hand-cued joint attention bids compared with non-autistic people. Moreover, both autistic and non-autistic people demonstrated an ability to adaptively use gaze information to aid coordination. Our findings suggest that the intersecting fields of autism and social neuroscience research may have overstated the role of eye gaze during coordinated social interactions.

Lay abstract

Autistic people have been said to have 'problems' with joint attention, that is, looking where someone else is looking. Past studies of joint attention have used tasks that require autistic people to continuously look at and respond to eye-gaze cues. But joint attention can also be done using other social cues, like pointing. This study looked at whether autistic and non-autistic young people use another person's eye gaze during joint attention in a task that did not require them to look at their partner's face. In the task, each participant worked together with their partner to find a computer-generated object in virtual reality. Sometimes the participant had to help guide their partner to the object, and other times, they followed their partner's lead. Participants were told to point to guide one another but were not told to use eye gaze. Both autistic and non-autistic participants often looked at their partner's face during joint attention interactions and were faster to respond to their partner's hand-pointing when the partner also looked at the object before pointing. This shows that autistic people can and do use information from another person's eyes, even when they don't have to. It is possible that, by not forcing autistic young people to look at their partner's face and eyes, they were better able to gather information from their partner's face when needed, without being overwhelmed. This shows how important it is to design tasks that provide autistic people with opportunities to show what they can do.

Keywords

eye contact, gaze, non-verbal communication, social interaction, virtual reality

---

[1]Macquarie University, Australia
[2]University of Cambridge, UK
[3]University College London, UK

Corresponding author:
Nathan Caruana, School of Psychological Sciences, Macquarie University, 16 University Ave, Sydney, NSW 2109, Australia.
Email: nathan.caruana@mq.edu.au

Joint attention is the capacity to coordinate our attention with another person so that we are attending to the same thing (Tomasello, 1999). This social ability emerges in the first year of life and allows humans to share experiences, learn socially, develop language and cooperate with others (Adamson et al., 2009; Baldwin, 1995; Charman, 2003; Mundy & Newell, 2007; Murray et al., 2008). In a typical joint attention experience, one person *initiates* joint attention by looking at, gesturing towards or describing an object, while a second person recognises this behaviour as communicative and responds by attending to the same thing. Delays in joint attention development are diagnostic markers for autism and are associated with poorer social, language and learning outcomes (Bruinsma et al., 2004; Lord et al., 2000).

Various social cues, such as speech and manual gestures, can be used to guide joint attention experiences by signalling where a person is attending. However, most experimental studies have examined joint attention using paradigms that require gaze-based interactions and, relatedly, have operationalised joint attention as the ability to attend to where another person is *looking* (see Caruana et al., 2017b, for review). Part of the reason for this historical focus on gaze-based joint attention is that differences and delays in orienting to another's gaze during social interactions are considered 'hallmarks' of being autistic (Zwaigenbaum et al., 2005) and fundamental to the social challenges experienced by autistic people (Baron-Cohen, 1995). However, despite more than 40 years of intense empirical investigation, it remains unclear how reliable the differences in gaze use are among autistic people, and how consequential these differences are during joint attention interactions. Answers to these questions are crucial for determining whether the current characterisation of gaze-use differences among autistic people as '*cognitive deficits*' does more harm (e.g. by stigmatising autism) than good (e.g. by informing cognitive or social interventions).

The empirical evidence for gaze-use differences in autistic people is mixed. For instance, several face-scanning studies suggest that autistic people are less likely to look at other faces than non-autistic people (Nakano et al., 2010; Pelphrey et al., 2002), and when they do, they spend less time than non-autistic people fixating on the eyes (Kliemann et al., 2010; Tanaka & Sung, 2016). However, other studies have failed to replicate these findings (Chita-Tegmark, 2016; Falck-Ytter & von Hofsten, 2011; Frazier et al., 2017). Similarly, a large body of research has used Posner-style gaze-cueing paradigms to measure the extent to which autistic individuals reflexively orient to gaze cues (see Gillespie-Lynch et al., 2013; Nation & Penny, 2008, for related reviews). These studies have attempted to determine whether a reduced sensitivity to gaze information might explain joint attention differences in autism. Although some studies have reported evidence for smaller valid cue-based effects in autistic people than in non-autistic people (i.e. faster reaction times to detect targets preceded by a valid, rather than an invalid gaze cue), others have failed to replicate these findings.

Interactive methods for studying joint attention

There is now growing recognition that a key explanation for these inconsistent findings – aside from autistic heterogeneity – is that most experimental studies of eye gaze information processing have failed to capture the dynamic and reciprocal aspects of social interactions in which gaze difficulties are likely to arise (Cañigueral & Hamilton, 2019; Caruana et al., 2017b; Schilbach et al., 2013). In response to this problem, the last decade has seen innovations in experimental methods – particularly involving the use of interactive eye-tracking and virtual reality – to examine how people use gaze in interactive contexts to achieve joint attention (Caruana et al., 2017b; Schilbach et al., 2013). Some interactive studies have presented tentative evidence for joint attention difficulties in autistic children (e.g. Little et al., 2016) and adults (e.g. Caruana et al., 2018), characterised by increased errors and response times during gaze-based interactions. However, one key limitation of these paradigms – and other joint attention paradigms beyond the field of autism research – is that people can *only* communicate via gaze signals (Caruana et al., 2015, 2017b; Leekam, 2016; Mundy, 2018; Oberwelland et al., 2016; Redcay et al., 2012; Wilms et al., 2010). Yet, in everyday interactions, a joint attention episode can be initiated by one person looking towards, pointing at or naming an object or event, sometimes all at once (Seibert et al., 1982). Indeed, to respond successfully to most joint attention bids, the responding partner must parse, filter or integrate social information cues that go well beyond just the eyes (Mundy et al., 2009; Siposova & Carpenter, 2019). To date, little is known about how humans achieve joint attention when multiple sources of social information are available.

Multi-gestural joint attention interactions

In their recent theoretical and empirical work, Yu and Smith (2013, 2017) highlight that different social cues can be used to achieve joint attention, the processing of which is likely to load onto different cognitive processes, and come with different costs and benefits. For instance, strategies involving responses to eye gaze cues may provide opportunities for rapid joint attention given that eye movements themselves are swift and constantly updated during face-to-face interactions. While eye movements can be highly informative about a social interlocutor's perspective (Gobel et al., 2015), their ubiquity and use for other

non-social purposes render gaze an ambiguous joint attention signal. As such, additional social cognitive processing is likely required to infer whether an eye movement is intentional or communicative (Caruana et al., 2017a, 2020; Senju & Johnson, 2009). Pointing gestures made with the hand, however, although slower to execute than eye movements, can be more spatially precise and, in most contexts, unambiguously convey communicative intent (Horstmann & Hoffmann, 2005; Pelz et al., 2001; Yoshida & Smith, 2008; Yu & Smith, 2013). Thus, hand-following may offer a less cognitively demanding pathway to joint attention. Indeed, across two recent studies, joint attention rates were commensurate across autistic and non-autistic children when observed using head-mounted trackers during parent-child interactions that afforded hand-object interactions (Perkovich et al., 2022; Yurkovic-Harding et al., 2022). This indicates that autistic people may effectively achieve joint attention using different strategies (e.g. hand-following). However, due to limitations in available experimental methodologies, we know very little about how humans integrate and prioritise different social cues during interactions to efficiently coordinate with others.

A recent study began to address this gap by interrogating how *adult dyads* integrate and prioritise multimodal (i.e. gaze and hand) non-verbal information to achieve joint attention (Caruana et al., 2021). This study implemented a novel immersive virtual reality task, which used simultaneous motion-tracking of the eye, head and hands to facilitate richer, more naturalistic social interactions. Virtual reality supported the implementation of objective eye and body movement measures without the need for an apparatus to obscure the virtual faces or bodies of participants, or that of their partner, during the interactive task. It also enabled a high degree of experimental control over the social interaction itself.

Caruana et al. had dyads engage in a collaborative task that required them to take turns initiating and responding to joint attention bids using *explicit* hand-pointing gestures. The authors examined three key questions: (1) Do adults display useful gaze information when initiating joint attention with a hand-pointing gesture? (2) Do adults consistently look at others' gaze during joint attention interactions? and (3) Do people integrate gaze information displayed by others to facilitate their responses, even when explicitly instructed to respond to a hand-pointing gesture? Marked variation was observed in both the extent to which initiators displayed informative eye movements in the lead-up to joint attention bids and how responders attended to the face of their partner when responding to joint attention bids. In short, across this sample of non-autistic adults, eye gaze was *not* a reliable source of information for guiding joint attention, and adults did not consistently orient to gaze information during interactions to guide their responses. Nevertheless, participants were influenced by the gaze information of their partner, even when they did not explicitly use a gaze-following strategy – they were faster to respond to hand-cued joint attention bids when the pointing gesture was preceded by a congruent (i.e. predictive), rather than incongruent (i.e. non-predictive), gaze shift to the target location.

The current study

The current study builds on the work of Caruana et al. (2021), investigating whether autistic young people spontaneously use gaze-following strategies during multimodal joint attention interactions, and whether different joint attention outcomes are observed between autistic and non-autistic participants. Given reports of gaze avoidance and experiences of distress and discomfort when looking at the eyes of others (Trevisan et al., 2017), we anticipated that autistic young people would be (1) *less likely* to orient to their partner's face when responding to joint attention bids and (2) *more likely* to implement a hand-following strategy to achieve joint attention. Importantly, we did not expect either (1) or (2) to reflect poorer joint attention outcomes. Rather, we expected joint attention response times to be faster among autistic than among non-autistic participants if they adopted a gaze-avoidance strategy because, in our paradigm, eye gaze was a less reliable joint attention cue than pointing.

An additional aim was to employ a multi-layered approach to understanding joint attention strategies used by autistic people, by combining objective measures of attention and behaviour with subjective measures of participants' experiences and strategies. This approach follows the suggestion that complex cognitive phenomena – especially those that manifest in dynamic interactions – cannot be fully understood using third-person observation alone (Jack & Roepstorff, 2002; Kingstone et al., 2008). As such, we have followed the example of previous studies of joint attention (e.g. Caruana et al., 2018), which have successfully used subjective measures to supplement the interpretation of objective measures of cognitive function. Specifically, in a post-experimental interview, we asked participants about their experiences, preferences and strategies used when both initiating and responding to joint attention bids. We were particularly interested in determining whether autistic people self-reported different preferences for non-social or non-gaze strategies than their non-autistic counterparts.

Finally, our sample targeted young people (older children, adolescents and young adults) for three reasons. First, joint attention and social cognition have been historically under-investigated in this age group. Second, adolescence marks an interesting period of social cognition development associated with increased opportunities for interaction in a growing social network that extends beyond a person's home and school, including exposure to independent interactions with strangers (Brizio et al.,

Table 1. Participant characteristics.

	Autistic (n = 22)			Non-autistic (n = 22)			Statistics		
	M	SD	Range	M	SD	Range	T(df)	p	BF$_{10}$
Age (years)	14.45	3.17	9–22	14.50	3.73	10–23	t(42) = 0.04	0.965	0.298
WASI-II									
VCI	113.17	16.47	70–135	124.90	17.67	97–156	t(37) = 2.13	0.040	1.79
PRI	114.94	16.10	77–138	116.14	12.51	80–133	t(37) = 0.26	0.795	0.321
FSIQ-4	117.78	14.17	99–137	122.86	14.42	91–142	t(37) = 1.19	0.240	0.547
STAI-B									
Before interaction	11.23	2.64	6–15	10.27	2.55	6–14	t(42) = 1.22	0.229	0.541
After interaction	11.05	3.24	6–15	10.59	3.05	6–15	t(42) = 0.48	0.635	0.326
ADOS-2 (autistic)									
Communication	1.94	5.56	0–6						
Social interaction	5.56	4.30	0–14						
Communication + Social interaction	7.50	5.32	0–18						
AQ-10 (non-autistic)				3.73	2.14	0–9			

Descriptive statistics are provided for sample characteristics including age in years and scores on the following standardised measures for both groups. Autistic participants also completed the ADOS-2 (Autism Diagnostic Observation Schedule; Lord et al., 2000). ADOS-2 scores were based on Module 4 observations, except for two autistic children who were observed using Module 3. Non-autistic participants completed the AQ-10 (10-item Autism Quotient; Allison et al., 2012). WASI-II: Wechsler Abbreviated Scale of Intelligence, Second Edition (Wechsler, 2011); VCI: Verbal Comprehension Index; PRI: Perceptual Reasoning Index; FSIQ-4: Full-Scale IQ using all four subtests; STAI-B: Brief 6-item state anxiety scale. M; mean; SD: standard deviation; p: p-value.

2015). Third, from a more practical perspective, we wanted to work with autistic young people without an intellectual disability who were old enough to tolerate virtual reality, and who were able to verbally communicate their task experiences, preferences and strategies during our post-experimental interview. This was informed by previous intervention research using head-mounted virtual reality (VR) social interaction tasks with autistic children as young as 9 years of age (e.g. Ke & Im, 2013; also see Newbutt et al., 2020).

Methods

Participants

Twenty-two autistic (n = 22, M_{age} = 14.45; SD = 3.17; 6 females) and non-autistic (n = 22, M_{age} = 14.5; SD = 3.73; 9 females) young people aged between 9 and 23 years took part in the study. Data from all participants were included in our final analyses. All participants had normal or corrected-to-normal vision (i.e. with clear contact lenses only) and had no history of neurological injury or impairment, as reported by adult participants or their caregivers. All participants provided written, informed consent prior to study completion, and parents of participants younger than 18 years also provided written consent for their children to take part. No participants reported experiencing any discomfort or tolerability issues when completing the task in VR. All procedures were carried out in accordance with the protocol approved by the Macquarie University Human Research Ethics Committee. Participants were reimbursed $20/h for their participation.

All efforts were made to ensure that both groups were as closely matched as possible for gender, age and IQ (see Table 1). We administered the Wechsler Abbreviated Scale of Intelligence, Second Edition, (WASI-II; Wechsler, 2011) and report the Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI) and Full-Scale IQ using all four subtests (FSIQ-4; see Table 1). Four autistic participants and one non-autistic participant declined to complete the WASI-II assessment. Autistic participants also completed the second edition of the Autism Diagnostic Observation Schedule (ADOS-2; Lord et al., 2000), although six participants declined.[1] Non-autistic participants completed the 10-item Autism Quotient (Allison et al., 2012) with relatively low scores observed across our sample (Table 1). We also administered a brief 6-item measure of state anxiety (STAI-B; Marteau & Bekker, 1992) for all participants immediately before and after they competed the social interaction task. This was to assess whether there were any differences between groups with respect to their emotional state before and after the interaction. We found no evidence for group differences across all measures, with the exception of VCI scores, which were significantly lower among our autistic than our non-autistic participants. Specifically, we found substantial evidence for a null effect of group on both age and our non-verbal measure of intelligence based on Lee and Wagenmakers' (2014) classification scheme for interpreting Bayes Factor scores (PRI; BF$_{10}$ < 0.33).

Task materials and procedure

Task overview. The current study adapted a novel dyadic joint attention paradigm developed by Caruana et al.

(2021), implementing one key change. Rather than recruiting naïve *dyads*, autistic and non-autistic participants were individually recruited to interact with a confederate, which offered a greater level of experimental control and minimised task-related differences across participants and between groups. Two confederates were used across the study, one male and one female, to sample a mix of male–male (autism: *n* = 9; non-autistic: *n* = 10), female–female (autism: *n* = 0; non-autistic: *n* = 2) and male–female interactions (autism: *n* = 13; non-autistic: *n* = 10). Confederates were visually cued in real time during the virtual interaction to precisely control the location and sequence of their eye movements, both within and across trials. Both confederates underwent extensive training to ensure they were accurately and reliably following these prompts (see *Online programming of confederate behaviour*). All other aspects of the task design and implementation, including stimulus features and apparatus, were identical to those of Caruana et al. (2021).

Task setup. After providing written consent, the experimenter began by explaining the apparatus and task to participants. To ensure that they engaged in the task as naturally as possible, specific instructions on *how* to interact with each other were not provided, and care was taken not to mention gaze as a focus of investigation. A slide deck with simple visual and text prompts was used to help standardise the delivery of task instructions (available on the Open Science Framework [OSF] project page https://osf.io/pkx9f/?view_only=f0011eabdb1146aaa09ce738db90804). These instructions depicted the virtual environment and provided an overview of two tasks that participants would complete. Participants were told that, in the first task, they would complete a task on their own that required them to search for a target among three cubes and point to the target cube as quickly and accurately as possible. In the second task, participants were instructed that they would complete a similar task together with the confederate. Instructions for both tasks were provided at the same time to avoid disrupting the participants' immersion in the virtual environment once the task began and the apparatus calibrated.

Participants were then introduced to the confederate. Participants were not deceived about the confederate's identity, including whether they were autistic or not, and were truthfully told that the confederate was a member of the research team conducting the study. The experimenter then attached a motion sensor to each participant's right index finger using masking tape and secured the virtual reality head-mounted display onto each participant's head, which was adjusted to ensure visual acuity, optimal ocular convergence and comfort. Participants were sent a study information sheet prior to their participation in the study that described the apparatus and setup procedure with images, so that they knew what to expect during the study session.
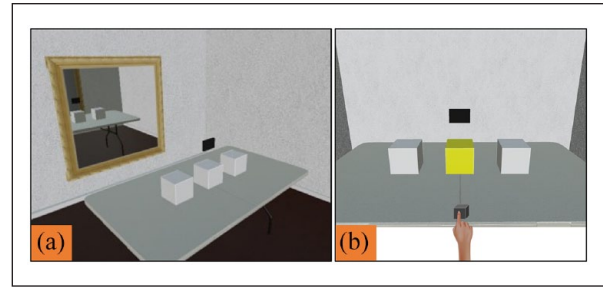

Figure 1. (a) Calibration environment and (b) non-social task stimuli.
Note. Figure 1(b) depicts all stimulus components used during the non-social block; however, during the task, the black rectangle disappeared when the trial began.

Based on the demographic information provided prior to the experiment, participants were assigned an avatar that matched their self-declared gender and were immersed into the virtual space. The experimenter then calibrated the eye-tracking device and centred the participant's position within their avatar and the virtual room. This allowed participants to see the virtual environment through their avatar's perspective and congruently move the avatar's body using their own natural body movements (eyes, head, hands). Following calibration, a virtual mirror appeared on the wall opposite the participant (Figure 1(a)). This provided participants with an opportunity to see their own avatar and the visual-motor congruity between their physical and virtual body movements. This promoted subjective embodiment and demonstrated the interface's capacity to accurately record and display their body movements in real time. The same calibration procedure was conducted for the confederate.

Non-social task. Following the calibration, the mirror was removed, and a row of three pale grey cubes (10 cm³ each) appeared on the table in front of participants. At the beginning of each trial, a small black cube appeared on the table in front of participants, and a black rectangle (15 × 10 cm) appeared above the three cubes (see Figure 1(b)). Participants were required to place their finger in the small cube and fixate on the rectangle to begin the trial. This served to standardise participants' eye, head and hand positions at the beginning of each trial and also provided an ongoing check for the eye- and motion-tracking calibration. The fixation rectangle position was selected to be analogous to the location of the confederate's avatar face on joint attention trials so that the starting positions were also matched across the non-social and joint attention tasks.

Participants were told that the small cube would turn green to indicate that their finger was correctly placed and the black rectangle would disappear once fixated, indicating the start of the trial. Then, one of the three large cubes would turn yellow. Participants were required to point to
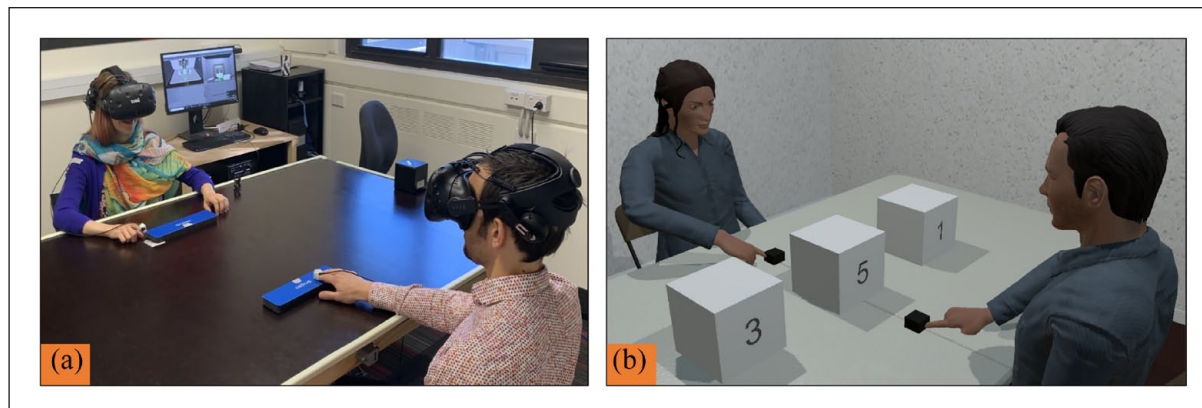
Figure 2. Dyads interacting in the (a) physical and (b) virtual laboratory.

the yellow cube as quickly and accurately as possible. Participants completed 27 consecutive trials (approximately 15–20 min in total), with the target location randomised and counterbalanced across trials. This task assessed whether there were any initial differences between autistic and non-autistic people when orienting to visual attention cues within the virtual environment. To this end, we measured and compared response times using both point reaction times (PRTs) and saccadic reaction times (SRTs; see *Responding behaviour* below).

Joint attention task. Upon completing the non-social task, participants were given an opportunity to rest their eyes for up to 5 min before completing the social joint attention task which comprised 72 trials (approximately 25–30 min in total) divided into two blocks of 36 trials to provide participants with a short break between blocks. During this task, the participant and confederate were immersed into the same virtual space, each represented by a gender-congruent avatar. The pair sat opposite each other at both the physical (Figure 2(a)) and virtual table (see Figure 2(b)).

Each trial began once both the participant and confederate had placed their right index finger into the starting cube and fixated on their partner's face (analogous to the black rectangle in the non-social task). This served to standardise initial gaze and finger positions and ensured that the participant and confederate were ready to commence each trial. Once both players were ready, a number between 1 and 6 appeared on each of the three cubes facing the participant. Participants were told that on trials where they saw three even numbers (i.e. 2, 4 and 6), their job was to initiate joint attention by pointing to the cube with the highest value number. If, however, they were presented with three odd numbers (i.e. 1, 3 and 5) or a combination of even or odd numbers (e.g. 1, 4 and 6), they were required to wait for their partner to initiate joint attention and respond accordingly by also pointing to the correct cube. Corrective feedback was provided on each trial with the target cube turning green or red for correct and incorrect

responses, respectively. Critically, participants could only see the information presented on their side of the cubes. This created a task context in which the dyad was required to simultaneously search among the three cubes to determine both their social role, as either the initiator or responder of the joint attention bid, as well as the target's location. Importantly, this also created an ecologically valid and dynamic scenario in which participants both displayed and were able to observe a range of eye movements in the lead-up to joint attention opportunities. To establish this reciprocal task context, it was essential that participants completed a combination of initiating and responding trials. In total, participants completed 72 joint attention trials (36 as initiator, 36 as responder). Trials were presented in randomised orders, with target location, cube number combinations and number location counterbalanced across trials (see Figure 3 for trial sequence).

Online programming of confederate behaviour. To standardise the perceptual information presented to participants on responding trials, we carefully curated the order in which the confederate searched the cubes and whether their final gaze shift was congruent with (i.e. predictive of) the subsequent pointing behaviour which signalled the location of the target. To achieve this, the confederate was visually cued within the virtual environment as to the order in which they should look at each cube during their search and where to hold their gaze while pointing to initiate joint attention. The confederate was cued to display congruent gaze-point behaviour on 50% of trials and incongruent gaze-point behaviour on the remaining 50% of trials. See Supplemental Material 1 for a detailed description of how the confederate's behaviour was programmed online, as well as how the confederate behaviour and the congruency manipulation were validated.

Stimulus. The immersive virtual reality task was developed using the Unity3D Game Engine (Version 2017.4.19f; San Francisco, CA, USA; see link for code https://github.
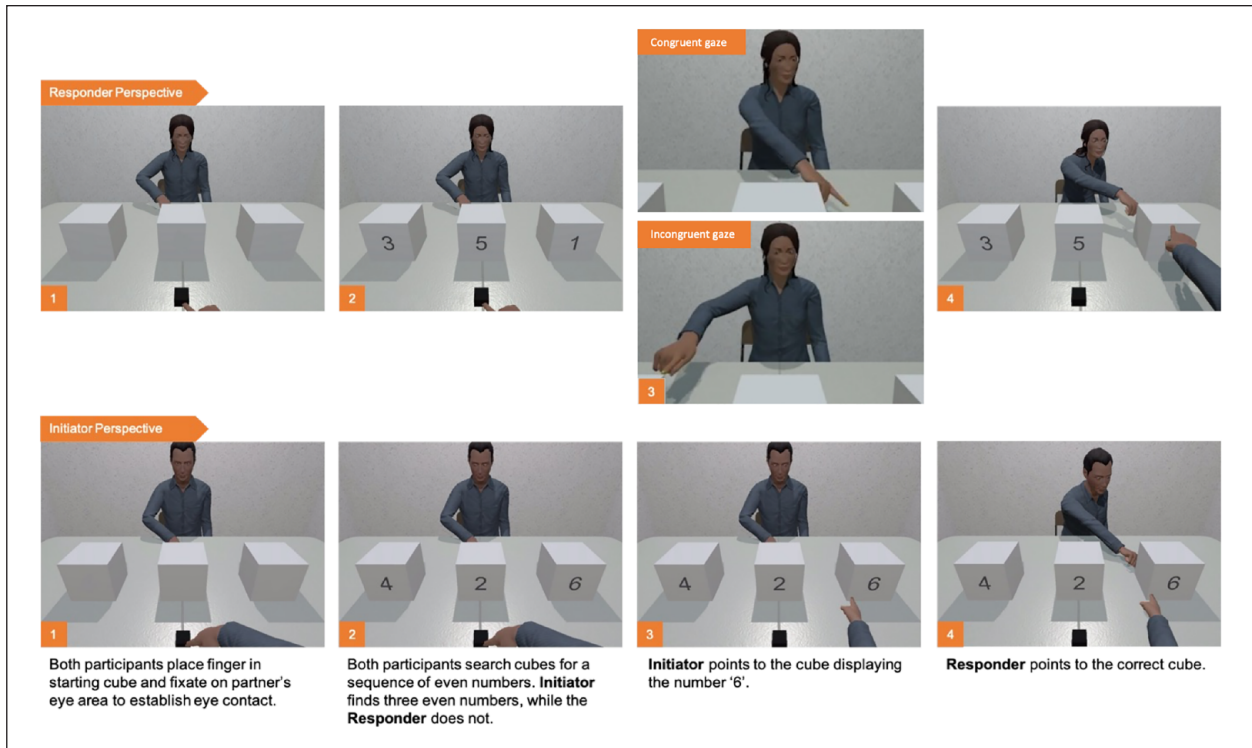
Figure 3. Trial sequence, as seen from the participants' perspective, as the responder or initiator of joint attention on each trial. Examples of congruent and incongruent gaze-point behaviour are depicted for responder trials in the top panel.

com/ShortFox/Autistic-People-Multigestural-Interactions). We used the exact same stimuli (i.e. virtual environment and avatars) as those reported by Caruana et al. (2021). The virtual room and its contents matched the location and size of their physical counterparts (room: 4.9 m width × 3 m depth; table: 1.5 m width × 1.15 m height × 0.98 m depth; chair seat height: 50 cm). A purpose-built wooden table and polypropylene resin chairs were used to avoid metallic interference with motion-tracking devices. Two anthropomorphic and ethnically ambiguous avatars, one male and one female, were used to represent male and female participants, respectively (see Caruana et al., 2021, for details on avatar generation and validation).

Analysis areas of interest. Analysis-related areas of interest (AOIs) were defined within the virtual environment to record when a participant viewed task-relevant stimuli (i.e. avatar gaze, hand, body, cubes and other). A look to the confederate's face was recorded when a participant's gaze intersected the 'face' AOI, defined as a spherical collider placed centrally on the avatar's head (sphere radius = 12 cm). The 'hand' AOI (10 × 20 cm), began at the avatar's right wrist and extended to the index finger. The confederate 'body' AOI began at the avatar's hips, spanned the width of the body and ended at the base of the neck (40 × 60 cm). An AOI was also defined for each of the three task-relevant cubes, which aligned with the parameters of each cube (10 cm³).

Apparatus. The current study used the exact apparatus and hardware described by Caruana et al. (2021). A HTC Vive head-mounted display (HTC Corporation, Taiwan; Valve Corporation, Bellevue, Washington, USA) with retrofitted Tobii Pro eye-trackers (Tobii Pro Inc., Sweden) were used to immerse participants in the virtual environment and record participant's head position/orientation and eye movements (i.e. saccades, fixations and blinks). These measures were not only critical for our objective analysis of social attention and coordination but also enabled participants to control the dynamics of their respective avatar using natural body movements. To this end, we also used a Polhemus G4 tracking system (Polhemus Inc, Vermont, USA) to track the movements of each participant's right index finger using a single sensor on the finger. An inverse kinematics calculator (RootMotion Inc., Estonia) was used to generate the remaining full-arm and upper-torso movement (Lamb et al., 2019, for a validation of this method; see Nalepka et al., 2019). The maximum display latency between the two participant/confederate's real-world movements and their movements in the virtual environment was 33 ms. See Caruana et al. (2021) for a more detailed account of stimulus and apparatus development and validation for this paradigm.

Eye- and hand-tracking measures

The current study used the same definitions for operationalising behaviour on initiating and responding trials as reported by Caruana and colleagues (2021).

Initiator behaviour. The initiation of a joint attention bid was operationalised as the confederate/participant moving their arm to point at the target location. We defined the onset of the pointing action as the point in time when the speed of the individual's arm movement towards the target reached 5% of the maximum speed reached during the trial. This 5% velocity threshold was modelled on previous approaches to exclude uninformative movements towards non-targeted objects, while including the earliest time that directional information could be conveyed by the movement (see Domkin et al., 2002, for a validation of this method). For initiating trials, we also conducted exploratory analyses to evaluate the frequency (across trials) and duration (within trials) of face-looking behaviour (i.e. overt fixations to the confederate's avatar face). Face-looking times were separately examined: (1) across the entire trial; (2) before the confederate responded to their joint attention bid and (3) in the period after the confederate responded to the participant's joint attention bid. For these analyses, we excluded trials where viewing times were less than 150 ms as these were unlikely to represent periods of meaningful face-processing.

Responding behaviour. SRT measures for responding to joint attention were calculated as the latency between the onset of the confederate's pointing action (as defined earlier) and the onset of the participant's subsequent saccade towards the target. Saccade onset, rather than fixation time on the cube, was used to eliminate variability in response times due to differences attributed to the distance of the target cube. Following the study by Caruana et al. (2020), SRTs shorter than 150 ms were excluded from analyses as they were likely to be anticipatory responses or false starts. PRT was also calculated using the same principles. The onset of the participant's point response was defined in the same way as pointing actions for initiating joint attention; the time at which the hand movement reached 5% of its maximum velocity. Our analyses primarily focused on SRT rather than PRT measures given that eye movement responses are likely to provide a more direct and sensitive measure of attention. As such, SRT analyses are reported below (see *Saccadic reaction times*) with the analogous PRT analyses included our RMarkdown documentation on the project's OSF project page (see the aforementioned link). Similar to the study by Caruana et al. (2021), we also measured the frequency with which participants overtly fixated the face of their partner while their partner was initiating a joint attention bid to the target.

Data and statistical analysis

Pre-processing data. The position and timing of the dyad's eye and hand and head movements were continuously recorded at 60 Hz. Interest area output and trial data were exported from Unity3D using a custom Matlab script (MATLAB R2017b; see link for code https://github.com/ShortFox/Gaze-Responsivity-Hand-Joint-Attention) and then screened and analysed in R 3.6.1 using a custom RMarkdown script (R Core Team, 2019-07-05; see RMarkdown document for a detailed description of each processing step, and all analysis output). No participants were removed from any analyses. However, individual trials were removed due to poor calibration (i.e. trials where the eye calibration fell below 95% accuracy; autistic: $M = 21.35$ trials, $SD = 30.45$; non-autistic: $M = 7.72$, $SD = 11.35$). Trials where participants responded incorrectly (see *Accuracy* analysis below) were also excluded from reaction time analyses. Moreover, trials in which responder SRTs were longer than 3000 ms were excluded, as they were unlikely to reflect an immediate response to the joint attention bid (Caruana et al., 2015, for a complementary approach, see 2020). A box-cox transformation (see Balota et al., 2013) was performed separately on SRT data and PRT data to determine objectively the appropriate transformation needed to meet the assumption of normally distributed residuals, which is critical for linear mixed effects (LME) modelling. The box-cox function implemented in our R code determined an appropriate value of λ, which was then used to transform the data (Box & Cox, 1964).

LME analysis. The effects of participant group and confederate initiator gaze-point congruency (for responding trials) were statistically analysed by estimating LME models, using the maximum likelihood estimation method within the *lme4* R package (Bates, 2005). LME models were employed instead of a traditional analysis of variance (i.e. ANOVA) to account for subject and trial-level variance (i.e. random effects) when estimating the fixed effect parameters (i.e. group, gaze-point congruency; see Caruana et al., 2021, for a detailed justification for the application of LME analyses in studies of dyadic interactions). We adopted a 'maximal' random factor structure, with random intercepts for the subject and trial (Barr et al., 2013). Specifically, we originally defined a saturated model including random slopes for the trial by subject factors. However, the 'maximum likelihood' of this complex model could not be estimated given the available data and 'failed to converge' (see Barr et al., 2013, p. 10 for an explanation). Therefore, we simplified our random-effects parameters to define the most saturated yet parsimonious model. The simplified model included random intercepts only for the trial and subject factors – again consistent with the approach taken by Caruana et al.

(2021). p Values were estimated using the *lmerTest* package (Kuznetsova et al., 2017), and a significance criterion of α < 0.05 was employed.

Community involvement statement

Community members were not involved in the design or implementation of the study or in the analysis or interpretation of the results.

Results

Gaze effects when responding to joint attention bids

Accuracy. High accuracy rates were observed across both autistic (M=94.54%, SD=11.84) and non-autistic participants (M=98.39%, SD=1.95), with no evidence for significant group differences (b=0.953, SE=0.624, z=1.527, p=0.127), across congruency conditions (b=0.538, SE=0.466, z=1.155, p=0.248) or associated interaction effects (b=0.944, SE=0.879, z=1.074, p=0.283).

Face-looking frequency. Compared to findings reported by Caruana et al. (2021), participants engaged in relatively more frequent instances of face-looking behaviour across trials when responding to joint attention bids, within both the autistic (M=73.09%, SD=29.80) and non-autistic (M=89.88%, SD=11.43) groups. Furthermore, the distribution of face-looking tendencies appeared to be commensurate across the two groups (see RMarkdown), with no evidence of significant group differences (b=0.645, SE=0.453, z=1.421, p=0.155).

Saccadic reaction times. We replicated the gaze-point congruency effect reported by Caruana et al. (2021) in which participants from both groups were significantly faster to respond to point-cued joint attention bids that followed congruent gaze shifts towards the target location than incongruent gaze shifts towards another location (see Figure 4). This congruency effect was significant when we modelled all responding trials (b=0.295, SE=0.106, t=2.78, p=0.008; Figure 4(a) and (b)) and when we separately modelled trials in which participants did (b=0.262, SE=0.111, t=2.36, p=0.022; Figure 4(c) and (d)) and did not (b=0.557, SE=0.216, t=2.58, p=0.013; Figure 4(e) and (f)) overtly attend to their partner's face.[2] The same pattern of findings was observed in the PRT data (see RMarkdown for complete analyses, including a comprehensive summary of descriptive statistics). We also found a significant group effect on SRT responses when we separately modelled trials where participants did *not* attend to their partner's face when responding (b=0.438, SE=0.184, t=2.39, p=0.025). In this instance, autistic participants showed faster overall response times than non-autistic participants (Figure 4(e) and (f)). However, this effect was not
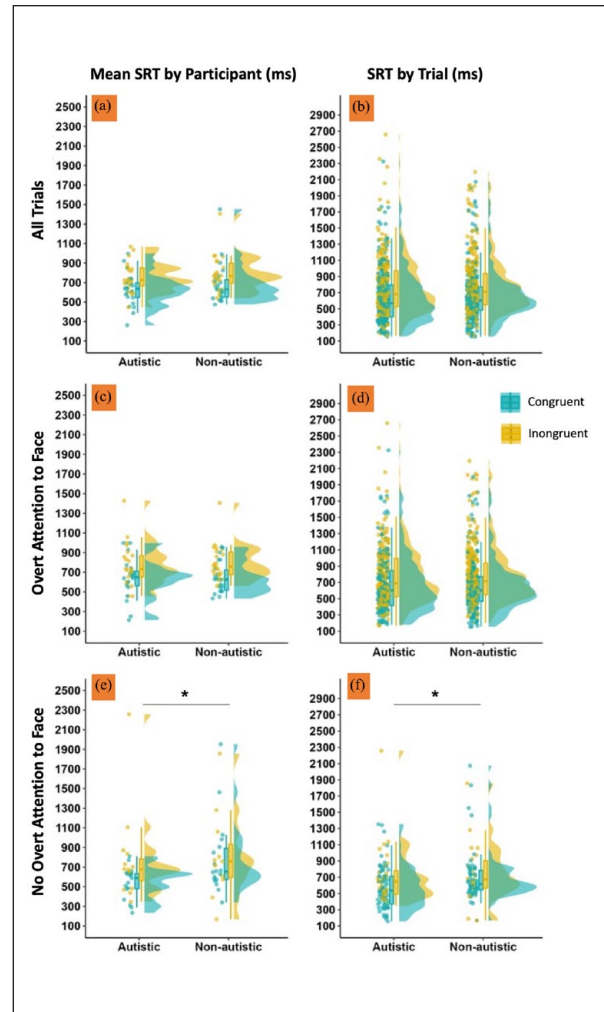

Figure 4. SRT (in milliseconds) when responding to joint attention bids is visualised separately to depict the data included in each LME model. (a and b) SRT data across all trials, irrespective of whether participants overtly attended to their partner's face. (c and d) Summarise trials in which participants overtly attended to their partner's face. (e and f) Trials in which participants did *not* overtly attend to their partner's face. Data have been visualised twice to fully represent the variation observed in our sample. (a, c and e) (left column) The variation across participants; data points depict mean SRT times for each participant, aggregated across trials. (b, d and f) (right column) Unaggregated trial SRTs. LME: linear mixed effects; SRT: saccadic reaction time.

significant in the PRT data. Finally, we found no evidence for any other significant group or group by congruency interaction effects in any of the planned SRT or PRT models (all ps > 0.097).

Gaze effects when initiating joint attention bids

We conducted a series of exploratory analyses to determine whether there were any significant group differences in face-looking behaviour when initiating joint attention.

Accuracy. Ceiling effects for accuracy were again observed across both autistic ($M=99.01\%$, $SD=3.25$) and non-autistic participants ($M=99.85\%$, $SD=0.71$), with all but three participants (two autistic, one non-autistic) attaining perfect accuracy scores.

Face-looking frequency. First, we examined the *frequency* of face-looking behaviour on initiating trials. We found no significant differences between autistic ($M=65.21\%$, $SD=22.75$) and non-autistic participants ($M=74.40\%$, $SD=16.57$; $b=0.414$, $SE=0.310$, $z=1.33$, $p=0.183$).

Face-looking time. We also examined the *duration* of face-looking across initiating trials. There was no evidence for any significant group differences in the time spent looking at the confederate's virtual face across the whole trial ($M_{autistic}=1173.69$ ms, $SD_{autistic}=978.10$, $M_{non-autistic}=1094.54$ ms, $SD_{non-autistic}=823.26$; $b=0.032$, $SE=0.078$, $t=0.408$, $p=0.687$) and when separately examining the interval *before* the confederate responded to the participant's joint attention bid ($M_{autistic}=1162.34$ ms, $SD_{autistic}=905.90$, $M_{non-autistic}=962.96$ ms, $SD_{non-autistic}=494.10$; $b=0.013$, $SE=0.086$, $t=0.156$, $p=0.878$). However, there was a significant group effect for face-looking times *after* the confederate's response ($b=0.014$, $SE=0.006$, $t=2.47$, $p=0.031$), characterised by longer face-looking times by non-autistic ($M=702.59$ ms, $SD=770.32$) than autistic participants ($M=313.29$ ms, $SD=117.39$). However, this effect should be interpreted with caution given it is likely biased by two non-autistic participants, who showed exceptionally long face-looking times. Furthermore, it is worth noting that when looking at face-looking durations *before* the confederate responded, the four participants who exhibited the longest face-looking times were all autistic (see Figure 5(c); a full summary of descriptive statistics reported in the RMarkdown). This goes against the suggestion that gaze avoidance is a characteristic behaviour among autistic people.

Non-social attention task

Accuracy. High accuracy rates were observed across both autistic ($M=95.68\%$, $SD=8.06$) and non-autistic ($M=98.75\%$, $SD=2.25$) participants. Our LME model indicated evidence for a significant group effect ($b=1.22$, $SE=0.588$, $z=2.07$, $p=0.038$), characterised by marginally higher error rates in the autistic group. However, this should be interpreted with caution given the marked ceiling effects, with all participants attaining accuracy scores on the non-social task above 90%.

Saccadic reaction times. We found no evidence for significant group differences in SRT responses when completing the non-social attention task within the same virtual environment ($b=0.0006$, $SE=0.002$, $t=0.286$, $p=0.777$; see
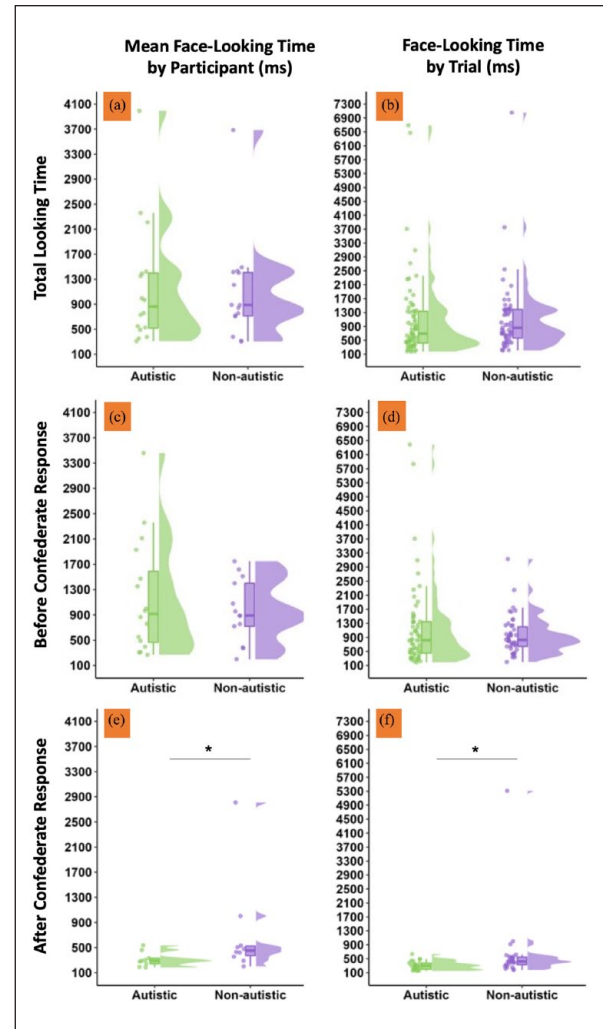

Figure 5. Face-looking times (in milliseconds) when initiating joint attention bids is visualised separately to depict the data included in each LME model. (a and b) Total face-looking times across the full trial. (c and d) Face-looking times *before* the confederate responded to the joint attention bid. (e and f) Face-looking times *after* the confederate responded to the joint attention bid. Data have been visualised twice to fully represent the variation observed in our sample. (a, c and e) (left column) The variation across participants; data points depict mean face-looking times for each participant, aggregated across trials. (b, d and f) (right column) Face-looking times across unaggregated trials. LME: linear mixed effects.

Figure 6). This was also the case when analysing PRT response measures ($b=0.913$, $SE=0.809$, $t=1.13$, $p=0.265$; see RMarkdown for full analyses and descriptive statistics on our OSF project page: https://osf.io/pkx9f/?view_only=f0011eabdb1146aaa09ce738db908046).

Subjective experiences

Subjective experience ratings. After completing the interactive task, participants were asked to rate their experience
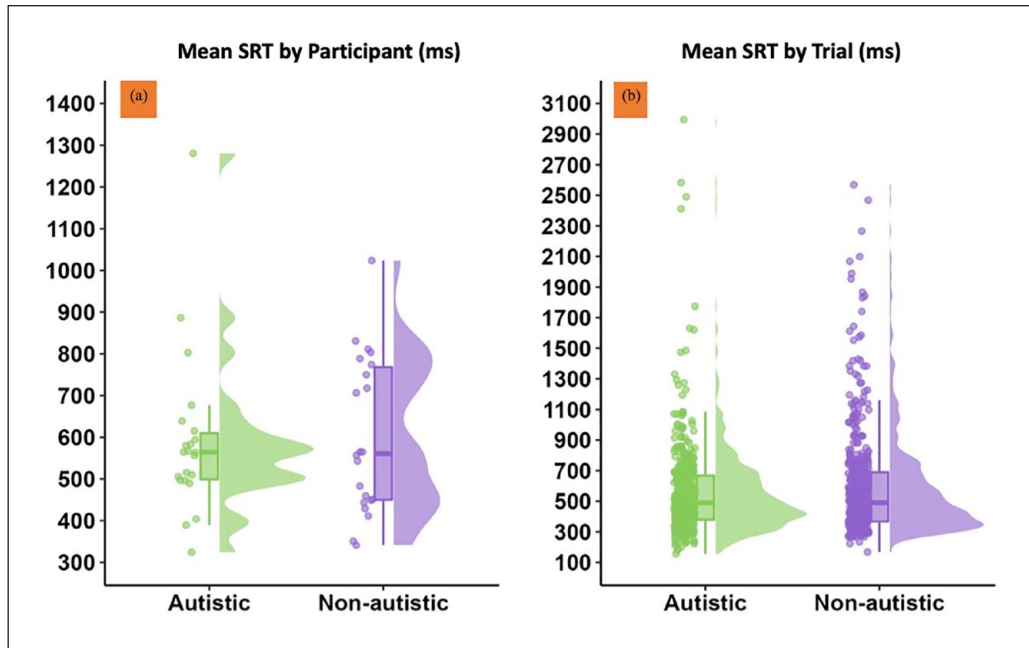
Figure 6. SRT (in milliseconds) when responding to non-social cues within the non-social task. (a) The variation across participants; data points depict mean SRT times for each participant, aggregated across trials. (b) Unaggregated trial SRTs. SRT: saccadic reaction time.

during the task, specifically on how *difficult*, *natural* and *pleasant* the interaction was, and how *helpful* they found their partner on a scale ranging from '1' (not at all) to '5' (extremely). For the most part, participants in both groups provided positive ratings of task experience, with no significant group differences across rating domains (all ps > 0.202). See the RMarkdown for a visualisation of subjective rating data by group (https://osf.io/pkx9f/?view_only=f0011eabdb1146aaa09ce738db908046).

Subjective reports of coordination strategies. During the post-experimental interview, we asked participants if they used any explicit strategies when initiating or responding to joint attention bids. A detailed report of our findings can be found in Supplemental Material 2. In sum, subjective reports of strategy use were strikingly similar across groups. For both groups, the use of an explicit gaze-based strategy was rare. In fact, no non-autistic participants reported using gaze-following strategies when responding to joint attention bids, whereas two autistic participants said they used this strategy explicitly. We also asked participants if they explicitly preferred to follow their partner's hands or eyes when responding to their joint attention bids. We observed a near-identical profile of preferences across groups, with an overwhelming preference across participants to use hand-based strategies for coordination.

Discussion

Existing experimental and objective studies of gaze processing and joint attention have not clearly elucidated whether or why social interactions can be challenging for autistic people, largely for methodological reasons. Specifically, they have not successfully captured the full complexity of social interactions that emerge from ordinary two-way exchanges between people. Previous studies reporting joint attention difficulties in autistic people used tasks that only allow participants to communicate via eye gaze. This might unfairly set up autistic people to appear 'impaired' when responding to joint attention given their self-reported experiences of discomfort when looking at the eyes of others (Trevisan et al., 2017). Identifying which non-verbal behavioural cues humans rely on to achieve joint attention in realistic and dynamic interactions is important because this affects our understanding of both the *cognitive* mechanisms involved during joint attention (Siposova & Carpenter, 2019) and the conditions that make social interactions difficult (or not) for autistic people.

In this study, we showed that autistic young people can and do adaptively attend to the gaze information displayed by others during coordinated social interactions, even when doing so is not a task requirement. Contrary to our expectations, both autistic and non-autistic participants exhibited largely commensurate face-looking behaviours towards their partner and were significantly faster to respond to pointing gestures that followed congruent, rather than incongruent, gaze shifts to the target location. This is particularly striking given that social coordination in our task did not explicitly require participants to look at their partner's face or use gaze information at all to effectively respond to hand-cued joint attention bids. It is

possible, however, that participants were sensitive to course spatial information in their peripheral vision conveyed by their partner's head and body movements, even when not looking directly at their partner's face. Such information, however crude, may act as a proxy signal for another's gaze. In any case, our data suggest that participants from both groups were sensitive to such information. This again highlights the importance of measuring social information processing using paradigms that involve realistic depictions of dynamic social information displays.

Furthermore, we anticipated that both subjective and objective measures would reveal stronger preferences for hand-following strategies among autistic than among non-autistic young people. However, both groups exhibited similar and overwhelming preferences for adopting an explicit hand-following strategy, with some participants from both groups also acknowledging (correctly) that a gaze-based strategy in this task did not appear to be optimal (see Supplemental Material 2 for an in-depth analysis). As such, this study provides evidence that both autistic and non-autistic people have the capacity to integrate gaze information to facilitate joint attention and are sensitive to when such a strategy may be more or less effective given the reliability of their social partner's gaze over time. This aligns with recent evidence from a concurrent, pre-registered study run in our lab, showing that non-autistic people adaptively abandon a gaze-following strategy when their partner's gaze is more incongruent with subsequent pointing behaviours than it is congruent (https://osf.io/d8b5t/?view_only=e91f29294fc14a06b27f3522c67fb 03a).

Our findings are at odds with earlier work examining gaze-based joint attention in autistic adults (Caruana et al., 2018). In this previous study, autistic adults were initially slower than non-autistic adults when responding to gaze-cued joint attention bids embedded in dynamic interactions where their social partner displayed both non-communicative and communicative gaze shifts. However, group differences on this social task diminished over time with an increase in task exposure and were not observed in a closely-matched non-social version of the task involving dynamic arrow stimuli. This was interpreted as evidence for a specific social difficulty in using ostensive eye contact signals to differentiate between communicative and non-communicative gaze signals. This was also corroborated by the subjective experiences of autistic participants who explained that they found eye gaze to be an 'intense' stimulus, which was often too difficult to interpret. Others commented on the 'intensity' of eye gaze, which led to feelings of discomfort and stress. These experiences are very much corroborated by other studies examining the subjective experiences of autistic people during everyday gaze-based interactions (Trevisan et al., 2017). It is possible that studies of gaze processing and joint attention that force participants to attend to, process and respond to eye

gaze stimuli might indirectly produce behavioural differences that resemble 'deficits' because they create an unrealistic social scenario of sustained gaze-based communication; a scenario that people almost never encounter in everyday situations, is known to be psychologically difficult for many autistic people and ignores emerging evidence that hand-following can provide a more accessible pathway to joint attention than gaze-following (Deák et al., 2014; Yu & Smith, 2013). In contrast, the interactive paradigm employed in the current study allowed participants to naturally signal, attend to and use gaze information without making this communicative approach mandatory. Indeed, we present evidence that our autistic participants were significantly faster than non-autistic participants (and thus presumably more effective) in responding to joint attention bids when we compared groups on trials where participants did not look at their partner's face during the joint attention episode. This offers further evidence that joint attention performance in autistic people may improve when it is operationalised beyond tasks that require the use of gaze. Accordingly, joint attention measures that impose gaze use may artificially conflate gaze processing with joint attention ability.

Adding to the aforementioned findings, we engaged in an exploratory examination of face-looking behaviour when initiating joint attention to determine whether there was any evidence for differences in face-looking behaviours across these different social role contexts. This analysis revealed that the top three participants who spent the most time looking at their partner's face were all autistic, with otherwise relatively commensurate levels of face-looking observed across the two groups. The current findings are particularly compelling given that, in this experimental task, joint attention can be effectively achieved without looking at the face of one's social partner, and so such behaviours are voluntary. These findings are at odds with past studies of joint attention and gaze-processing in autism, which typically force participants to attend to, decipher and respond to gaze cues displayed by either interactive or non-interactive social partners (see Caruana et al., 2017b; Mundy, 2018, for reviews). In this way, the current study's paradigm – although arguably much more ecologically-valid, dyadic and complex in its use of multiple gestures – provided a less-demanding context for social interaction for autistic people. This in turn may have made it easier for participants to intermittently gather and respond to gaze information during the interaction.

It is also possible, however, that the gaze information conveyed by animated avatars is less perceptually overwhelming for our autistic participants than physically embodied face-to-face interactions. This possibility is consistent with recent findings in which autistic children demonstrated better emotion categorisation accuracy when viewing animated avatar faces than photographs of real faces (Pino et al., 2021). As such, whether these findings

generalise to embodied face-to-face interactions requires further scrutiny. However, the challenge of examining such interactions outside virtual reality is in the loss of control over environmental and interpersonal factors, and in gathering objective eye movement measures without obstructing the faces of interlocutors with tracking equipment. Finally, the participants in the current study were all of at least average intellectual ability. As such, more research is needed to investigate coordination strategies using this paradigm across a more diverse autistic sample.

Rethinking social information processing in autistic people

Our findings are important for reshaping how we conceptualise and understand social differences or 'deficits' in autism. Historically, autism has been conceptualised under a medical model that characterises differences in social and communicative behaviours as 'deficits'. However, growing empirical research, and testimony from autistic adults themselves, has highlighted that these conceptualizations of social ability in autism can be both inaccurate and harmful (Jaswal & Akhtar, 2019; Pellicano & Houting, 2022). An alternative view suggests that the social interaction difficulties described by autistic people might be better characterised as a mismatch of interaction capabilities, preferences or styles, rather than autism-specific deficits (Davis & Crompton, 2021). This does not disregard the possibility that autistic people can sometimes experience difficulties with communication – or engage in different communication strategies compared with many non-autistic people – but rather challenges the notion that the communication 'problem' lies solely with the autistic person. Specifically, the double empathy problem framework proposes that reciprocal interactions can breakdown due to a lack of mutual understanding when people process sensory (and social) information differently (Milton, 2012; Milton et al., 2018). Building on Milton's ideas, the dialogical misattunement hypothesis dismisses the idea that social difficulties in autism stem from 'disordered' autistic brains, but rather from a cumulative mismatch in the interpersonal dynamics between two people (Bolis et al., 2017).

Despite the relative nascency of these interpretative frameworks, they have already inspired fruitful research programmes promoting inclusivity and reducing stigma and prejudice (Sasson & Morrison, 2019; Scheerer et al., 2022) to ultimately improve social outcomes for autistic people by improving how non-autistic people approach interactions with autistic people (see Davis & Crompton, 2021, for review). Bolis et al. further argued that two-person experiments are key to understanding sensory processes and interpersonal coordination in autism. The current study meets this challenge and establishes a new paradigm for conducting this research in a way that helps characterise how multi-gestural interactions are navigated

by autistic and non-autistic dyads. For instance, this paradigm can also be used to examine whether, in line with the double empathy problem, non-autistic people adapt their signalling behaviours differentially when knowingly interacting with an autistic or non-autistic partner. Consistent with these accounts, our findings suggest that autistic people are likely to coordinate effectively with others when they have access to social signals that match their social information processing preferences.

This paradigm also presents unique opportunities for future work which carefully probes other factors that likely shape the adoption of adaptive coordination strategies during interactions among both autistic and non-autistic people. For instance, the reliability of gaze and hand information can be manipulated over time to probe the extent to which autistic and non-autistic individuals adapt their behaviours based on the certainty of social information. Predictive coding accounts of social information processing in autism suggest that there is a low-level difference in how autistic people predict sensory events in their environment depending on their volatility compared to non-autistic people (Lawson et al., 2014; Pellicano & Burr, 2012; Van De Cruys et al., 2013, 2014). These low-level differences in processing information may then result in higher-level mismatches in social interaction between autistic and non-autistic people (Lawson et al., 2017). Autistic people may be hyper-sensitive to uncertain social information, and adaptations of the current paradigm can be used to test whether they can adapt their social information processing strategies (i.e. gaze use) depending on the reliability of another's gaze displays over time. The application of this novel paradigm for testing social predictive coding accounts will be invaluable for informing a new model of social responsivity and information processing in autistic people.

Rethinking the importance of eye gaze during coordinated interactions

Eye gaze has long been considered among the most important social signals humans use to understand and coordinate with each other during social interactions (Cañigueral & Hamilton, 2019; Conty et al., 2016; Gobel et al., 2015; Senju & Johnson, 2009; Tomasello, 1999). However, the current study reveals that by capturing the multimodal dynamics of joint attention interactions – and moving away from conventional gaze-only paradigms of joint attention – eye gaze may play a less-privileged role in social coordination outcomes for both autistic and non-autistic people. Specifically, we show that although eye gaze can be used to facilitate responsivity during joint attention, it may be less important for achieving joint attention when other communicative gestures (e.g. hand-pointing) are available. Furthermore, our study shows that not only do autistic people seem able to use gaze

information during interactions, gaze-following strategies in this context do not appear to be the 'preferred' or 'default' strategy engaged by non-autistic people. This finding also aligns with the suggestion of Yu and Smith (2013) that hand-following may offer a less cognitively demanding strategy for coordination than gaze-following. This is because situationally relevant hand gestures can be less volatile and dynamic than gaze and also provide less ambiguous spatial and communicative signals (Horstmann & Hoffmann, 2005; Pelz et al., 2001; Yoshida & Smith, 2008; Yu & Smith, 2013). Consequently, hand-following strategies may be less likely to engage the higher-level socio-cognitive mechanisms that are needed to make inferences about the communicative significance and spatial relevance of eye movements (Senju & Johnson, 2009).

Conclusion

We found that our autistic and non-autistic participants alike intuitively and voluntarily used gaze information to facilitate coordination, even when more salient hand-pointing gestures were available to them. Furthermore, many autistic participants were able to recognise if their partner was providing useful gaze information over time that could be used to inform their coordination strategies. Finally, we showed that autistic and non-autistic people were just as likely to voluntarily look at the partner's gaze when responding to joint attention. Critical to these new insights was the implementation of an interactive task that did not force participants to continuously look and exclusively use eye gaze cues to achieve joint attention. Taken together, this study powerfully challenges a large body of research suggesting social information processing deficits in autism. Correcting this bias in how we experimentally test and conceptualise social ability in autism is critically important for addressing the stigmatisation of autism – and the poor social outcomes that likely result from this stigma (Davis & Crompton, 2021).

Author contributions

Nathan Caruana: Conceptualisation, Methodology, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualisation, Supervision, Project administration, Funding acquisition. **Patrick Nalepka:** Methodology, Software, Data curation, Writing – review & editing, Funding acquisition. **Glicyr A Perez:** Investigation, Formal analysis, Visualisation. **Christine Inkley:** Investigation, Project administration. **Courtney Munro:** Investigation, Project administration. **Hannah Rapaport:** Investigation, Writing – review & editing. **Simon Brett:** Investigation, Writing – review & editing. **David M Kaplan:** Conceptualisation, Writing – review & editing, Funding acuisition. **Michael J Richardson:** Conceptualisation, Writing – review & editing, Resources, Funding acuisition. **Elisabeth Pellicano:** Conceptualisation, Methodology, Resources, Writing – review & editing, Funding acuisition.

ORCID iDs

Nathan Caruana [iD] https://orcid.org/0000-0002-9676-814X
Elizabeth Pellicano [iD] https://orcid.org/0000-0002-7246-8003

Notes

1. Two of these participants provided reports from clinical psychologists who had administered the ADOS-2. One included comparison scores for *communication* and *social interaction*, which are included in the descriptive statistics reported in Table 1. The other only provided assessment results but noted that the child met criteria for an autism diagnosis with marked difficulties with social communication and interaction.
2. We also estimated additional exploratory models on the SRT data, including *age* and *confederate* as random effects. This revealed the same pattern of results described here. These analyses are comprehensively reported in the RMarkdown published alongside this article.

Supplemental material

Supplemental material for this article is available online.

References

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. (2009). Joint engagement and the emergence of language in children with autism and down syndrome. *Journal of Autism and Developmental Disorders*, *39*(1), 84–96. https://doi.org/10.1007/s10803-008-0601-7

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief 'Red flags' for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(2), 202–212.e7. https://doi.org/10.1016/j.jaac.2011.11.003

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Lawrence Erlbaum.

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, *39*(5), 1563–1571. https://doi.org/10.1037/a0032186

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D. (2005). Fitting linear mixed models in R. *R News*, *5*(1), 27–30.

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity. *Psychopathology*, *50*(6), 355–372. https://doi.org/10.1159/000484353

Brizio, A., Gabbatore, I., Tirassa, M., & Bosco, F. M. (2015). 'No more a child, not yet an adult': Studying social cognition in adolescence. *Frontiers in Psychology*, *6*, Article 1011. https://doi.org/10.3389/fpsyg.2015.01011

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*(3), 169–175. https://doi.org/10.1002/mrdd.20036

Box, G.E. and Cox, D.R. (1964) An Analysis of Transformation (with Discussion). *Journal of the Royal Statistical Society B*, *26*, 211–252.

Cañigueral, R., & Hamilton, A. F. d.C. (2019). The role of eye gaze during natural social interactions in typical and autistic people. *Frontiers in Psychology*, *10*, Article 560. https://doi.org/10.3389/fpsyg.2019.00560

Caruana, N., Alhasan, A., Wagner, K., Kaplan, D. M., Woolgar, A., & McArthur, G. (2020). The effect of non-communicative eye movements on joint attention. *Quarterly Journal of Experimental Psychology*, *73*(12), 2389–2402. https://doi.org/10.1177/1747021820945604

Caruana, N., Brock, J., & Woolgar, A. (2015). A frontotemporoparietal network common to initiating and responding to joint attention bids. *NeuroImage*, *108*, 34–46. https://doi.org/10.1016/j.neuroimage.2014.12.041

Caruana, N., Inkley, C., Nalepka, P., Kaplan, D. M., & Richardson, M. J. (2021). Gaze facilitates responsivity during hand coordinated joint attention. *Scientific Reports*, *11*(1), 21037. https://doi.org/10.1038/s41598-021-00476-3

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017a). Detecting communicative intent in a computerised test of joint attention. *PeerJ*, *5*, e2899. https://doi.org/10.7717/peerj.2899

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017b). Simulating social interactions for the experimental investigation of joint attention. *Neuroscience & Biobehavioral Reviews*, *74*, 115–125. https://doi.org/10.1016/j.neubiorev.2016.12.022

Caruana, N., Stieglitz Ham, H., Brock, J., Woolgar, A., Kloth, N., Palermo, R., & McArthur, G. (2018). Joint attention difficulties in autistic adults: An interactive eye-tracking study. *Autism*, *22*(4), 502–512. https://doi.org/10.1177/1362361316676204

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1430), 315–324. https://doi.org/10.1098/rstb.2002.1199

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*, *48*, 79–93. https://doi.org/10.1016/j.ridd.2015.10.011

Conty, L., George, N., & Hietanen, J. K. (2016). Watching Eyes effects: When others meet the self. *Consciousness and Cognition*, *45*, 184–197. https://doi.org/10.1016/j.concog.2016.08.016

Davis, R., & Crompton, C. J. (2021). What do new findings about social interaction in autistic adults mean for neurodevelopmental research? *Perspectives on Psychological Science*, *16*, 649–653. https://doi.org/10.1177/1745691620958010

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, *17*(2), 270–281. https://doi.org/10.1111/desc.12122

Domkin, D., Laczko, J., Jaric, S., Johansson, H., & Latash, M. L. (2002). Structure of joint variability in bimanual pointing tasks. *Experimental Brain Research*, *143*(1), 11–23. https://doi.org/10.1007/s00221-001-0944-1

Falck-Ytter, T., & von Hofsten, C. (2011). How special is social looking in ASD: A review. *Progress in Brain Research*, *189*, 209–222. https://doi.org/10.1016/B978-0-444-53884-0.00026-9

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(7), 546–555. https://doi.org/10.1016/j.jaac.2017.05.005

Gillespie-Lynch, K., Elias, R., Escudero, P., Hutman, T., & Johnson, S. P. (2013). Atypical gaze following in autism: A comparison of three potential mechanisms. *Journal of Autism and Developmental Disorders*, *43*(12), 2779–2792. https://doi.org/10.1007/s10803-013-1818-7

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359–364. https://doi.org/10.1016/j.cognition.2014.11.040

Horstmann, A., & Hoffmann, K.-P. (2005). Target selection in eye–hand coordination: Do we reach to where we look or do we look to where we reach? *Experimental Brain Research*, *167*(2), 187–195. https://doi.org/10.1007/s00221-005-0038-6

Jack, A. I., & Roepstorff, A. (2002). Introspection and cognitive brain mapping: From stimulus–response to script–report. *Trends in Cognitive Sciences*, *6*(8), 333–339. https://doi.org/10.1016/S1364-6613(02)01941-1

Jaswal, V. K., & Akhtar, N. (2019). Being versus appearing socially uninterested: Challenging assumptions about social motivation in autism. *Behavioral and Brain Sciences*, *42*, e82. https://doi.org/10.1017/S0140525X18001826

Ke, F., & Im, T. (2013). Virtual-reality-based social interaction training for children with high-functioning autism. *The Journal of Educational Research*, *106*(6), 441–461. https://doi.org/10.1080/00220671.2013.832999

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340. https://doi.org/10.1348/000712607X251243

Kliemann, D., Dziobek, I., Hatri, A., Steimke, R., & Heekeren, H. R. (2010). Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. *The Journal of Neuroscience*, *30*(37), 12281–12287. https://doi.org/10.1523/JNEUROSCI.0688-10.2010

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lamb, M., Nalepka, P., Kallen, R. W., Lorenz, T., Harrison, S. J., Minai, A. A., & Richardson, M. J. (2019). A hierarchical behavioral dynamic approach for naturally adaptive human-agent pick-and-place interactions. *Complexity*, 2019, *5964632*. https://doi.org/10.1155/2019/5964632

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, *8*, 1–10, Article 302. https://doi.org/10.3389/fnhum.2014.00302

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Leekam, S. (2016). Social cognitive impairment and autism: What are we trying to explain? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1686), 20150082. https://doi.org/10.1098/rstb.2015.0082

Little, G. E., Bonnar, L., Kelly, S. W., Lohan, K. S., & Rajendran, G. (2016). Gaze contingent joint attention with an avatar in children with and without ASD. *In 2016 joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (pp. 15–20). https://ieeexplore.ieee.org/document/7846780

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223.

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Milton, D. E. M. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society*, *27*(6), 883–887. https://doi.org/10.1080/09687599.2012.710008

Milton, D. E. M., Heasman, B., & Sheppard, E. (2018). Double empathy. In F. R. Volkmar (Ed.), *Encyclopedia of autism spectrum disorders* (pp. 1–8). Springer. https://doi.org/10.1007/978-1-4614-6435-8_102273-1

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, *47*(6), 497–514. https://doi.org/10.1111/ejn.13720

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research*, *2*(1), 2–21. https://doi.org/10.1002/aur.61

Murray, D. S., Creaghead, N. A., Manning-Courtney, P., Shear, P. K., Bean, J., & & Prendeville, J.-A. (2008). The relationship between joint attention and language in children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *23*(1), 5–14. https://doi.org/10.1177/1088357607311443

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., Ohta, H., Kato, N., & Kitazawa, S. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1696), 2935–2943. https://doi.org/10.1098/rspb.2010.0587

Nalepka, P., Lamb, M., Kallen, R. W., Shockley, K., Chemero, A., Saltzman, E., & Richardson, M. J. (2019). Human social motor solutions for human–machine interaction in dynamical task contexts. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(4), 1437–1446. https://doi.org/10.1073/pnas.1813164116

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology*, *20*(1), 79–97. https://doi.org/10.1017/S0954579408000047

Newbutt, N., Bradley, R., & Conley, I. (2020). Using virtual reality head-mounted displays in schools with autistic children: Views, experiences, and future directions. *CyberPsychology, Behavior, and Social Networking*, *23*(1), 23–33. https://doi.org/10.1089/cyber.2019.0206

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., Herpertz-Dahlmann, B., Konrad, K., & Schulte-Rüther, M. (2016). Look into my eyes: Investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *NeuroImage*, *130*, 248–260. https://doi.org/10.1016/j.neuroimage.2016.02.026

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Pellicano, E., & Houting, J. (2022). Annual Research Review: Shifting from 'normal science' to neurodiversity in autism science. *Journal of Child Psychology and Psychiatry*, *63*(4), 381–396. https://doi.org/10.1111/jcpp.13534

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, *32*(4), 249–261. https://doi.org/10.1023/A:1016374617369

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, *139*(3), 266–277. https://doi.org/10.1007/s002210100745

Perkovich, E., Sun, L., Mire, S., Laakman, A., Sakhuja, U., & Yoshida, H. (2022). What children with and without ASD see: Similar visual experiences with different pathways through parental attention strategies. *Autism & Developmental Language Impairments*, *7*, 1–16. https://doi.org/10.1177/23969415221137293

Pino, M. C., Vagnetti, R., Valenti, M., & Mazza, M. (2021). Comparing virtual vs real faces expressing emotions in children with autism: An eye-tracking study. *Education and Information Technologies*, *26*(5), 5717–5732. https://doi.org/10.1007/s10639-021-10552-w

Redcay, E., Kleiner, M., & Saxe, R. (2012). Look at this: The neural correlates of initiating and responding to bids for joint attention. *Frontiers in Human Neuroscience*, *6*, Article 169. https://doi.org/10.3389/fnhum.2012.00169

Sasson, N. J., & Morrison, K. E. (2019). First impressions of adults with autism improve with diagnostic disclosure and increased autism knowledge of peers. *Autism*, *23*(1), 50–59. https://doi.org/10.1177/1362361317729526

Scheerer, N. E., Boucher, T. Q., Sasson, N. J., & Iarocci, G. (2022). Effects of an educational presentation about autism on high school students' perceptions of autistic adults. *Autism in Adulthood*, *4*(3), 203–213. https://doi.org/10.1089/aut.2021.0046

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Seibert, J. M., Hogan, A. E., & Mundy, P. C. (1982). Assessing interactional competencies: The Early Social-Communication Scales. *Infant Mental Health Journal*, *3*(4), 244–258. https://doi.org/10.1002/1097-0355(198224)3:4<244::AID-IMHJ2280030406>3.0.CO;2-R

Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–134. https://doi.org/10.1016/j.tics.2008.11.009

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, *189*, 260–274. https://doi.org/10.1016/j.cognition.2019.03.019

Tanaka, J. W., & Sung, A. (2016). The 'eye avoidance' hypothesis of autism face processing. *Journal of Autism and Developmental Disorders*, *46*(5), 1538–1552. https://doi.org/10.1007/s10803-013-1976-7

Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*(1), 509–529. https://doi.org/10.1146/annurev.anthro.28.1.509

Trevisan, D. A., Roberts, N., Lin, C., & Birmingham, E. (2017). How do adults and teens with self-declared Autism Spectrum Disorder experience eye contact? A qualitative analysis of first-hand accounts. *PLOS ONE*, *12*(11), Article e0188446. https://doi.org/10.1371/journal.pone.0188446

Van De Cruys, S., de-Wit, L., Evers, K., Boets, B., & Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to Pellicano and Burr (TICS, 2012). *i-Perception*, *4*(2), 95–97. https://doi.org/10.1068/i0580ic

Van De Cruys, S., Evers, K., Van Der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II)*. Pearson.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes – Using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *5*(1), 98–107. https://doi.org/10.1093/scan/nsq024

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*(3), 229–248. https://doi.org/10.1080/15250000802004437

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLOS ONE*, *8*(11), Article e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development*, *88*(6), 2060–2078. https://doi.org/10.1111/cdev.12730

Yurkovic-Harding, J., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., Yu, C., & Kennedy, D. P. (2022). Children with ASD establish joint attention during free-flowing toy play without face looks. *Current Biology*, *32*(12), 2739–2746.e4. https://doi.org/10.1016/j.cub.2022.04.044

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, *23*(2–3), 143–152. https://doi.org/10.1016/j.ijdevneu.2004.05.001