



Interacting with autistic virtual characters: intrapersonal synchrony of nonverbal behavior affects participants' perception

Carola Bloch^{1,2} · Ralf Tepest² · Sevim Koeroglu² · Kyra Feikes² · Mathis Jording³ · Kai Vogeley^{2,3} · Christine M. Falter-Wagner¹

Received: 29 June 2023 / Accepted: 18 December 2023 / Published online: 25 January 2024
© The Author(s) 2024

Abstract

Temporal coordination of communicative behavior is not only located between but also within interaction partners (e.g., gaze and gestures). This intrapersonal synchrony (IaPS) is assumed to constitute interpersonal alignment. Studies show systematic variations in IaPS in individuals with autism, which may affect the degree of interpersonal temporal coordination. In the current study, we reversed the approach and mapped the measured nonverbal behavior of interactants with and without ASD from a previous study onto virtual characters to study the effects of the differential IaPS on observers ($N=68$), both with and without ASD (crossed design). During a communication task with both characters, who indicated targets with gaze and delayed pointing gestures, we measured response times, gaze behavior, and post hoc impression formation. Results show that character behavior indicative of ASD resulted in overall enlarged decoding times in observers and this effect was even pronounced in observers with ASD. A classification of observer's gaze types indicated differentiated decoding strategies. Whereas non-autistic observers presented with a rather consistent eyes-focused strategy associated with efficient and fast responses, observers with ASD presented with highly variable decoding strategies. In contrast to communication efficiency, impression formation was not influenced by IaPS. The results underline the importance of timing differences in both production and perception processes during multimodal nonverbal communication in interactants with and without ASD. In essence, the current findings locate the manifestation of reduced reciprocity in autism not merely in the person, but in the interactional dynamics of dyads.

Keywords Autism spectrum disorder · Interaction · Nonverbal · Gaze · Gestures · Joint attention

Introduction

Temporal structures that emerge in social interactions between individuals constitute a shared rhythm that influences the mutual evaluation and success of the interaction

[1–6]. However, temporal structures already occur at the subpersonal level as *intrapersonal synchrony* (IaPS) that can be defined as the intrapersonal temporal coordination of multimodal signals (e.g., gaze, gestures, facial expressions, speech) within individuals while engaged in interactions [7]. It is assumed that IaPS is a pre-requisite for interpersonal alignment [7, 8]. IaPS includes implicit processes acquired from infancy on through, for example, behavioral imitation and time-sensitive communication structures in caregiver–child interactions [4, 9–14].

Considering IaPS in an interpersonal scenario, *production* and *perception* of IaPS have to be distinguished, see simplistic scheme in Fig. 1. The intrapersonal coordination of multimodal signals will influence, which and how series of multimodal signals are perceived as a signal-unit through temporal coherence of signal expressions (e.g., as operationalized in the current study, gaze and pointing gesture coupling). The interaction partner can then in turn

Kai Vogeley and Christine M. Falter-Wagner share last-authorship.

✉ Carola Bloch
carola.bloch@uk-koeln.de

¹ Department of Psychiatry and Psychotherapy, Medical Faculty, LMU Clinic, Ludwig-Maximilians-University, 80336 Munich, Germany

² Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, 50937 Cologne, Germany

³ Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Forschungszentrum Juelich, 52425 Juelich, Germany

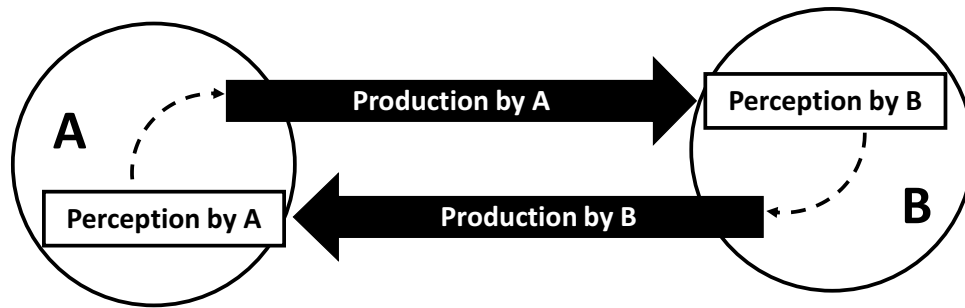


Fig. 1 Simplistic illustration of intrapersonal synchrony (IaPS) production and perception in a dyadic setting. Two individuals **A** and **B** are schematically displayed as circles. The production of intrapersonally synchronized behavior is perceived by the interaction partner. The dashed lines indicate the assumption that the perception systematically affects response behavior within the respective observer who

respond effectively to these signal-units. Thus, an interaction sequence could be conceived as a circular process in which the production of IaPS depends on its perception in the counterpart and vice versa.

Autism spectrum disorder (ASD) is a pervasive developmental disorder that is being clinically diagnosed on the basis of various criteria including characteristics in non-verbal communication and social interactions [15]. Studies have shown reduced interpersonal temporal alignment during dyadic interactions of individuals with and without ASD [16–18]. On the individual level, deviations in the production of IaPS are reflected by diagnostic items in the ADOS [19] and by studies that reported increased and more variable delays in the intrapersonal production of multimodal signals [20, 21]. These differences in IaPS production could potentially affect communication dynamics and explain why interactions between individuals with and without ASD are temporally less aligned and are judged less favorable by typically developed (TD) individuals [20, 22–26]. Having identified systematic group differences in the production of IaPS [20, 21, 27] and assuming that the production of IaPS will affect both perception and production (response behavior) in interactants (see Fig. 1), we systematically investigate the communicative effects of ASD-specific differences in the production of IaPS.

Considering the perception of IaPS, it is plausible to assume perceptual automatisms acquired from infancy [28–32]. In this context of highly automatized communication behaviors, deixis (i.e., making a spatial reference from a personal perspective [32, 33] and the more complex multimodal establishment of *joint attention* through deictic signals (i.e., gaze and pointing gestures) are of particular interest. Infants and caregivers already use this type of communication behaviors before any verbal exchange to spatially reference objects to each other [34, 35]. Deictic actions subserve the establishment of joint attention [33] and are central

will also produce multimodal communicative signals. A simplistic reciprocal interaction sequence could be regarded as a circular process in which production on side (**A**) influences perception and subsequently production on side (**B**), which in turn influences perception and production on side (**A**)

for language acquisition in infants [32, 36] as well as for the development of social cognition [37, 38]. Indeed, atypical perception of joint attention and atypical response behavior during joint attention sequences have been frequently reported in ASD samples [30, 39–42].

To our knowledge, it has not been investigated if and how different IaPS production levels during multimodal joint attention affect observers. In a virtual interaction paradigm, Caruana et al. [43] showed that gaze movements that are congruent with a pointing gesture are automatically integrated in TD observer’s perceptual process and lead to faster responses. Whether and how attenuated gaze–gesture coupling affects responses in observers is unknown. Furthermore, it remains elusive whether there might be an in-group advantage of observers with ASD when observing gaze–gesture coordination that is prototypical of the production in ASD [21].

Study aims

The aim of the current study was to quantify the effects of differences in gaze-pointing delays (i.e., IaPS production levels) on communicative efficiency in a crossed design with ASD and TD observers. This design allowed to study “in-group” and “out-group” effects referring to the two observer groups. IaPS production levels were based on original data from a prior study [21]. The real-life parameters from the previous study were mapped onto two virtual characters in the current study. In a virtual communication task, participants had to decode the nonverbal behavior displayed by the virtual characters and respond to it. Effects of different IaPS production levels on three perceptual domains were measured, namely (i) decoding speed (response times), (ii) visual information search behavior (gaze recordings), and (ii) post hoc impression formation. We assumed that gaze–gesture

delays as produced by TD individuals would lead to effective and thus fast signal decoding in TD observers and that signal encodings outside the TD range would slow down response times and negatively affect post-hoc impression formation ratings. Furthermore, we assumed that these effects would be diminished for observers with ASD when observing IaPS levels as produced by TD individuals.

Methods

Participants

Inclusion criteria were age between 18 and 60 years and normal or corrected to normal vision. An ASD diagnosis (F84.5 or F84.0 according to ICD-10 [15]) was a mandatory inclusion criterion for the ASD group. Depression and use of antidepressants were no exclusion criteria for the ASD group because of the high co-occurrence of depression in ASD and our aim to recruit a representative sample [44]. Beyond that, any current or past psychiatric or neurological disorder was an exclusion criterion as well as current intake of psychoactive medication. All inclusion criteria were preregistered. ASD participants were recruited via the outpatient clinic for autism in adulthood at the University Hospital Cologne. ASD diagnosis was provided by two independent clinicians according to the German national health guidelines for ASD diagnostics [45]. TD participants were recruited as age- (± 5 years) and gender-matched pairs via social media and the hospital's intranet.

Prior to study realization, a power analysis was run in G*Power [46] for sample size planning. Aiming to obtain a

power of 0.95 to detect an effect size f^2 of 0.35, we planned to recruit 25 individuals per group. Due to a technical issue, nine participants with ASD and eight TD participants were unintentionally recorded with missing trials (see Sect. “Data preprocessing”). After fixing the problem, we decided to recruit 10 more individuals per group and subsequently test whether the technical issue had affected the results. In total, we tested 36 participants with ASD from which one person's data had to be excluded from final analysis due to their high diopter having interfered with the eye-tracking system. One further person with ASD was excluded who had too many missing responses (49.4% of response data missing).

The final sample included 34 individuals with confirmed ASD diagnosis (22 identified as male, 11 as female, one as diverse) and 34 TD individuals (23 identified as male, 11 as female). All individuals in the ASD group who enrolled for testing had a confirmed F84.5 diagnosis. All participants provided written informed consent prior to testing and received financial compensation for their participation. Sample characteristics and group comparisons on neuropsychological and autism-screening scales are depicted in Table 1.

Study protocol

The study protocol was approved by the ethics review board of the medical faculty at the University of Cologne (case number: 16–126) and preregistered in the German register for clinical studies (reference number: DRKS00011271) and the Open Science Framework (OSF) (osf.io/dt6vh). The study was performed in line with the principles of the Declaration of Helsinki.

Table 1 Sample characteristics and group comparisons

	TD		ASD		Statistic	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Age ^B	37.77	12.73	39.52	12.80	$U=618$	0.630
AQ ^B	13.39	5.89	39.59	5.77	$U=1116$	<0.001
EQ ^B	47.06	13.77	15.41	8.09	$U=32$	<0.001
SQ ^A	26.35	10.66	39.18	15.08	$t(59.4)=4.05$	<0.001
SPQ	55.15	14.56	53.50	15.06	$t(66)=-0.459$	0.648
ADC ^B	17.41	9.39	51.32	14.80	$U=1126$	<0.001
BDI-II ^B	4.79	5.70	15.77	11.48	$U=928$	<0.001
D2	104.35	11.05	109.44	12.19	$t(66)=1.80$	0.076
vIQ	107.03	8.88	109.88	11.39	$t(65)=1.14$	0.258

AQ Autism Spectrum Quotient, EQ Empathy Quotient, SQ Systemizing Quotient, SPQ Sensory Perception Quotient, ADC Adult Dyspraxia Checklist, BDI Beck's Depression Inventory II, D2 test of concentration abilities, vIQ verbal Intelligence Quotient assessed by 'Wortschatztest' Results of Student's *t*-tests with $\alpha=0.05$.

^AUnequal variances indicated by Bartlett tests and Welch approximation of *df* applied

^BViolation of normality indicated by Shapiros test and results of Wilcoxon *U*-tests reported
AQ data and vIQ data missing for one TD participant each

Prior to the experiment, participants completed German versions of the Autism Spectrum Quotient [47], Empathy Quotient [48], Systemizing Quotient [49], Sensory Perception Quotient [50], and Adults Dyspraxia Checklist [51] as standard neuropsychological testing in the autism outpatient clinic. Testing began with a brief introduction and obtaining written informed consent. Demographic data were collected and the Beck's Depression Inventory [52] was administered. Participants then performed the virtual communication task and answered a post-hoc questionnaire after the task. Afterward they performed a perceptual simultaneity task, which is reported here for completeness but not discussed further. Finally, the 'Wortschatztest' [53] and the D2 [54] were administered. 40 participants (24 with ASD) participated in another study on social interactions and person judgment in ASD on the same day, with a break between studies. The order of the studies was randomized and counterbalanced to prevent systematic order effects.

Animation

Two virtual characters were pre-selected by a pilot study (see Supplementary Material 1 and Supplementary Table 1). Animation of the virtual characters was conducted in Autodesk® Motion Builder in which spatial and temporal parameters from real behavioral parameters in a previous study [21] were closely implemented. In this prior study, participants' task was to communicate the position of a target stimulus that appeared on the left or right side of an opposite screen to an experimenter using both gaze and pointing gestures [21]. For specific information on animation, see Supplementary Material 2.

Virtual communication task design

The virtual communication task consisted of 108 interaction sequences in 6 blocks of 18 trials each. The 6 blocks were divided into 3 blocks per character and condition. All participants interacted with both characters and observed both IaPS conditions. The two characters prototypically displayed different IaPS levels: One displayed a temporal coordination of gaze and gestures that was based on measures in TD individuals (IaPS_{TD}) and the other displayed behavior measured in individuals with ASD (IaPS_{ASD}). The assignment of the two virtual characters to one of the IaPS_{TD} or IaPS_{ASD} conditions and the order in which participants interacted with the two characters was randomized and counterbalanced within groups.

In each trial, the characters selected one of two items located on screens available in the scenery on the left and right side. The preferred item was communicated by a gaze shift and a pointing gesture similar to a joint attention sequence. Participants were instructed to quickly and

accurately select the indicated object by keypress, once they knew what their partner had selected. Participants were not informed in advance that the virtual characters were prototypical of the production modes of the two groups (ASD and TD).

Trial procedure and experimental manipulation of IaPS

The trial procedure of the virtual communication task is depicted in Fig. 2. Each trial started with a fixation cross that was displayed for an average duration of 2500 ms (± 1500 ms, uniformly distributed). Then the characters appeared with the hand resting on the table and gazing toward the observer. This eye-contact position of the character was held for an average duration of 400 ms (SD = 100, normally distributed). The temporal variability in these initial trial sections was implemented to bring a temporal dynamic to the communication task. Next, an object appeared on the left and another on the right screen respectively. After a duration of 200 ms, the character "chose an object" and indicated the preferred object with deictic nonverbal signals to the observer. Thus, a gaze shift to one of the two objects with a duration of 75 ms occurred. The pointing gesture started after a variable delay after the gaze shift onset: In the IaPS_{TD} condition, temporal delays ranged from 100 to 300 ms in steps of 25 ms. This resulted in nine temporal conditions (100, 125, 150, 175, 200, 225, 250, 275, 300 ms) that were displayed once per side (left, right) in a randomized order per block. In the IaPS_{ASD} condition, temporal delays ranged from 250 to 650 ms in steps of 50 ms. This resulted in eight temporal conditions (250, 300, 350, 400, 450, 500, 550, 600, and 650 ms) that were displayed once per side (left, right) in a randomized order per block. The pointing gesture had a duration of 700 ms from onset to final linger position. The trial ended with the character looking and pointing to the target object. Overall, each trial had a fixed duration of 1700 ms from the appearance of the objects to the disappearance of the character and the start of a new trial.

Apparatus & stimuli

The experiment was conducted in a windowless room with steady illumination. The experimental script was run in PsychoPy [55], running on a HP desktop computer. Stimuli were presented on an ASUS 27 inch LCD widescreen monitor with a resolution of 2560 × 1440 pixels and a refresh rate of 120 fps. An EyeLink 1000 Plus eye-tracker by SR Research® in a desktop-mounted mode recorded participants' monocular gaze movements with 1000 fps temporal resolution. Gaze recordings were adjusted to the screen using a 9-point calibration procedure. Participants' heads

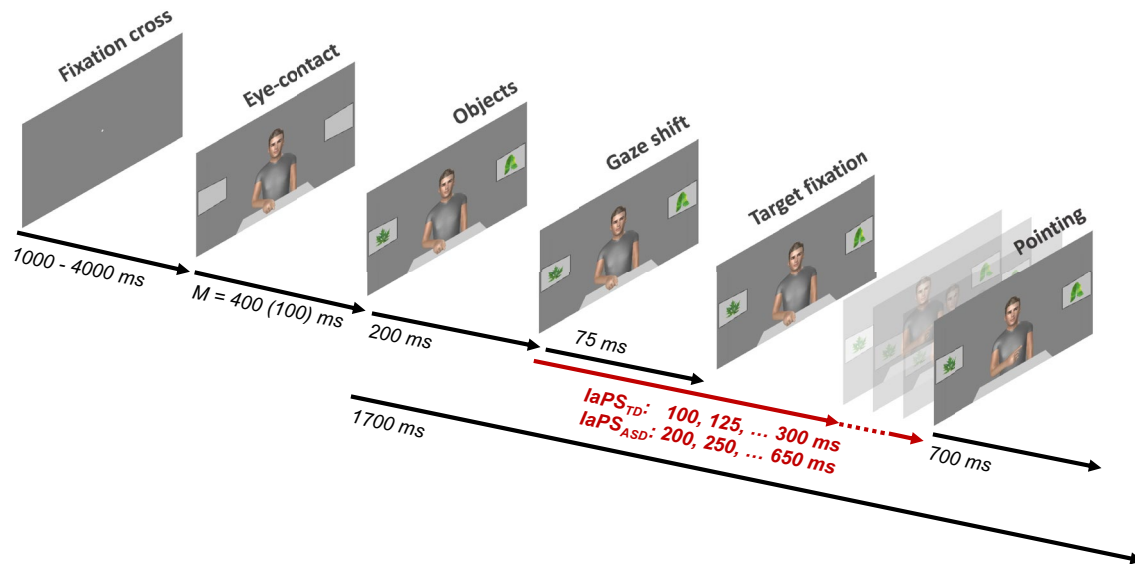


Fig. 2 Trial procedure during virtual communication task. Each trial started with a fixation cross displayed for an average duration of 2500 ms (± 1500 ms, uniformly distributed). The virtual character appeared, facing participants, for an average duration of 400 ms (SD=100 ms, normally distributed). The objects appeared and after a latency of 200 ms a gaze shift (lasting 75 ms) occurred toward one

of the two objects. The respective object was then fixated until the end of the trial. The pointing gesture onset occurred after the gaze shift onset with a variable delay, that accounted for the IaPS conditions (see red arrow). The pointing gesture itself took 700 ms from onset to peak. From object appearance until the end, trials had a fixed duration of 1700 ms

were positioned on a chin-rest (SR Research®) that also stabilized the forehead in a distance of 94 cm from the edge of the table to the screen. A keyboard (German layout) was positioned in front of participants so that the V key was centered in front of the chin rest. Subjects were instructed to place the index finger of the left hand on the Y key and the index finger of the right hand on the M key. Once the virtual partner had selected an object, participants were to press one of the two keys (Y key for left object, M key for right object) in order to select the chosen object for their interaction partners. The time of the key press (measured from object appearance, see Fig. 2) was additionally logged in the gaze data using functions provided by the PyLink Module (SR Research®).

The stimuli used in the study consisted of videos of the animated characters (see Sect. “Animation”). The objects were chosen from an object library provided by Konkle et al. [56]. They appeared on the two screens (left and right) and were derived from the same object category (e.g., butterflies) that had a similar appearance in color and form. The two screen frames were drawn in a way that the same perspective was obtained symmetrically on both sides.

Impression formation ratings

After the virtual communication task, participants were asked to rate each character with regard to nine items about impression formation [24] on a polar slider scale. Items

included clumsiness, likeability, involvement, uncertainty, strangeness of communication, clarity of communication, willingness to spend time with this person, willingness to have a conversation with this person, and willingness to ask this person for help. The order of rating items was randomized and the order of characters was randomized and counterbalanced between participants.

Data preprocessing

Response time data

The time of the key press from the appearance of the objects was recorded. Response times thus indicated the time at which participants decoded the nonverbal communication by their virtual partner. Due to a technical error, one video of one character (rightward gaze-gesture delay of 350 ms) was not presented. This resulted in the systematic loss of three trials (one per block) for 17 participants (9 ASD, 8 TD) in the IaPS_{ASD} condition. We corrected this issue and subsequent participants were tested with the corrected setup. A binary variable in the subsequent analysis was used to dummy-code participants tested before or after correction (see Table 2, model 5).

Trials with missing responses were excluded. In the TD group, there were $n = 48$ (1.3%) missing responses in total (range 1–7 in 18 subjects) and in the ASD group, there were $n = 132$ (3.6%) missing responses (range 1–22 in 29

Table 2 Linear mixed effects models predicting response times (models 1 – 6) and virtual character ratings (model 7) with fixed and random factors

Mod.	Fixed factors	Random factors
1 *	Group * IaPS condition	(IaPS condition subject) + (1 target side) + (1 object category)
2 *	Group * delay	(1 subject) ^A + (1 target side) + (1 object category) + (1 IaPS condition)
3 *	Group * poly(delay, 3)	(1 subject) ^A + (1 target side) + (1 object category) + (1 IaPS condition)
4	Group * IaPS condition * block	(IaPS condition * block subject) + (1 target side) + (1 object category)
5	Group * IaPS condition * correction	(IaPS condition subject) + (1 target side) + (1 object category)
6 ^B	gaze type	(1 subject) + (1 target side) + (1 object category) + (1 IaPS condition)
7*	Group * IaPS condition	(IaPS condition subject) + (1 avatar) + (1 rating item)

*preregistered on OSF (<https://osf.io/dt6vh>)

^AIncluding random slopes for delays resulted in a convergence failure

^BThis model was calculated separately for both groups because of the potential confounding of group and gaze type

All models were fit using the BOBYQA optimizer. Model compositions are given in *lme4* grammar [60]

subjects). Furthermore, trials with > 1 keypresses were excluded. In the TD group, there were $n = 39$ (1.1%) trials with multiple responses (range 1–18 in 14 subjects). In the ASD group, there were $n = 31$ (0.9%) trials with multiple responses (range 1–12 in 11 subjects). Of a maximum of 108 trials, 81–108 trials per person were analyzed in the TD group and 86–108 trials per person were analyzed in the ASD group.

Gaze data

For the gaze data, only the previously mentioned missing trials (see Sect. “[Response time data](#)”) were excluded. Gaze recordings started with object appearance and regions of

interest (RoI) were set for the virtual characters eye area (RoI *gaze*), their gesture trajectory (RoI *gesture*), and the area located between the gaze and gesture areas (RoI *upperBody*). Additionally, the two object areas were defined trial-wise as a target-object area (RoI *target*), and a distractor-object area (RoI *distractor*). Gaze events (saccade onsets/offsets; fixation starts/ends) were read out from the sample data using the online parser provided by EyeLink that applied a velocity threshold for saccades of 30°/second. Dwell times on each predefined RoI were read out per trial using the software Dataviewer provided by SR Research®. Dwell times in regions other than those defined (i.e., random regions) were excluded. After that, relative RoI dwell times were calculated for each IaPS condition and each subject as the

percentage of dwell time on the respective RoI relative to the sum of dwell times in all RoI. These percentages of dwell times in RoIs per IaPS condition were used to classify individuals into gaze types (see Supplementary Material 3 for specifics on the gaze type classification).

Post hoc impression formation

The post hoc impression formation judgements were made on a semi-continuous slider scale ranging from 0 to 100 in single steps. For three rating items (clumsiness; strangeness of communication; insecurity), ratings were rescaled so that high values represented positive valence (e.g., less clumsy; less strange; less insecure) and low values indicated negative valence.

Statistical analysis

Data analysis was conducted in RStudio version 1.4.1103 [57] using the R language version 4.0.3 [58] and functionalities of the *tidyverse* package library [59].

Mixed effects models

Mixed effects models were fitted and analyzed using the packages *lme4* [60] and *afex* [61], see Table 2.

Model assumptions (no multicollinearity, homoscedasticity, and normality of residuals) were visually inspected with functions included in the *performance* package [62]. Likelihood ratio tests (LRT) of nested model were conducted in order to infer if a fixed factor improved model fit above chance level against an alpha of 0.05 using the *afex* function `mixed()`. Type-III sum of squares was used for all tests. Sum coding was applied to all factors and continuous factors were scaled and centered before implementation in the model in order to retain interpretable main and interaction effects [63, 64].

Results

Decoding speed

Response times in both observer groups and IaPS conditions are depicted in Fig. 3A, where the response times have been centered on the start of the virtual character pointing gestures for illustration purposes. In the TD group, responses in the IaPS_{ASD} condition (Fig. 3A, right panel) partly even preceded gesture onsets which was less the case for observers with ASD. In line with that, LRTs of coefficients in Model 1 (see Table 2) showed that there was a significant effect of IaPS condition, with slower responses for characters displaying the IaPS_{ASD} condition ($\chi^2(1) = 13.43, p < 0.001$).

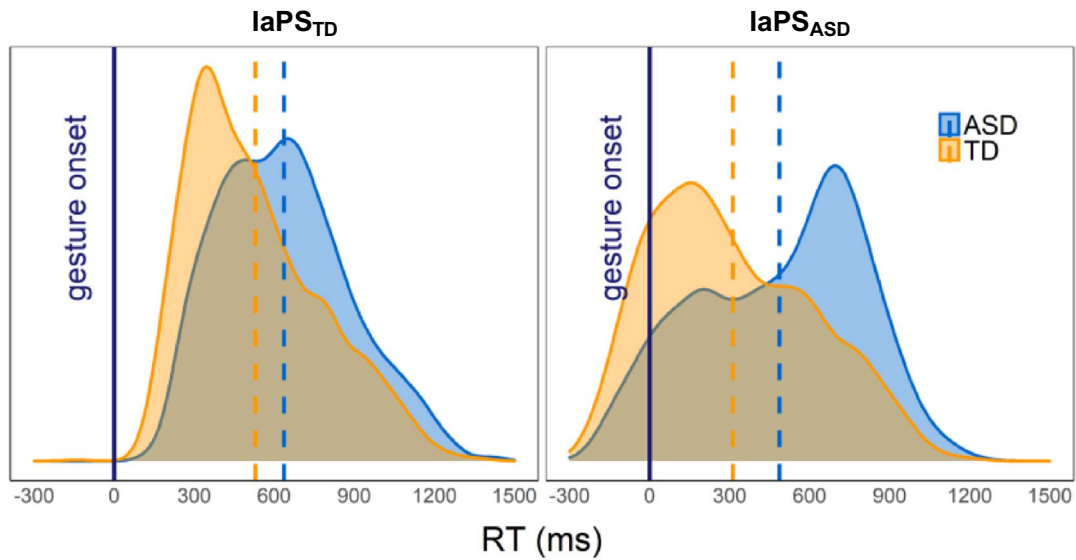
This IaPS effect was enlarged for individuals with ASD, as indicated by a marginally significant interaction term ($\chi^2(1) = 3.85, p = 0.050$), which contradicts a putative in-group advantage. Additionally, there were overall slower decoding speeds in the ASD group ($\chi^2(1) = 7.85, p = 0.005$). Current symptoms of depression, indicated by BDI-II scores, did not significantly affect decoding speed ($\chi^2(1) = 2.15, p = 0.142$) neither did the interaction of BDI-II scores with group ($\chi^2(1) = 0.99, p = 0.321$).

Decoding speeds across different gaze–gesture delays are depicted in Fig. 3B. Since gaze was always the primary signal, a flat line in these graphs would have suggested that observers consistently responded based on the gaze signal only and that varying the delay of the subsequent gesture had no effect. Conversely, a line parallel to the gray guide line in Fig. 3B (annotated ‘pointing peak’) would have suggested that a particular gesture event (e.g., peak of pointing gesture) was consistently used as a decision anchor across delay levels. Any other trajectory indicates that, on a group level, responses were not made based on unimodal signal anchors only, but rather on some form of integration of gaze and gesture signals. In fact, the trajectories in both observer groups indicate integration of signals during decoding, with response times in both groups increasing more or less across delay levels. LRTs (see Model 2 in Table 2) showed a considerable effect of delay levels ($\chi^2(1) = 233.09, p < 0.001$) that was more pronounced in the ASD group (interaction group * delays; $\chi^2(1) = 86.63, p < 0.001$). Including a polynomial term (see Table 2 Model 3) did not improve the model fit (comparison of Model 2 and Model 3, see Table 2) ($\chi^2(4) = 4.93, p = 0.294$). The increase of response times with IaPS delays thus rather presented as a gradually increasing linear relationship without a specific delay at which decoding speeds were particularly faster or slower. The linear increase was however steeper in the ASD group, as indicated by the significant interaction.

As a next step, the influence of experimental blocks was included in the models in order to investigate possible training effects in both observer groups. Response times decreased over blocks (see Supplementary Fig. 1). LRTs (see Table 2 Model 4) indicated that this was a significant effect ($\chi^2(1) = 42.99, p < 0.001$) with no significant difference between groups (i.e., non-significant two-way interaction term group * block; $\chi^2(1) = 0.04, p = 0.838$). Group, condition, and group * condition effects remained significant with the inclusion of block (see Supplementary Table 2 for model estimates). Thus, there appears to have been a training effect, with observers responding faster over time, but this effect was equally pronounced in observers with and without ASD. Hence, the difference between the groups in the effect of IaPS levels cannot be ascribed to a difference in training.

The technical error (i.e., missing condition for some participants, see Sect. “Data preprocessing”) did not

A Response times (pointing onset = 0) in groups and IaPS conditons



B Response times in groups across gaze-gesture delay levels

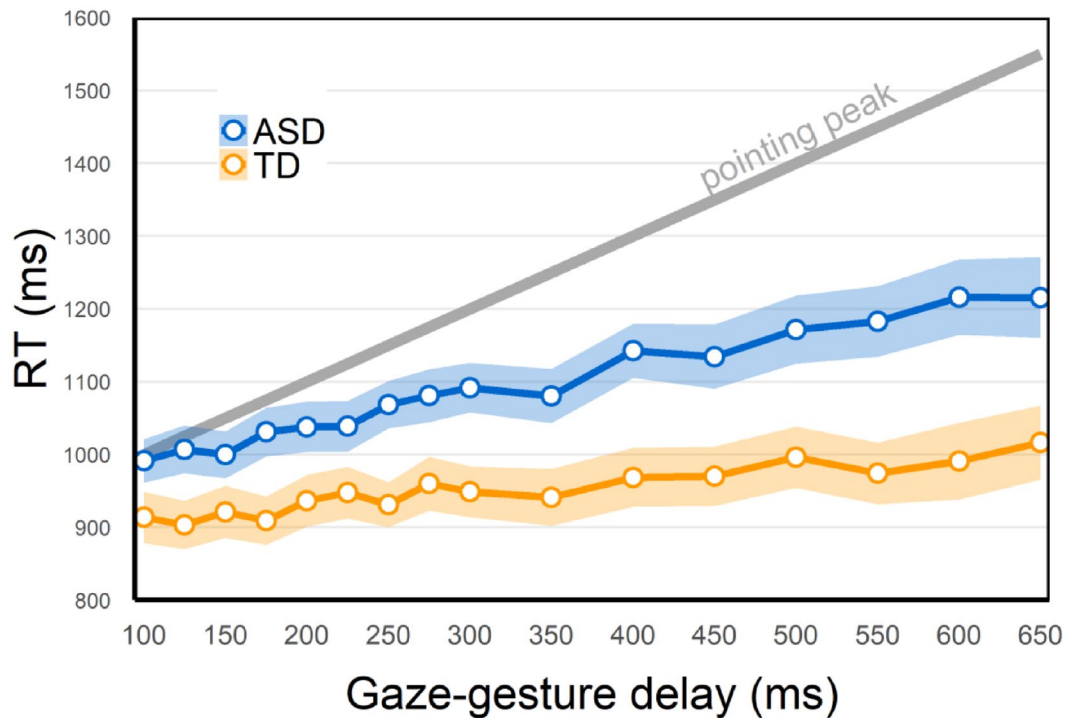


Fig. 3 Decoding speeds at varying intrapersonal synchrony (IaPS) levels. **A** Raw response times are displayed centered to the pointing gesture onsets of virtual characters (=zero x -axis intercept) for illustration purposes. Responses to the character displaying gaze-gesture delays prototypical of TD (IaPS_{TD}) are displayed in the left panel; responses to gaze-gesture coupling prototypical of ASD (IaPS_{ASD}) are displayed in the right panel. TD observer group (orange) and ASD observer group (blue). Dashed lines represent observer group means. Negative response times indicate that observers responded before the virtual character started the pointing gesture.

B Response times averaged across gaze-gesture delay levels (x -axis) as white dots per observer group (TD group in orange; ASD group in blue). Average response times were calculated per delay level independent of the nesting in IaPS condition (e.g., delay of 250 ms was equally present in IaPS_{TD} and IaPS_{ASD} and the mean for this delay was calculated by pooling values over IaPS conditions). Ribbons show standard errors of the marginal means. Gray diagonal line represents the point at which the characters pointing gestures peaked (i.e., reached steady linger position with the extended index finger pointing toward target) relative to each delay condition.

significantly affect response times, as including the binary correction variable (see Table 2 Model 5) did not improve model fit above chance level ($\chi^2(1)=0.11$, $p=0.745$). Group and IaPS effects remained significant, only the interaction term in this model was slightly above significance threshold ($\chi^2(1)=3.64$, $p=0.056$), possibly due to decreased power in the model with more predictors (see Supplementary Table 3 for model estimates).

Exploratory gaze analysis

Summary statistic and group comparisons of the gaze recordings are reported in Table 3. Individuals with and without ASD differed significantly in how much they fixated gaze, gesture, and upper body regions. Gaze types in groups and IaPS conditions are displayed in Fig. 4. First, we conducted a sanity check for the classification. In order to test if the 70% threshold resulted in comparably good representation of gaze patterns in both groups, the percentage values that exceeded the 70% threshold were retrieved per subject (e.g., if for a *gaze_target* type the two areas summed up to 80%, then the fit would have been 10% (=80%–70%). According to a Wilcoxon U test for non-normal data, these fits were comparable between groups ($M_{TD}=12.62\%$ ($SD_{TD}=7.29$)/ $M_{ASD}=12.54\%$ ($SD_{ASD}=7.97$); $U=2288$, $p=0.919$, $r=0.01$).

Two gaze types appeared as dominant in the TD group: 83.8% of TD participants were categorized as *gaze* or *gaze_target* types, irrespective of the IaPS condition. In the ASD group, only 36.76% of participants were categorized into these two types across both IaPS conditions. Overall, gaze types were more heterogeneous in the ASD group with increased numbers of gaze types that excluded the eye region of the character (i.e., RoI *gaze*; see Fig. 4). Furthermore, there were more within-subject gaze type changes in the ASD group, 61.8% of participants with ASD changed gaze types with IaPS conditions, whereas in the TD group, these were 44.1%, while from that 60.0% of changes were again between *gaze* and *gaze-target* types.

Extracting gaze coordinates per trial at the time point at which observers responded revealed again a preference for gaze in the TD group that was less pronounced in the ASD group, see Table 4. The gaze area was mostly fixated during keypresses in the TD group, whereas fixated areas during decision-making were more differentiated in the ASD group with largest percentages in the target RoI.

As visual information search strategies possibly affected decoding speeds, models predicting response times by gaze types were fitted for each group (see Table 2, Model 6). LRTs showed that gaze type explained a considerable amount of variance in the data in both groups (TD: $\chi^2(1)=141.07$, $p<0.001$; ASD: $\chi^2(1)=273.89$, $p<0.001$).

Especially gaze types that excluded the characters eye region yielded fast response times, see Supplementary Fig. 2.

In sum, exploratory gaze analysis showed distinguishable visual information search behavior in both groups. TD individuals seemed to have favored the eye region of their virtual interaction partners whereas individuals with ASD presented with more diversified gaze types that add to the explanation of the group effect in decoding speeds.

Impression formation

Aggregated mean ratings per subject in both IaPS conditions are depicted in Fig. 5. LRTs (Table 2, model 7) showed that the IaPS condition did not affect impression formation ratings above chance level ($\chi^2(1)=0.74$, $p=0.390$). TD individuals rated the virtual characters overall more positive than individuals with ASD ($\chi^2(1)=4.73$, $p=0.030$), however, this group effect did not interact with IaPS condition ($\chi^2(1)=0.10$, $p=0.747$).

Discussion

Arguably, the intrapersonal temporal coordination of multimodal communication signals (e.g., gaze and gestures) is essential for communication efficiency. Given gaze–gesture coordination was shown to be particularly delayed and more variable in interactants with ASD [21, 27], the current study aimed to investigate the effects of group-specific intrapersonal synchrony (IaPS) of deictic gaze and pointing gestures. In a crossed design, we measured decoding speed, visual search behavior, and post-hoc impression formation in observers with and without ASD who interacted with virtual characters exhibiting group-specific couplings, or “in-group/out-group” patterns, of deictic gaze and pointing gestures.

Results showed that observing a virtual character prototypical of TD behavior (IaPS_{TD}) yielded faster response times, hence, a higher communication efficiency compared to responding to a virtual character with ASD behavior (IaPS_{ASD}). This effect was both found in observers with and without ASD and was even pronounced in observers with ASD, which contradicts an in-group advantage following the Double Empathy Hypothesis [65–67]. These results are consistent with findings that interpersonal synchrony is also reduced between individuals with ASD [16], so there is converging evidence against a within-group advantage for individuals with ASD at the temporal level of communication. Since gaze was always the first signal, as in the vast majority of real-life measures in Bloch et al. [21], and in accordance with literature of eye–hand coordination [68–71], a plausible outcome would have been that individuals would use gaze as a unimodal decision anchor. However, on a group-level, observers’ response times increased linearly with increasing

Table 3 Summary statistic and group comparisons of dwell times (in ms) in groups and Regions of Interest (RoI)

	TD			ASD			Statistic	<i>p</i>	<i>effsize</i>
	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>M</i>	<i>SD</i>	<i>Md</i>			
Gaze	1077	375	1091	580	487	465	252	<0.001	0.485
Gesture	81	145	22	260	387	94	757	0.029	0.266
Target	278	210	271	332	206	293	642	0.439	0.095
Distractor	94	73	74	98	87	80	568	0.908	0.015
Upper body	156	192	75	307	277	259	797	0.007	0.326

Means, standard deviations, and medians of dwell times in groups (TD and ASD) and RoI (gaze, gesture, target, distractor, upper body). Results of Wilcoxon tests for group comparisons per RoI are reported

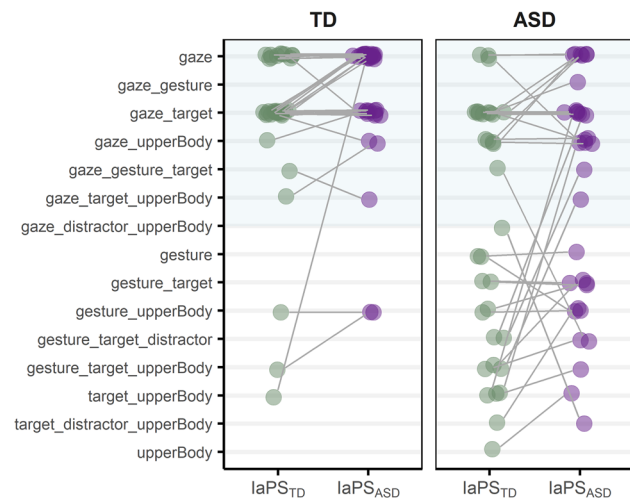


Fig. 4 Gaze types (y-axis) in groups of observers (TD, ASD) and IaPS conditions (IaPS_{TD}, IaPS_{ASD}). Dots display gaze types in IaPS conditions for each subject (IaPS_{TD} left; IaPS_{ASD} right) in both groups of observers (TD left panel; ASD right panel). Two dots per participant, one in each IaPS condition, are connected with gray lines. Horizontally aligned gray lines show individuals whose gaze type did not vary across conditions, vertical lines show individuals whose gaze type varied with IaPS condition. Light blue area in the upper parts of both panels highlight gaze types that included the characters eye region

Table 4 Percentages of fixated Regions of Interest (RoI) at the time a decision was made by keypress

RoI	TD		ASD	
	<i>IaPS_{TD}</i>	<i>IaPS_{ASD}</i>	<i>IaPS_{TD}</i>	<i>IaPS_{ASD}</i>
Gaze	47.58	56.46	18.87	24.39
Point	4.78	6.35	14.07	12.88
Rbody	5.45	4.14	13.62	8.96
Target	35.80	27.15	37.86	37.57
Distractor	1.39	1.19	4.24	3.20
Random	5.00	4.71	11.33	13.00

All values in %. Largest values per group and IaPS condition (i.e., most fixated RoI during keypress) in bold

gaze-gesture delays (Fig. 3B). There was no threshold of delays that made a qualitative difference but rather a graded effect of gaze-gesture delays. Again, the levels of increase contradicted the possibility that a particular gesture event (e.g., pointing peak) affected decoding speeds as a unimodal decision anchor. Instead, the data in both groups suggest an integration of gaze and pointing signals into one complex multimodal signal-unit during decoding. The size of the temporal delays thereby influenced decoding times, but to different degrees in observers with and without ASD.

We further investigated the possibility that the intrapersonal temporal coupling of multimodal signals could yield communicative effects that go beyond low-level decoding processes. In their study, de Marchena & Eigsti [20] found magnified delays between semantic aspects of speech and co-linguistic gestures in adolescents with ASD, which were associated with poorer ratings of communication quality by TD raters. In contrast, in the current study, there was no effect of the IaPS conditions on the impression that observers gained from their virtual interaction partners. As such, the global assumption that deviant temporal subtleties in multimodal nonverbal communication could influence higher-order impression formation was not supported by our data. These results further indicate that the differences between the two virtual characters’ effects on observers were not due to mere personal bias. Thus, the possibility that a negative overall impression of the virtual character influenced the efficiency of communication is not supported here. Especially since the subjects did not receive any information about the differences between the two characters before interacting with them in this rudimentary communication. It is therefore possible to infer that IaPS as a behavioral cue has no effect on impression formation on its own. Arguably, impression formation could rather be driven by more emotional behaviors, as in the stimulus material in Sasson et al. [24]. However, it should be noted that the high similarity in appearance of the two virtual characters and their highly reduced behavioral repertoire could have contributed to the absence of any such “personalized” effects. In addition, the interactive nature of the task was limited in that the virtual

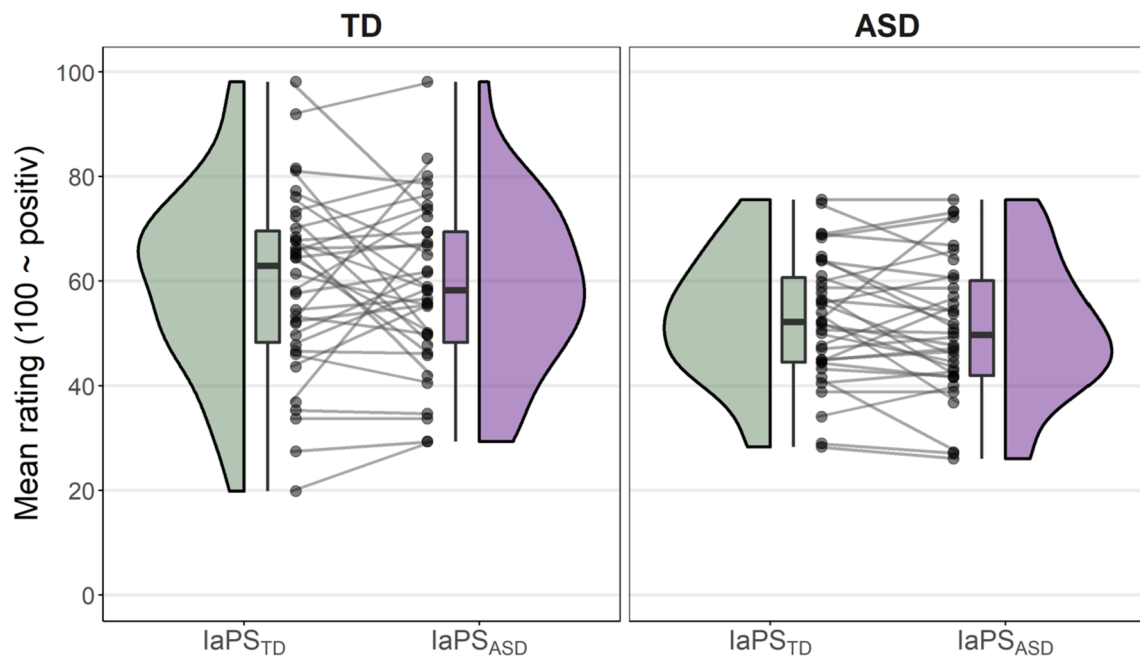


Fig. 5 Mean post hoc character ratings per observer groups (TD, ASD) and IaPS conditions (IaPS_{TD}, IaPS_{ASD}). All character ratings were averaged per subject and IaPS condition. Mean ratings are displayed as black dots. Connected dots display within-subject difference between IaPS conditions. Ratings were scaled so that “100” displays

most positive impressions and “0” displays most negative impressions. Additionally, boxplots and density graphs show distributions of mean ratings in groups (TD left panel; ASD right panel) and IaPS conditions (IaPS_{TD} in green; IaPS_{ASD} in purple)

characters did not respond to the subjects’ choices. In this respect, the difference between the characters may have been too subtle to elicit differences in impression ratings.

Exploratory analysis of gaze behavior suggested that the group difference in decoding speeds was influenced by group-specific strategies of visual information retrieval. TD observers seemed to have deployed a rather homogeneous *eyes-focused* decoding strategy. This is in line with the assumption that a pointing gesture represents a spatially less ambiguous supplement to the primary and faster gaze signal with a potentially confirming feature [28, 29, 43]. It is also in line with the gaze being a special attractor for attention already from infancy onwards [72–76]. Furthermore, it should be taken into account that the eyes represent a spatially smaller signal, and thus focusing on the eyes is an efficient way of multimodal acquisition, since gestures could still be perceived peripherally with covert attention but eyes potentially could not. Thus, the gaze shift represented a fast directional signal, which, if performed at an appropriate temporal delay, could be supported or subsequently confirmed by a congruent pointing gesture. This temporal delay could be influenced by a suggested time window of about 350 ms opening after the gaze shift for the evaluation of social relevance and the intention to establish joint attention [77]. Interestingly, there was a sizeable TD subgroup who were classified as *gaze_target* types in the IaPS_{TD} condition,

and switched to *gaze* types in the IaPS_{ASD} condition (see Fig. 4). This suggests that a temporally coherent pointing gesture, as produced by characters in the IaPS_{TD} condition, may have supported a shift of gaze to targets (i.e., responding to joint attention) and thus triadic attentional processes.

Decoding strategies turned out to be substantially different in observers with ASD. Individuals with ASD generally took longer to respond to the nonverbal communication. However, this group effect should be regarded under consideration of the gaze type analysis, which indicated a reduced focus on the partners’ gaze signal (i.e., more gaze types excluding *RoI gaze*). This is in line with other studies showing reduced eye region focus in observers with ASD [78–82]. As the gaze signal was the faster signal, a reduced focus on the partner’s eye region could at least partially explain the deceleration of response times. Notably, results showed that the increase of response times across different gaze-gesture delays was pronounced in individuals with ASD who were thus more affected by the variation in intrapersonal temporal alignment of multimodal signals. In analogy to the Double Empathy Hypothesis, one might have expected that individuals with ASD might decode nonverbal behavior faster and finally evaluate it better, which is consistent with the mode of production in ASD. However, the results in decoding times and visual information search strategies suggest here that observational behavior

in ASD cannot be explained by an in-group effect. Taken together, the results indicate that individuals with ASD rather deployed strikingly *variable* decoding strategies. An increased involvement of the delayed pointing gesture region during visual search likely have contributed to the attenuation of decoding speeds. This group deviation is in line with ASD entailing atypicalities in attending to and processing gaze cues during joint attention episodes [30, 41, 42, 76, 83–85]. Possibly, such early peculiarities in gaze processing could explain the development of alternative strategies in multimodal communication, as exemplified in the highly variably decoding strategies shown in this study.

Using different strategies to decode multimodal nonverbal signals could potentially contribute to the understanding of reduced interpersonal synchrony in mixed dyads of individuals with and without ASD [16–18]. Koban et al. suggest that behavioral synchronization occurs due to the brain's optimization principle [1]. Accordingly, the core functional principle is to reduce prediction errors through the matching of produced and perceived behavior. Deploying similar strategies to decode multimodal nonverbal behavior as shown for TD observers may probably result in more similar reciprocal response behavior, less prediction errors and ultimately more interpersonal synchrony. This contributes to timing differences in the production of multimodal nonverbal signals [21] that may further lead to a reciprocal mismatch of produced and perceived timed nonverbal behavior and thus to violations of priors in the sense of a Bayesian brain principle [1, 86–88].

Interpersonal synchrony has mostly been studied as a dependent variable, rather than as an independent variable [6]. Nevertheless, it is important to note that other factors influencing interpersonal synchrony have been studied or discussed (e.g., [89, 90]). Studies that investigate a direct relationship between IaPS and interpersonal synchrony are not yet available and should include these aspects as covariates.

Limitations

As limitation aspects, it should be pointed out that the usage of virtual characters, besides the merits of high experimental control and standardization, limits the naturalism and ecological validity of the communication task. Thus, the generalizability of the present results to perceptual and response behavior in real-life scenarios is unclear.

Furthermore, the communication task should be clearly distinguished from naturalistic interaction tasks or tasks with a higher degree of interactivity that do not show the repetitive pattern of gaze and gestural behavior as implemented in this study.

Furthermore, impairments in executive functions in ASD could have added to response delays in the current study

[91], yet could not explain differences between conditions, that were present in both observers with and without ASD.

The generalizability of the results is limited to adults with a F84.5 diagnosis according to ICD-10 and insofar it is unclear how the inclusion of a wider spectrum and age range would have affected the results.

The idea has been put forward that the majority or all psychiatric disorders might be associated with social impairments (e.g., depression or schizophrenia as cited in [92, 93]). While beyond the scope of the current study, future studies might investigate IaPS in a comparative fashion with other clinical groups to clarify specific coordination patterns.

Conclusion

Characteristics of interactions between adults with and without ASD can be quantified by deviations in interpersonal synchronization [16, 17]. Differences in temporal communication patterns between multimodal signals *within* interactants (IaPS) could provide an explanation for such interpersonal deviations. To better understand interpersonal misalignment will require focusing on how interacting individuals process and produce temporal information during multimodal communication. The current results show that the multimodal communication of a virtual character is less efficient, if gaze and gesture are more detached and less integrated. This was the case for both observers with and without ASD. The contrast between a shared eyes-focused decoding strategy in TD observers versus heterogeneous decoding strategies in observers with ASD contribute to a reciprocal mismatch in the temporal dynamics of social interactions between individuals with and without ASD. Essentially, the current results place the symptom of reduced reciprocity in autism not in the individual, but directly in the interaction dynamics of a dyad.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00406-023-01750-3>.

Acknowledgements We kindly thank all those who participated in this study.

Author contributions CB: Conceptualization, methodology, investigation, data curation, software, formal analysis, writing—original draft, writing—review & editing, visualization, project administration. RT: Software, formal analysis, visualization, writing—review & editing. SK: Data acquisition, writing—review & editing. KF: Data acquisition, writing—review & editing. MJ: Conceptualization, funding acquisition, writing—review & editing. KV: Conceptualization, funding acquisition, supervision, project administration, writing—review & editing. CMFW: Conceptualization, funding acquisition, supervision, project administration, writing—review & editing.

Funding Open Access funding enabled and organized by Projekt DEAL. CB and CFW were supported by the DFG (*Deutsche Forschungsgemeinschaft*) grant numbers FA 876/3–1, FA 876/5–1. KV and MJ

were supported by the EC, Horizon 2020 Framework Programme, FET Proactive [Project VIRTUALTIMES; Grant agreement ID: 824128]. KV and SK were supported by the German Ministry of Research and Education [Project SIMSUB; grant ID 01GP2215].

Data and materials availability The conditions of our ethics approval do not permit public archiving of pseudonymized study data. Readers seeking access to the data should contact the corresponding author. Access to data that do not breach participant confidentiality will be granted to individuals for scientific purposes in accordance with ethical procedures governing the reuse of clinical data, including completion of a formal data sharing agreement. The scripts used for data acquisition and analysis are available from the corresponding author. The stimulus material cannot be published for licensing reasons, but can be requested from the corresponding author for scientific purposes.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Koban L, Ramamoorthy A, Konvalinka I (2019) Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle. *Soc Neurosci* 14(1):1–9
- Hove MJ, Risen JL (2009) It's all in the timing: interpersonal synchrony increases affiliation. *Soc Cogn* 27(6):949–960
- Hale J, Ward JA, Buccheri F, Oliver D, de Hamilton AF (2020) Are you on my wavelength? Interpersonal coordination in dyadic conversations. *J Nonverbal Behav* 44(1):63–83
- Feldman R (2007) Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *J Child Psychol Psychiatry* 48(3–4):329–354
- Cacioppo S, Zhou H, Monteleone G, Majka EA, Quinn KA, Ball AB et al (2014) You are in sync with me: neural correlates of interpersonal synchrony with a partner. *Neuroscience* 277:842–858
- Bente G, Novotny E (2020) Bodies and minds in sync: forms and functions of interpersonal synchrony in human interaction. In: Floyd K, Weber R (eds) *The Handbook of Communication Science and Biology*, 1st edn. Routledge, New York, pp 416–428
- Bloch C, Vogeley K, Georgescu AL, Falter-Wagner CM (2019) INTRApersonal synchrony as constituent of INTERpersonal synchrony and its relevance for autism spectrum disorder. *Front Robot AI* 6(73):1–8
- Cañigueral R, de Hamilton AF (2018) The role of eye gaze during natural social interactions in typical and autistic people. *Front Psychol* 10:1–18
- Rochat P, Querido JG, Striano T (1999) Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Dev Psychol* 35(4):950–957
- Jaffe J, Beebe B, Feldstein S, Crown C, Jasnow M, Rochat P et al (2001) Rhythms of dialogue in infancy: coordinated timing in development. *Monogr Soc Res Child Dev* 66(2):1–149
- Beebe B (1982) Micro-timing in mother-infant communication. Key MR, editor. *Nonverbal communication today: Curr Res* 169–195
- Chartrand TL, Van Baaren R (2009) Human mimicry. *Adv Exp Soc Psychol* 41:219–274
- Todisco E, Guijarro-Fuentes P, Collier J, Coventry KR (2020) The temporal dynamics of deictic communication. *Sage First Lang* 41(2):154–178
- Parladé MV, Iverson JM (2015) The development of coordinated communication in infants at heightened risk for autism spectrum disorder. *J Autism Develop Disord* 45(7):2218–2234
- World Health Organization (1993) *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*
- Georgescu AL, Koeroglu S, Hamilton A, Vogeley K, Falter-Wagner CM, Tschacher W (2020) Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: a motion energy study. *Mol Autism* 11(1):1–15
- Koehler J, Georgescu AL, Weiske J, Spangemacher M, Burghof L, Falkai P et al (2021) Brief report: specificity of interpersonal synchrony deficits to autism spectrum disorder and its potential for digitally assisted diagnostics. *J Autism Dev Disord* 52:3718–3726
- Noel JP, De Niar MA, Lazzara NS, Wallace MT (2018) Uncoupling between multisensory temporal function and nonverbal turn-taking in autism spectrum disorder. *IEEE Trans Cogn Dev Syst* 10(4):973–982
- Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, Dillavore PC et al (2000) The autism diagnostic observation schedule—generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev* 30(3):205–223
- de Marchena A, Eigsti IM (2010) Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. *Autism Res* 3(6):311–322
- Bloch C, Tepest R, Jording M, Vogeley K, Falter-Wagner CM (2022) Intrapersonal synchrony analysis reveals a weaker temporal coherence between gaze and gestures in adults with autism spectrum disorder. *Sci Rep* 12(20417):1–12
- Edey R, Cook J, Brewer R, Johnson MH, Bird G, Press C (2016) Interaction takes two: typical adults exhibit mind-blindness towards those with autism spectrum disorder. *J Abnorm Psychol* 125(7):879
- Morrison KE, DeBrabander KM, Jones DR, Faso DJ, Ackerman RA, Sasson NJ (2020) Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. *Autism* 24(5):1067–1080
- Sasson NJ, Faso DJ, Nugent J, Lovell S, Kennedy DP, Grossman RB (2017) Neurotypical peers are less willing to interact with those with autism based on thin slice judgments. *Sci Rep* 7(40700):1–10
- Grossman RB (2015) Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. *Autism* 19(5):580–587
- Wehrle S, Cangemi F, Janz A, Vogeley K, Grice M (2023) Turn-timing in conversations between autistic adults: typical short-gap transitions are preferred, but not achieved instantly. *PLoS One* 18(4):1–22
- Bloch C, Viswanathan S, Tepest R, Jording M, Falter-Wagner CM, Vogeley K (2023) Differentiated, rather than shared, strategies

- for time-coordinated action in social and non-social domains in autistic individuals. *Cortex* 166:207–232
28. Yu C, Smith LB (2017) Hand-eye coordination predicts joint attention. *Child Dev* 88(6):2060–2078
 29. Yu C, Smith LB (2013) Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS One* 8(11):1–10
 30. Senju A, Tojo Y, Dairoku H, Hasegawa T (2004) Reflexive orienting in response to eye gaze and an arrow in children with and without autism. *J Child Psychol Psychiatry Allied Discip* 45:445–58
 31. Mundy P, Sullivan L, Mastergeorge AM (2009) A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Res* 2(1):2–21
 32. Diessel H (2006) Demonstratives, joint attention, and the emergence of grammar. *Cogn Linguist* 17(4):463–489
 33. Stukenbrock A (2020) Deixis, Meta-perceptive gaze practices, and the interactional achievement of joint attention. *Front Psychol* 11:1–23
 34. Salo VC, Rowe ML, Reeb-Sutherland B (2018) Exploring infant gesture and joint attention as related constructs and as predictors of later language. *Infancy* 23(3):432–452
 35. Franco F (2005) Infant pointing: Harlequin, Servant of two masters. In: Eilan N, Hoerl C, McCormack T, Roessler J, editors. *Joint attention: Communication and other minds*. Oxford University Press, 129–164
 36. Brooks R, Meltzoff AN (2008) Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *J Child Lang* 35(1):207–220
 37. Mundy P, Newell L (2007) Attention, joint attention, and social cognition. *Curr Dir Psychol Sci* 16(5):269–274
 38. Mundy P (2018) A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *Eur J Neurosci* 47(6):497–514
 39. Redcay E, Dodell-Feder D, Mavros PL, Kleiner M, Pearrow MJ, Triantafyllou C et al (2013) Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Hum Brain Mapp* 34(10):2511–2523
 40. Charman T (2003) Why is joint attention a pivotal skill in autism? *Philos Trans R Soc Lond B Biol Sci* 358(1430):315–324
 41. Dawson G, Toth K, Abbott R, Osterling J, Munson J, Estes A et al (2004) Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Dev Psychol* 40(2):271–283
 42. Caruana N, Stieglitz Ham H, Brock J, Woolgar A, Kloth N, Palermo R et al (2018) Joint attention difficulties in autistic adults: an interactive eye-tracking study. *Autism* 22(4):502–512
 43. Caruana N, Inkley C, Nalepka P, Kaplan DM, Richardson MJ (2021) Gaze facilitates responsiveness during hand coordinated joint attention. *Sci Rep* 11(1):1–11
 44. Hollocks MJ, Lerh JW, Magiati I, Meiser-Stedman R, Brugha TS (2019) Anxiety and depression in adults with autism spectrum disorder: A systematic review and meta-analysis. *Psychol Med* 49(4):559–572
 45. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. *Autismus-Spektrum-Störungen im Kindes-, Jugend- und Erwachsenenalter - Teil 1: Diagnostik - Interdisziplinäre S3-Leitlinie der DGKJP und der DGPPN sowie der beteiligten Fachgesellschaften, Berufsverbände und Patientenorganisationen Langversion*. AWMF online. 2016.
 46. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2):175–191
 47. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E (2001) The Autism-Spectrum Quotient (AQ): evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* 31(1):5–17
 48. Baron-Cohen S, Wheelwright S (2004) The empathy quotient: an investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *J Autism Dev Disord* 34(2):163–175
 49. Baron-Cohen S, Richler J, Bisarya D, Gurunathan N, Wheelwright S (2003) The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos Trans Royal Soc B: Biol Sci* 358(1430):361–374
 50. Tavassoli T, Hoekstra RA, Baron-Cohen S (2014) The Sensory Perception Quotient (SPQ): development and validation of a new sensory questionnaire for adults with and without autism. *Mol Autism* 5(1):1–10
 51. Kirby A, Edwards L, Sugden D, Rosenblum S (2010) The development and standardization of the Adult Developmental Co-ordination Disorders/Dyspraxia Checklist (ADC). *Res Dev Disabil* 31(1):131–139
 52. Beck AT, Steer RA, Brown GK (1996) *Beck Depression Inventory, 2nd edn*. Manual. 2nd ed. San Antonio, TX: The Psychological Corporation
 53. Schmidt KH, Metzler P (1992) *Wortschatztest*. Beltz
 54. Brickenkamp R (1981) *Test d2 - Aufmerksamkeits-Belastungs-Test*. Handbuch psychologischer und pädagogischer Tests 270–273
 55. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H et al (2019) *PsychoPy2: experiments in behavior made easy*. *Behav Res Methods* 51(1):195–203
 56. Konkle T, Brady TF, Alvarez GA, Olivia A (2010) Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J Exp Psychol Gen* 139(3):558–578
 57. RStudioTeam (2020) *RStudio: Integrated Development for R*. Boston
 58. RCoreTeam (2019) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing
 59. Wickham H, Averic M, Bryan J, Chang W, McGowan LD, François R et al (2019) Welcome to the Tidyverse. *J Open Source Softw* 4(43):1686
 60. Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* <https://doi.org/10.18637/jss.v067.i01>
 61. Singmann H, Bolker B, Westfall J. *Afex: Analysis of Factorial Experiments*. 2021.
 62. Lüdtke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. *performance: An R Package for Assessment, Comparison and Testing of Statistical Models*. *J Open Source Softw*. 2021;6(60).
 63. Singmann H, Kellen D (2019) An introduction to mixed models for experimental psychology. *New Methods Cognit Psychol*. <https://doi.org/10.4324/9780429318405-2>
 64. Brown VA (2021) An introduction to linear mixed-effects modeling in R. *Adv Methods Pract Psychol Sci* 4(1)
 65. Milton DEM (2012) On the ontological status of autism: the ‘double empathy problem.’ *Disabil Soc* 27(6):883–887
 66. Mitchell P, Sheppard E, Cassidy S (2021) Autism and the double empathy problem: implications for development and mental health. *Br J Dev Psychol* 39(1):1–18
 67. Crompton CJ, Ropar D, Evans-Williams CVM, Flynn EG, Fletcher-Watson S (2020) Autistic peer-to-peer information transfer is highly effective. *Autism* 24(7):1704–1712
 68. Horstmann A, Hoffmann KP (2005) Target selection in eye-hand coordination: do we reach to where we look or do we look to where we reach? *Exp Brain Res* 167(2):187–195
 69. Prablanc C, Echallier JE, Jeannerod M, Komilis E (1979) Optimal response of eye and hand motor systems in pointing at a

- visual target – II. Static and dynamic visual cues in the control of hand movement. *Biol Cybern.* 35(3):183–187
70. de Brouwer AJ, Flanagan JR, Sperling M (2021) Functional use of eye movements for an acting system. *Trends Cogn Sci* 25(3):252–263
 71. Sailer U, Eggert T, Ditterich J, Straube A (2000) Spatial and temporal aspects of eye-hand coordination across different tasks. *Exp Brain Res* 134(2):163–173
 72. Argyle M, Cook M (1976) *Gaze and mutual gaze*. Cambridge U Press
 73. Jording M, Hartz A, Bente G, Schulte-Rüther M, Vogeley K (2018) The ‘Social Gaze Space’: a taxonomy for gaze-based communication in triadic interactions. *Front Psychol* 9:1–8
 74. Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci Biobehav Rev* 24(6):581–604
 75. Georgescu AL, Kuzmanovic B, Schilbach L, Tepest R, Kulbida R, Bente G et al (2013) Neural correlates of ‘social gaze’ processing in high-functioning autism under systematic variation of gaze duration. *Neuroimage Clin* 3:340–351
 76. Senju A, Johnson MH (2009) The eye contact effect: mechanisms and development. *Trends Cogn Sci* 13(3):127–134
 77. Caruana N, de Lissa P, McArthur G (2015) The neural time course of evaluating self-initiated joint attention bids. *Brain Cogn* 98:43–52
 78. Robles M, Namdarian N, Otto J, Wassiljew E, Fellow NN, Falter-wagner C et al (2022) A virtual reality based system for the screening and classification of autism. *IEEE Trans Vis Comput Graph* 28(5):2168–2178
 79. Roth D, Jording M, Schmee T, Kullmann P, Navab N, Vogeley K (2020) Towards computer aided diagnosis of autism spectrum disorder using virtual environments. *Proceedings - 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR* 115–122
 80. Papagiannopoulou EA, Chitty KM, Hermens DF, Hickie IB, Lagopoulos J (2014) A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Soc Neurosci* 9(6):610–632
 81. Tanaka JW, Sung A (2016) The, “Eye Avoidance” hypothesis of autism face processing. *J Autism Dev Disord* 46(5):1538–1552
 82. Madipakkam AR, Rothkirch M, Dziobek I, Sterzer P (2017) Unconscious avoidance of eye contact in autism spectrum disorder. *Sci Rep* 7(1):13378
 83. Caruana N, McArthur G, Woolgar A, Brock J (2017) Detecting communicative intent in a computerised test of joint attention. *PeerJ* 2017(1):1–16
 84. Nyström P, Thorup E, Bölte S, Falck-Ytter T (2019) Joint attention in infancy and the emergence of autism. *Biol Psychiatry* 86(8):631–638
 85. Pierno AC, Mari M, Glover S, Georgiou I, Castiello U (2006) Failure to read motor intentions from gaze in children with autism. *Neuropsychologia* 44(8):1483–1488
 86. van de Cruys S, Evers K, van der Hallen R, van Eylen L, Boets B, de-Wit L, et al (2014) Precise minds in uncertain worlds: Predictive coding in autism. *Psychol Rev* 121(4):649–75
 87. Parr T, Friston KJ (2019) Generalised free energy and active inference. *Biol Cybern [Internet]* 113(5–6):495–513
 88. Bolis D, Dumas G, Schilbach L (2023) Interpersonal attunement in social interactions: from collective psychophysiology to interpersonalized psychiatry and beyond. *Philos Trans R Soc Lond B Biol Sci* 378(20210365):1–13
 89. Brambilla M, Sacchi S, Menegatti M, Moscatelli S (2016) Honesty and dishonesty don’t move together: trait content information influences behavioral synchrony. *J Nonverbal Behav* 40(3):171–186
 90. Lumsden J, Miles LK, Richardson MJ, Smith CA, Macrae CN (2012) Who syncs? Social motives and interpersonal coordination. *J Exp Soc Psychol* 48(3):746–751
 91. Demetriou EA, Lampit A, Quintana DS, Naismith SL, Song YJC, Pye JE et al (2018) Autism spectrum disorders: a meta-analysis of executive function. *Mol Psychiatry* 23(5):1198–1204
 92. Schilbach L (2016) *Towards a second-person neuropsychiatry*. Philosophical Transactions of the Royal Society B: Biological Sciences. Royal Society of London
 93. Vogeley K (2018) Communication as Fundamental Paradigm for Psychopathology. In: Newen A, De Bruin L, Gallagher S (eds) *The Oxford Handbook of 4E Cognition*, 1st edn. Oxford University Press, UK, pp 805–820
 94. Kenny L, Hattersley C, Molins B, Buckley C, Povey C, Pellicano E (2016) Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism* 20(4):442–462
 95. Buijsman R, Begeer S, Scheeren AM (2022) ‘Autistic person’ or ‘person with autism’? Person-first language preference in Dutch adults with autism and parents. *Autism* 27(3):788–795
 96. Bury SM, Jellet R, Haschek A, Wenzel M, Hedley D, Spoor JR (2022) Understanding language preference: Autism knowledge, experience of stigma and autism identity. *Autism* 0(0):1–13
 97. Monk R, Whitehouse AJO, Waddington H (2022) The use of language in autism research. *Trends Neurosci* 45(11):791–793
 98. Tepest R (2021) The meaning of diagnosis for different designations in talking about autism. *J Autism Dev Disord* 51(2):760–761