

Altered Gesture and Speech Production in ASD Detract from In-Person Communicative Quality

Laura M. Morett^{1,2}  · Kirsten O'Hearn¹ · Beatriz Luna^{1,2} · Avniel Singh Ghuman^{1,2,3}

© Springer Science+Business Media New York 2015

Abstract This study disentangled the influences of language and social processing on communication in autism spectrum disorder (ASD) by examining whether gesture and speech production differs as a function of social context. The results indicate that, unlike other adolescents, adolescents with ASD did not increase their coherency and engagement in the presence of a visible listener, and that greater coherency and engagement were related to lesser social and communicative impairments. Additionally, the results indicated that adolescents with ASD produced sparser speech and fewer gestures conveying supplementary information, and that both of these effects increased in the presence of a visible listener. Together, these findings suggest that interpersonal communication deficits in ASD are driven more strongly by social processing than language processing.

Keywords Gesture · Language · Social communication · Dialogue · Adolescence

Preliminary results from this work were presented in May 2014 at the 13th meeting of the International Forum for Autism Research (IMFAR), Atlanta, GA.

✉ Laura M. Morett
morett@pitt.edu

¹ Department of Psychiatry, University of Pittsburgh, S906 Scaife Hall, 3550 Terrace St., Pittsburgh, PA 15261, USA

² Center for the Neural Basis of Cognition, Pittsburgh, PA, USA

³ Department of Neurosurgery, University of Pittsburgh, Pittsburgh, PA, USA

Introduction

According to the DSM-5, autism spectrum disorder (ASD) is characterized by impairments in two major areas: communication and social interaction, and restricted and repetitive behaviors and interests (American Psychiatric Association 2013). Normal communication relies on both language processing and social processing; thus, both of these factors likely contribute to communication impairments in high-functioning ASD, although they may do so in different ways. Typically developing (TD) individuals use co-speech gesture to communicate information in addition to the information that they convey via speech to one another face-to-face. Thus, in TD individuals, gesture is a key component of language processing, as well as social interaction between listeners, and is affected by each of these factors (McNeill 2005; Mol et al. 2012). When the social context of communication does not allow people to see one another (e.g., communication via phone or computer), gesture does not reflect social interaction, given that speakers are aware that their listeners cannot see their gestures (Alibali et al. 2001; Bavelas et al. 2008). Moreover, people speak more slowly and less fluently when they cannot see one another, indicating that speech production requires greater effort or care in the presence of a non-visible listener (Alibali et al. 2001). By manipulating the social context of communication via listener visibility, we examined the differentiability of the impacts of social context and language processing on speech and gesture abnormalities in high-functioning adolescents with ASD. By doing so, this study provides insight into sensitivity to the social context of communication in ASD via speech, language, and gesture, clarifying the role of key contributors to ineffective communication via these modalities in ASD.

In the DSM-5, the previously separate diagnostic criteria of ASD for speech and language processing and social interaction were combined, implying that deficits in these domains are inherently related. Nevertheless, previous work has not fully elucidated whether abnormalities in the communication of individuals with ASD stem more from problems with language processing or understanding the social context of communication. Language processing requires both linguistic aptitude and appropriate social attention to listeners' speech, making disentangling the social and linguistic aspects of communication difficult. Indeed, it has been postulated that the ability to interpret the intentions and emotions of others is intertwined with successful language development, explaining the deficits in language acquisition that many children with ASD demonstrate (Astington and Baird 2005; Baron-Cohen 1997; Tomasello 2009). In support of this point, deficits in the ability of children with ASD to focus on object concurrently with other individuals (termed joint attention) are associated with deficits in language development (Dawson et al. 2004; Mundy et al. 1990). Moreover, in high-functioning adolescents and adults with ASD, there is evidence that deficits in interpreting intentions and emotions are closely related to deficits in language comprehension and communication, explaining the difficulty that individuals with ASD encounter comprehending figurative language (Happé 1993; Martin and McDonald 2004). Together, these findings highlight the difficulty of disambiguating the influences of language and social context on communication in ASD. Also, given that the findings concerning language development in ASD are based on developmentally disabled children, they highlight the need for research examining the influences of these factors during development in high-functioning individuals with ASD.

Speech production abnormalities in high-functioning ASD tend to fall primarily into two categories: prosodic errors, which concern rhythmic features of speech such as stress and intonation; and pragmatic errors, which concern the context of speech. Both types of errors pervade the speech of individuals with ASD from early childhood into adolescence and adulthood (Baltaxe 1977; Landa 2000; McCann and Peppe 2003; Shriberg 2001). In school-aged children with ASD, prosodic errors are related to expressive and receptive language processing ability (McCann et al. 2007), and in adolescence, some prosodic abnormalities, such as stress and hypernasality, are related to ratings of sociability and communicative competence (Paul et al. 2005). Together, these studies indicate that abnormal processing of language and social context both contribute to the development of communication deficits of ASD, albeit differentially. However, research on speech production has not fully differentiated impacts of language processing and social context on communication deficits in ASD.

Given that they reflect language processing and social context and are not formally taught, the production of co-speech gestures—meaningful hand motions that accompany speech (hereafter referred to simply as gestures)—provide a particularly informative medium for examination of the underpinnings of communication impairments in high-functioning ASD. According to the most popular gesture classification scheme (McNeill 1992, 2005), which is based on the assumption that gesture and speech arise from unitary mental representations and are a component of language, gestures can be grouped into one of four categories. These categories are as follows: iconic gestures, which convey the physical affordances of concrete entities or actions (e.g., sweeping motions accompanying the word *broom*); metaphorical gestures, which convey an abstract idea by physically expressing concrete attributes associated with it (e.g., moving the hands apart horizontally to convey length); beat gestures, which are simple rhythmic movements reflecting speech prosody or emphasis (e.g., finger taps produced on stressed syllables of an utterance), or deictic gestures, which direct attention or convey directionality through their handshape, consisting of one or more fingers extended in the direction of a concrete or abstract entity (e.g., pointing to the wrist, where a watch is worn, to indicate the time). Although gestures are consistent with the structure and meaning of language, they do not convey these attributes *in toto* like speech, and they often provide information supplemental to that conveyed by speech, such as direction of motion or procedural information (Cook and Tanenhaus 2009; Kita and Özyürek 2003). Furthermore, because they are learned primarily through observation and experience, gestures allow one to examine communication outside of the context of formally taught language knowledge.

Within normal conversation, gestures are produced regardless of whether they will be seen; however, TD adults produce more and larger gestures, and more iconic gestures, in the presence of a visible listener (Alibali et al. 2001; Bavelas et al. 2008). These differences in the number, size, and iconicity of gestures suggest that TD speakers are aware of the information that gestures convey, and that they adjust the quantity and type of gestures that they produce accordingly. In addition to these communicative functions, there is evidence that gesture reflects—and can even facilitate—speakers' own cognition. Speakers recall more information from working memory when they gesture during an unrelated task than when they do not gesture during that task (Goldin-Meadow et al. 2001), and restricting gesture production hinders speakers' ability to recall words from long-term memory (Frick-Horbury and Guttentag 1998). Together, these findings demonstrate that, for TD individuals, gesture serves both social and cognitive/linguistic functions, and that the amount of gestures

produced, the attributes of gesture (e.g., iconicity), and the information conveyed by it (e.g., semantic, affective, etc.) reflect these functions.

To date, most research on gesture production in ASD has focused on production of early prelinguistic gestures (e.g., pointing and showing gestures) (Charman et al. 2003; Colgan et al. 2006; Loveland and Landry 1986; Stone et al. 1997), mainly by developmentally disabled children, in the presence of an in-person listener. Developmentally disabled children with ASD are less likely to respond to deictic (pointing) gestures than developmentally disabled TD children (Loveland and Landry 1986; Mundy et al. 1990). Indeed, even high-functioning adults with ASD allocate less attention to gestures than to accompanying speech, as measured via eyetracking (Silverman et al. 2010). In terms of production, relative to the gestures of TD children, representational gestures (i.e., gestures depicting semantic content through their form, placement, and/or motion) produced by children with ASD are less frequent (Loveland et al. 1988), less varied (Colgan et al. 2006), and communicate less information not explicitly expressed through speech (Attwood et al. 1988; Loveland and Landry 1986). Children with ASD also use fewer deictic (pointing) gestures than their TD counterparts (Osterling and Dawson 1994), particularly to share experiences with listeners (Camaioni et al. 1997), and their attention-directing speech and gesture varies less across social contexts (Landry and Loveland 1989). By adolescence, individuals with ASD produce representational gestures with frequency similar to that of TD individuals, consistent with a larger trend of normalization of behavioral differences in cognition from childhood into adolescence in ASD (O’Hearn et al. 2010, 2011). However, there is evidence that, relative to gestures produced by TD adolescents, gestures produced by adolescents with high functioning ASD are more temporally asynchronous with speech (de Marchena and Eigsti 2010). These findings suggest that individuals with ASD encounter difficulty understanding how to use gesture to communicate effectively in different social contexts, which is evident in the relationship between their gesture and speech.

To our knowledge, no study to date has compared speech and gesture production by high-functioning individuals with ASD in the presence or absence of a visible listener to directly probe the influence of social context on communication. Here, we examine how gesture and speech production reflect impairments processing language and social context in ASD by examining gesture and speech produced by TD and ASD adolescents as they re-tell a story, sometimes in the presence of a visible listener and sometimes without seeing the listener. When relating narratives, people gesture spontaneously while speaking, communicating semantic and affective information to their

listener, even if they cannot be seen. Manipulating listener visibility allows the influences of language processing and social context on speech and gesture production to be examined, providing insight into the impact of these factors on communication deficits in ASD. The results from these manipulations will indicate how these factors reflect the origins of communication deficits in ASD.

We predicted that the presence of a visible listener would negatively affect the quality of high functioning ASD adolescents’ communications via speech and gesture, whereas it would not affect the quality of TD adolescents’ communications. This hypothesis is based on findings suggesting that the communication of individuals with ASD is superior in the presence of non-human agents relative to human listeners (Tartaro and Cassell 2008). Additionally, we predicted that adolescents with ASD would produce speech containing more dysfluencies in the presence of visible than non-visible listeners, whereas no such difference would be found in the speech of TD adolescents. This prediction is based on work indicating that the speech produced by adolescents with ASD during in-person interactions contains pragmatic and prosodic abnormalities (Baltaxe 1977; Landa 2000; Paul et al. 2005; Shriberg 2001). Together, these predicted results would demonstrate that social context significantly impacts communication via speech and gesture in ASD.

Methods

Participants

41 adolescents (18 ASD, 23 TD) participated in this task, in addition to other cognitive tasks. All participants were fluent English speakers. Participants were group matched on chronological age, gender, and verbal IQ (see Table 1 for demographic information), with a cut-off in IQ at 85, one standard deviation below the mean. All participants in the ASD group were originally diagnosed with autistic disorder (with the exception of one participant diagnosed with Asperger’s Disorder) via clinical judgment based on DSM-IV-TR criteria. Because DSM-5 criteria for autism spectrum disorder encompass all pervasive developmental disorders (PDDs) except for Rett’s Syndrome, the findings of this study apply to a subset of individuals who would be diagnosed with ASD under DSM-5—namely, high functioning individuals with more severe symptoms. Thus, given that about 91 % of individuals who meet DSM-IV criteria for PDDs other than Rett’s also meet DSM-5 criteria for autism spectrum disorder (Huerta et al. 2014), it may be the case that not all of the participants would be diagnosed with autism spectrum disorder under DSM-5 criteria. Using DSM-IV-TR criteria, diagnoses were

Table 1 Participants' demographic information and scores on the ADOS, ADI-R, and WAIS-III

| | Group | | <i>t</i> | <i>p</i> | <i>d</i> |
|--|-----------------------------|-----------------------------|----------|----------|----------|
| | ASD <i>M</i> (SD) Range | TD <i>M</i> (SD) Range | | | |
| <i>N</i> | 18 | 21 | | | |
| Gender (M:F) | 15:3 | 15:6 | .68 | .50 | .23 |
| Age (years) | 15.17 (2.75) 10.44–19.49 | 15.81 (2.42) 11.52–19.36 | .79 | .44 | .25 |
| <i>WAIS-III</i> | | | | | |
| Verbal IQ | 105.6 (14.71) 87–144 | 106.0 (11.48) 92–132 | .10 | .92 | .03 |
| Performance IQ | 101.56 (14.15) 77–124 | 109.76 (8.88) 92–128 | 2.16 | .04 | .69 |
| Full scale IQ | 104.83 (14.33) 85–134 | 108.65 (8.00) 96–130 | 1.08 | .29 | .33 |
| <i>ADOS</i> | | | | | |
| Communication (C) | 4.25 (1.24) 2–6 | | | | |
| Social responsiveness (SR) | 8.19 (1.76) 5–12 | | | | |
| Full scale (C + SR) | 12.44 (2.42) 8–17 | | | | |
| <i>ADI-R</i> | | | | | |
| Communication (C) | 16.0 (4.31) 8–23 | | | | |
| Social functioning (SF) | 19.69 (4.53) 13–29 | | | | |
| Restrictive and repetitive behaviors (RRB) | 5.61 (1.94) | | | | |

confirmed in the ASD group and ruled out of the TD group using the ADI-R (Lord et al. 1994) and the ADOS (Lord et al. 2000). Exclusion criteria for participants with ASD included the following conditions: epilepsy, meningitis, encephalitis, diabetes, childhood disintegrative disorder, and PDD-NOS. Other conditions comorbid with ASD were not recorded. In the TD group, only participants without comorbid psychiatric disorders were included, and none of the first-degree relatives of TD participants were diagnosed with any pervasive developmental disorder. This study was approved by the University of Pittsburgh Institutional Review Board. Prior to experimental sessions, written consent and assent for participation, including video recording, was obtained from all participants and their caregivers.

Measures

The Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994) and Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000) were used to confirm the diagnosis of participants with ASD. For the ADOS, participants

completed either Module 3 ($n = 9$) or Module 4 ($n = 9$), depending on their maturity level. Modules 3 and 4 provide comparable scores, so scores were collapsed across modules. As can be seen in Table 1, all participants in the ASD group scored at or above the cutoff for a diagnosis in both the communicative and social domains (8 and 10, respectively) on both the ADI-R and ADOS. In the restrictive and repetitive behavior domain, participants also scored above the cutoff for a diagnosis in the restricted and repetitive behaviors domain (3) on the ADI-R. For the ADOS, restricted and repetitive behavior scores were not collected given that there is no established cutoff score for this domain.

The Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler 1999) was used to measure the intelligence quotient of all participants. As can be seen in Table 1, the TD and ASD groups did not differ significantly in VIQ or FSIQ. Given that the narrative recounting task employed in this study relies heavily on verbal communication and minimally on performance-related skills, the minor discrepancy in PIQ between the ASD and TD groups was deemed non-problematic for the purposes of this study. All

analyses produced similar results when entering PIQ as a covariate.

Procedure

Participants watched two 4-minute clips of a “Tweety and Sylvester” cartoon, “Canary Row,” which has been used in numerous other studies of gesture production (Alibali and Don 2001; Alibali et al. 2001; McNeill 2005). Each of the clips contained 4 events in which Sylvester (a cat) attempts to catch Tweety (a canary) in a different way. Prior to watching the clips, participants were instructed to try to remember the events in them as best they could, since they would be asked to recount them afterward. Following each segment, participants were asked to describe to the experimenter what happened in the cartoon in as much detail as possible.

If participants’ retellings were extremely brief, the experimenter prompted them by asking, “What else happened in the cartoon?”. Pilot testing indicated that occasional prompting was necessary to ensure that all participants produced a sufficient amount of speech for analysis, given that a subset of participants provided 1–2 sentence responses. In the final sample, 25 participants were prompted at least once, and participants with ASD were more likely to be prompted than TD participants [mean ASD = .31 (SD = .48); mean TD = .59 (SD = .50); $F(1, 39) = 4.88, p = .03, \eta_p^2 = .11$]. Prompting was not significantly different between visibility conditions, and there was no significant diagnosis by visibility interaction. The greater probability of prompting with the individuals with ASD means that the results reported here underestimates what the difference between groups would have been without prompting, particularly with regards to count variables, such as total words, total gestures, speech duration, etc.

The critical independent variable in this study was listener visibility, which was manipulated by placing or removing an opaque cardboard screen on a table between the participant and the listener (the experimenter). When recounting the events of one clip, participants spoke to the listener face-to-face (visible condition). When recounting the events of the other clip, the screen was placed so that the listener could not be seen behind it (non-visible condition). The order of visibility conditions and clip presentation was counterbalanced across participants. In both cases, participants were aware of the presence of the listener, who provided the initial and any subsequent verbal prompts. When recounting the events of the video clips, participants were discreetly video recorded by a webcam located at an angle perpendicular to the table at which participants and the experimenter were seated.

Transcription and Coding

The story recounting task was used to assess the quantity, quality, and characteristics of speech and gesture spontaneously produced by TD and ASD adolescents, as well as the relationships in meaning (semantic) and timing (temporal) between them. Both coders and raters were research assistants unaware of the experimental design, hypotheses, and participant diagnoses at the time of rating.

Speech Transcription and Analysis

All retellings were transcribed word-for-word in ELAN (Max Planck Institute for Psycholinguistics; <http://tla.mpi.nl/tools/tla-tools/elan>) from digital video files recorded by the webcam during experimental sessions. Utterances were segmented based on a combination of linguistic and suprasegmental cues. Speech dysfluencies such as discourse markers (um, uh, etc.), word fragments, revisions, and repeated words were transcribed in light of evidence that they indicate awareness of shared knowledge in ASD, as well as TD (de Marchena and Eigsti 2015). Silent pauses equal to or greater than two seconds were coded, as were word and phrase repetitions (e.g., “Tweety... Tweety Bird climbs the spout,”) and revisions (“Tweety... The bird climbs the spout,”).

Several measures of speech complexity were calculated using transcribed speech. Text analysis measures were used as a metric of speech complexity because there are no comparable automated measures specifically designed for speech analysis. Most measures consisted of simply counting and averaging the number of occurrences of syllables, words, or phrases in each retelling. Speech duration was calculated by summing the length (in seconds) of all utterances produced by a participant. Mean length of utterance was calculated by dividing the number of morphemes (the smallest grammatical unit of language) by the number of utterances. Lexical density was calculated by dividing the number of distinct words produced by the total number of words produced, and multiplying by 100. Flesh-Kincaid grade level was calculated using the following standard formula: $.39 (\text{total words}/\text{total sentences}) + 11.8 (\text{total syllables}/\text{total words}) - 15.59$. Measures of speech complexity were calculated using the Readability Score.com website (<https://readability-score.com/>), based on aspects of speech (words, utterances, sentences, etc.) transcribed manually as described above.

Gesture Identification and Classification

Gestures were defined as meaningful hand motions produced in conjunction with speech. Similar to speech, all gestures were also coded using ELAN from digital videos

recorded during experimental sessions. To control for differences in gesture production stemming from differences in speech production, the number of gestures produced per 100 words (g/100) was used as the dependent variable for gesture production. Gestures were identified and coded as one of four types, based on McNeill's (2005) scheme: iconic, metaphorical, beat, or deictic. Iconic gestures ($n = 147$; 28 %) were defined as gestures that depict concrete referents (e.g., a flapping motion to convey a bird). Metaphorical gestures ($n = 156$; 29 %) were defined as gestures that depict abstract referents metaphorically (e.g., moving a fist outward to indicate forward motion). Beat gestures ($n = 121$; 23 %) were defined as simple, rhythmic gestures that do not depict semantic content. Deictic gestures ($n = 60$; 11 %) were defined as gestures directing attention through their form, usually consisting of an extended finger or hand. Several gestures ($n = 69$; 13 %) were classified as combinations of two of these four types (e.g., a combination iconic/beat gesture that represented the referent concretely and was moved in time to speech prosody); all possible combinations were represented. Combination gestures are not mutually exclusive with gestures from the other categories listed (iconic, metaphorical, beat, deictic), given that these gestures comprise gestures from two separate categories. Emblem gestures ($n = 5$; <1 %), defined as culturally specific, quotable gestures (thumbs up, peace, etc.) were also coded. Gestures that could not be classified as any of these types, due to reasons such as unintelligibility of accompanying speech or partially hidden hands, were classified as unclear ($n = 41$; 8 %). Neither emblems nor unclear gestures were analyzed due to their infrequency and functional ambiguity, respectively.

Gesture–Speech Synchrony

Temporal gesture–speech synchrony magnitude was calculated by identifying the lexical affiliate of each gesture (i.e., the word or phrase depicted by the gesture),¹ and measuring the difference in time in milliseconds (ms) between the beginning of the lexical affiliate and the beginning of the gesture's stroke (identifiable portion; i.e., timing of speech onset). Additionally, the number of gestures with an onset difference exceeding 200 ms from speech onset was counted. Temporal gesture–speech synchrony was classified as positive (gesture preceding speech) or negative (speech preceding gesture), in accordance with the direction of onset time differences. Several

attributes of representational gestures were also coded, including size, location, and semantic information (same or different from accompanying speech).

Ratings of Communicative Quality

All story retellings were rated holistically to assess two aspects of communicative quality: coherency and engagement. To evaluate these aspects of communication, the following questions were posed: How well could you follow the narrative? (coherency) and how engaging did you find the narrative? (engagement). Each retelling was assigned an integer score for each of these questions on a Likert scale ranging from (1) not at all to (7) extremely. Half of the retellings were rated based on both audio and video of the retellings, and half were rated based only on audio. (No significant differences in ratings based on presentation modality (audio/video vs. audio only) were found; thus, analyses reflect ratings collapsed across rating modality.) Two raters (different than the listeners and the coders who coded the gestures) rated all stories for communicative quality. These raters were also unaware of the experimental design, hypotheses, and participant diagnoses at the time of rating.

Interrater Reliability

Two coders independently transcribed and coded all of the speech and gestures produced by participants. To evaluate inter-rater reliability, 10 randomly selected participants' transcripts (5 ASD; 5 TD) were examined in both experimental conditions (visible listener; non-visible listener). For all continuous measures, a two way mixed intraclass correlation (ICC) was used to gauge interrater reliability for average measures. For speech, this analysis yielded scores of .99 for number of words spoken, .86 for utterance segmentation, .79 for timing of speech onset, and .90 for dysfluencies (discourse markers, silent pauses, revisions, and repetitions; range: .78–.97). For gesture, this analysis yielded scores of .98 for gesture identification, and scores of .89 for gesture onset and .95 for lexical affiliate onset synchrony. For the categorical measure of gesture classification, Cohen's Kappa was used to assess interrater reliability based on raw data. This analysis yielded a score of .61, indicating substantial agreement between raters (Landis and Koch 1977). Raters agreed 78 % in their identification of lexical affiliates of gestures.² These values are

¹ Lexical affiliates were identified for all gesture types (iconic, metaphorical, beat, deictic), which reduced interrater agreement compared to other similar studies (e.g., de Marchena and Eigsti 2010) where not all gesture types were scored.

² This value was lower than others because there were unlimited possibilities, rather than a fixed scale or alternatives to choose between. In order to be scored as in agreement, both coders' response were required to be identical.

generally consistent with those of other similar studies of gesture production in ASD (de Marchena and Eigsti 2010).

For communicative quality, inter-rater reliability was evaluated using ratings for the same subset of 10 participants that was randomly selected for evaluation of inter-rater reliability for the two primary coders (see above). For coherency, the ICC was .90, and for engagement, it was .87, indicating very high agreement in ratings between raters. Collapsed across questions, the ICC was .91 for ASD participants' retellings and .82 for TD participants' retellings, and was .90 for the visible listener condition and .88 for the non-visible listener condition, with no significant differences between groups.

Results

Differences by Diagnosis

We first examined the characteristics of speech produced by ASD and TD participants during cartoon retellings (see Table 2). Relative to TD adolescents, adolescents with ASD produced fewer words [mean ASD = 185.86 (SD = 84.34); mean TD = 245.35 (SD = 72.61); $F(1, 39) = 5.88, p = .02, \eta_p^2 = .13$], syllables [ASD = 228.81 (113.59); TD = 293.24 (93.50); $F(1, 39) = 4.55, p = .04, \eta_p^2 = .11$], and shorter utterances [ASD = 17.52 (9.22); TD = 26.92 (13.33); $F(1, 39) = 6.49, p = .02, \eta_p^2 = .14$]. We next examined the complexity of speech produced by TD and ASD participants, which we quantified using Flesch–Kincaid (F–K) grade level (simplicity as measured by words per sentence and syllables per word; (Kincaid et al. 1975) and lexical density (content per functional or lexical units) (Ure 1971).³ There was a nonsignificant trend for adolescents with ASD to produce less complex speech than TD adolescents [ASD = 6.40 (3.63); TD = 8.04 (5.41); $F(1, 39) = 2.59, p = .09, \eta_p^2 = .10$]; however, no difference between ASD and TD adolescents was found for lexical density [ASD = 63.31 (5.78); TD = 63.29 (9.32); $F < 1$]. Finally, we examined production of silent pauses and discourse markers (*um, uh*, etc.), the latter of which serve the pragmatic function of indicating an upcoming pause to listeners (Swerts et al. 1996). ASD adolescents produced fewer discourse markers [ASD = 3.89 (4.00); TD = 9.96 (7.59); $F(1, 39) = 9.43, p = .004, \eta_p^2 = .20$], and more silent pauses greater than 2 s [ASD = 2.81 (1.86); TD = 1.11 (1.18); $F(1, 39) = 12.66, p = .001, \eta_p^2 = .25$], supporting the findings of Lake et al. (2011). Together, these findings indicate that the speech of

adolescents with ASD is sparser, simpler, and contains more silent pauses and fewer discourse markers than the speech of TD adolescents.

We next examined gestures produced by ASD and TD adolescents during cartoon retellings. To control for the differences in verbosity between ASD and TD adolescents detailed above, the number of gestures produced per 100 words (g/100) was used as the dependent variable. For the amount of gesture produced, there was no main effect of diagnosis [ASD = 3.07 (3.42); TD = 4.68 (4.15); $F < 1$]. However, in contrast with previous work examining gesture production in adolescents with ASD (de Marchena and Eigsti 2010), we found an interaction of gesture type by diagnosis [$F(3, 117) = 2.74, p = .05, \eta_p^2 = .07$]. Specifically, adolescents with ASD produced fewer metaphorical (ASD = .48 (.63); TD = 1.08 (1.62); $t(39) = 1.99, p = .05, d = .49$), and beat [ASD = .26 (.60); TD = 1.14 (1.76); $t(39) = 2.13, p = .04, d = .67$], gestures than TD adolescents. No differences between ASD and TD adolescents were found for production of iconic [ASD = .79 (.25); TD = .77 (.25); $t < 1$] or deictic [ASD = .36 (.16); TD = .25 (.10); $t < 1$] gestures (see Fig. 1). Considered as a whole, these results indicate that adolescents with ASD produce fewer non-concrete (beat and metaphorical) gestures than TD adolescents, but that ASD and TD adolescents produce similar amounts of concrete (deictic and iconic) gestures.

We next examined temporal congruency between speech and gesture produced by participants. Temporal congruency was assessed in two ways: By examining the magnitude of the difference in timing between the strokes (i.e., meaningful portions) of representational gestures and their lexical affiliates (i.e., accompanying words or phrases expressing their meaning), and by counting the number of gestures with stroke onsets preceding or following lexical affiliate (speech) onsets by more than 200 ms. Consistent with the findings of Morrel-Samuels and Krauss (1992), on average, the strokes of all adolescents' gestures preceded lexical affiliates in time rather than following them [preceding = 270 ms (260 ms); following = 210 ms (330 ms); $F(1, 22) = 6.96, p = .01, \eta_p^2 = .15$], and were more likely to precede lexical affiliates, relative to strokes following lexical affiliates [preceding = 2.0 g/100 (2.0 g/100); following = 1.0 g/100 (1.0 g/100); $F(1, 22) = 5.06, p = .03, \eta_p^2 = .12$]. Although there were no main effects or interactions of diagnosis for the magnitude in time of temporal gesture asynchrony, adolescents with ASD produced a greater quantity of temporally asynchronous gestures than TD adolescents [ASD = 4.0 g/100 (4.0 g/100); TD = 2.0 g/100 (2.0 g/100); $F(1, 22) = 4.14, p = .05, \eta_p^2 = .10$], providing some support for the findings of de Marchena and Eigsti (2010). These results provide some evidence that the gestures of adolescents with ASD are

³ Text analysis measures were used as a metric of speech complexity because there are no comparable automated measures specifically designed for speech analysis.

Table 2 Production of various speech attributes by (and to) ASD and TD adolescents in the visible and non-visible listener conditions

| Statistics group | Descriptive | | | | Inferential | | | | |
|--------------------------------|--------------------------------|-------------------------------|--------------------------------|-------------------------------|--|--|--|--|--|
| | ASD | | TD | | Diagnosis | | Visibility | | Dx*Visibility <i>p</i> , effect size |
| | Visible <i>M</i> (SD) Range | Non-visible <i>M</i> Range | Visible <i>M</i> (SD) Range | Non-visible <i>M</i> Range | <i>p</i> , effect size | <i>p</i> , effect size | <i>p</i> , effect size | | |
| <i>Speech characteristics</i> | | | | | | | | | |
| Words spoken | 174.39 (91.69) 47–377 | 197.33 (87.73) 71–390 | 237.78 (72.70) 72–377 | 252.91 (86.85) 133–399 | $F(1, 39) = 5.88$ $p = .02, \eta^2 = .13$ | $F(1, 39) = 3.48$ $p = .07, \eta^2 = .08$ | $F < 1$ | $F < 1$ | $F < 1$ |
| Mean length of utterance (MLU) | 102.75 (96.45) 7.36–303 | 130.62 (105.19) 11.5–305 | 208.48 (100.80) 57.5–377 | 204.33 (99.92) 19–399 | $F(1, 39) = 9.98$ $p = .003, \eta^2 = .20$ | $F < 1$ | $F(1, 39) = 1.31$ $p > .10, \eta^2 = .03$ | $F < 1$ | $F < 1$ |
| Speech duration | 82.49 (41.14) 30.00–167.6 | 93.89 (46.08) 25.34–190.7 | 95.03 (32.91) 32.42–152.6 | 104.34 (34.91) 55.8–168.7 | $F(1, 39) = 1.01$ $p > .10, \eta^2 = .03$ | $F(1, 39) = 6.55$ $p = .01, \eta^2 = .14$ | $F < 1$ | $F < 1$ | $F < 1$ |
| Syllable count | 217.22 (112.32) 60–457 | 240.39 (114.87) 84–511 | 282.13 (83.19) 81–427 | 304.35 (103.80) 159–488 | $F(1, 39) = 4.55$ $p = .04, \eta^2 = .11$ | $F(1, 39) = 3.69$ $p = .06, \eta^2 = .09$ | $F < 1$ | $F < 1$ | $F(1, 39) = 7.17$ $p = .01, \eta^2 = .16$ |
| Words per sentence | 14.78 (8.19) 3.6–32.8 | 20.27 (11.72) 8.0–49.8 | 28.83 (16.45) 8.4–72.0 | 25.00 (12.90) 9.4–63.4 | $F(1, 39) = 6.49$ $p = .02, \eta^2 = .14$ | $F < 1$ | $F < 1$ | $F < 1$ | $F < 1$ |
| <i>Dysfluencies</i> | | | | | | | | | |
| Discourse markers | 3.17 (.82) 0–10 | 4.61 (1.28) 0–22 | 9.34 (1.62) 0–32 | 10.57 (1.75) 2–27 | $F(1, 39) = 9.43$ $p = .004, \eta^2 = .20$ | $F(1, 39) = 2.71$ $p = .11, \eta^2 = .07$ | $F < 1$ | $F < 1$ | $F < 1$ |
| Silent pauses | 2.78 (.53) 0–7 | 2.83 (.45) 0–6 | 1.13 (.36) 0–6 | 1.09 (.29) 0–6 | $F(1, 39) = 12.66$ $p = .001, \eta^2 = .25$ | $F < 1$ | $F < 1$ | $F < 1$ | $F < 1$ |
| Repetitions | .94 (0.26) 0–4 | 2.00 (.61) 0–7 | 1.09 (.28) 0–4 | 1.13 (.39) 0–8 | $F < 1$ $p = .05, \eta^2 = .09$ | $F(1, 39) = 3.97$ $p = .07, \eta^2 = .08$ | $F < 1$ | $F < 1$ | $F < 1$ |
| Revisions | 2.89 (.91) 0–16 | 3.0 (1.0) 0–15 | 2.13 (.28) 0–4 | 2.74 (.47) 0–7 | $F < 1$ $p > .10, \eta^2 = .03$ | $F(1, 39) = 1.01$ | $F < 1$ | $F < 1$ | $F < 1$ |
| <i>Speech complexity</i> | | | | | | | | | |
| F-K grade level | 4.58 (3.42) 0–12 | 6.48 (4.52) .9–17.9 | 9.46 (6.27) 2.2–25.8 | 8.12 (4.96) 2.9–23.4 | $F(1, 39) = 2.59$ $p = .09, \eta^2 = .10$ | $F < 1$ | $F < 1$ | $F(1, 39) = 3.68$ $p = .03, \eta^2 = .14$ | $F < 1$ |
| Lexical density | 66.68 (11.33) 47.4–93.1 | 61.82 (7.34) 46.0–78.4 | 63.09 (8.74) 51.1–81.3 | 62.30 (7.42) 49.7–77.5 | $F < 1$ | $F < 1$ | $F < 1$ | $F(1, 39) = 3.19$ $p = .05, \eta^2 = .12$ | $F < 1$ |

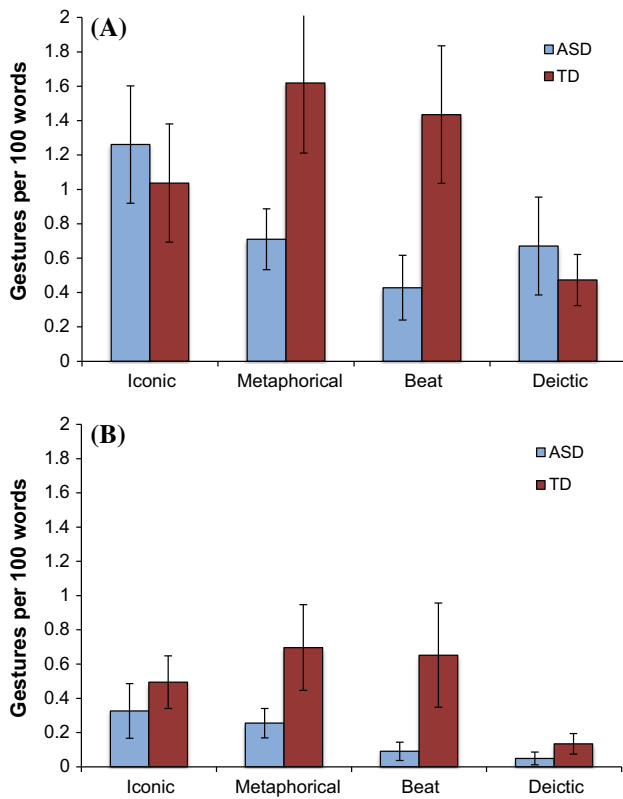


Fig. 1 Production of different gesture types by ASD and TD adolescents in the **a** visible listener and **b** non-visible listener conditions (error bars represent standard error)

more likely to be temporally incongruous with accompanying speech than the gestures of TD adolescents.

Finally, to gauge communicative quality, we examined coherency and engagement ratings for story retellings of ASD and TD adolescents. This analysis revealed that there was a nonsignificant trend for the retellings of adolescents with ASD to be perceived as less coherent than the retellings of TD adolescents [ASD = 3.65 (1.46); TD = 4.40 (1.00); $t(22) = 3.52, p = .07, d = .09$]. Furthermore, it revealed that the retellings of adolescents with ASD were perceived as less engaging than the retellings of TD adolescents [ASD = 3.41 (1.19); TD = 4.33 (.70); $t(22) = 11.42, p = .002, d = .23$]. These findings provide evidence that the adolescents with ASD communicate in a less engaging and coherent manner than their TD peers.

Differences by Listener Visibility

Next, we examined how gesture, speech, and communication differ as a function of listener visibility, and whether this effect varies by diagnosis. Consistent with previous work in TD adults (Alibali et al. 2001), the duration of all adolescents' speech was longer [visible = 89.53 s (36.81 s); non-visible = 99.75 s (40.00 s); $F(1, 39) = 6.55, p = .01,$

$\eta_p^2 = .14$] and showed a nonsignificant trend towards containing more words [visible = 209.95 words (86.57 words); non-visible = 228.51 words (90.55 words); $F(1, 39) = 3.48, p = .07, \eta_p^2 = .08$] in the presence of a non-visible listener than a visible listener (see Table 2). Moreover, there was a diagnosis by visibility interaction for the number of words produced per sentence. Simple main effects analysis revealed that adolescents with ASD produced more words per sentence in the presence of a non-visible listener than a visible listener [visible = 14.78 words (8.19 words); non-visible = 20.27 words (11.72 words); $F(1, 17) = 7.88, p = .01, \eta_p^2 = .32$], whereas no such difference was found for TD adolescents [visible = 28.83 words (16.45 words); non-visible = 25.00 words (12.90 words); $F(1, 17) = 7.88, p > .1$]. These results extend prior findings (Alibali et al. 2001; de Marchena and Eigsti 2010) by demonstrating that adolescents with ASD increase their speech to a greater extent in the presence of non-visible listeners than visible listeners.

Although speech complexity did not show a main effect of listener visibility, it showed a diagnosis by visibility interaction (see Table 2). Simple main effects analysis revealed that the speech complexity of adolescents with ASD differed significantly as a function of listener visibility, whereas that of TD adolescents did not. Specifically, speech produced by adolescents with ASD exhibited lower F–K grade level [visible = 4.58 (3.42); non-visible = 6.48 (4.52); $F(1, 17) = 3.68, p = .03, \eta_p^2 = .14$] and higher lexical density [visible = 66.68 (11.33); non-visible = 61.82 (7.34); $F(1, 17) = 3.19, p = .05, \eta_p^2 = .12$] in the presence of a visible listener than in the presence of a non-visible listener (see Fig. 2). These results indicate that adolescents with ASD produce simpler speech in the presence of a visible interlocutor than a non-visible interlocutor, whereas the complexity of TD adolescents' speech does not vary as a function of interlocutor visibility.

We next examined differences in the amount of gesture produced by ASD and TD adolescents during cartoon retellings as a function of interlocutor visibility. In terms of gesture, consistent with the findings of previous work examining gesture production in adults and children (Alibali and Don 2001; Alibali et al. 2001), all adolescents produced more gestures when speaking to a visible listener than a non-visible listener [visible = .97 g/100 (1.45 g/100); non-visible = .36 g/100 (.83 g/100), $F(1, 39) = 18.18, p < .001, \eta_p^2 = .32$; see Fig. 1]. However, overall gesture production did not show a diagnosis by interlocutor visibility interaction ($F < 1$). Considered in conjunction with the previous result concerning speech complexity, this result indicates that social context differentially affects the speech of adolescents with ASD more than their gesture production.

Fig. 2 Average Flesh–Kincaid grade level (a) and lexical density (b) of speech produced by ASD and TD adolescents (error bars represent standard error)

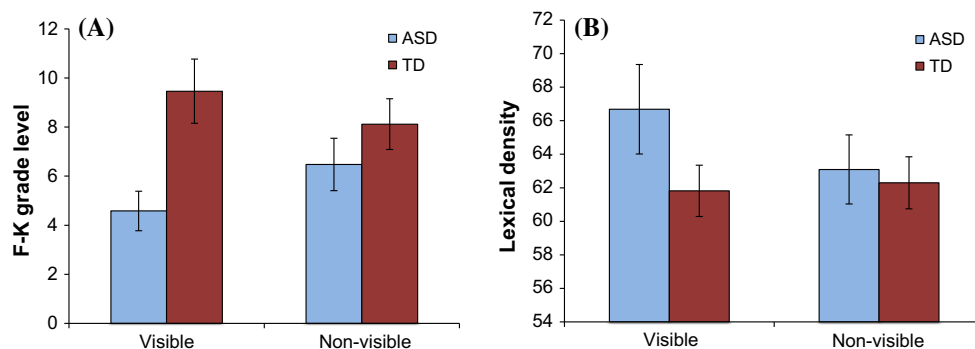


Table 3 Semantic and temporal gesture–speech synchrony by ASD and TD adolescents in the presence of a visible and non-visible listener

| Statistics group | Descriptive | | | | Inferential | | |
|---|--------------------------------|------------------------------------|--------------------------------|------------------------------------|--|--|--|
| | ASD | | TD | | Diagnosis <i>p</i> , effect size | Visibility <i>p</i> , effect size | Dx*Visibility <i>p</i> , effect size |
| Condition Factor | Visible <i>M</i> (SD) Range | Non-visible <i>M</i> (SD) Range | Visible <i>M</i> (SD) Range | Non-visible <i>M</i> (SD) Range | | | |
| <i>Semantic</i> | | | | | | | |
| Similar information (relative to speech) | 3.06 (3.28) 0–10 | 1.06 (1.70) 0–6 | 3.35 (3.83) 0–14 | 1.78 (2.21) 0–6 | $F < 1$ | $F(1, 39) = 10.89$ $p = .002, \eta_p^2 = .22$ | $F < 1$ |
| Different information (relative to speech) | .94 (1.11) 0–4 | .39 (.70) 0–2 | 3.30 (3.64) 0–12 | 1.18 (2.15) 0–9 | $F(1, 39) = 6.11$ $p = .02, \eta_p^2 = .14$ | $F(1, 39) = 14.13$ $p = .001, \eta_p^2 = .27$ | $F(1, 39) = 4.86$ $p = .03, \eta_p^2 = .11$ |
| <i>Temporal</i> | | | | | | | |
| Positive magnitude (gesture precedes speech) | .34 (.33) 0–.95 | .18 (.26) 0–.67 | .22 (.21) 0–.75 | .21 (.24) 0–.8 | $F < 1$ | $F(1, 39) = 2.41$ $p > .10, \eta_p^2 = .06$ | $F(1, 39) = 1.67$ $p > .10, \eta_p^2 = .04$ |
| Negative magnitude (gesture follows speech) | .17 (.20) 0–.70 | .29 (.44) 0–1.36 | .23 (.27) .23–.27 | .21 (.40) 0–1.67 | $F < 1$ | $F < 1$ | $F < 1$ |
| Positive difference count (gesture precedes speech) | 6.52 (5.78) 0–19 | 3.0 (3.84) 0–13 | 3.39 (3.16) 0–10 | 1.11 (1.71) 0–6 | $F(1, 39) = 6.08$ $p = .02, \eta_p^2 = .14$ | $F(1, 39) = 14.34$ $p = .001, \eta_p^2 = .27$ | $F < 1$ |
| Negative difference count (gesture follows speech) | 5.26 (5.55) 0–21 | 2.13 (2.70) 0–10 | 2.61 (3.01) 0–8 | .88 (1.45) 0–5 | $F(1, 39) = 4.33$ $p = .04, \eta_p^2 = .10$ | $F(1, 39) = 13.59$ $p = .001, \eta_p^2 = .26$ | $F(1, 39) = 1.15$ $p > .10, \eta_p^2 = .03$ |

All values represent gesture count

To more directly explicate the semantic relationship between speech and gestures produced by participants, we examined the timing of gesture and speech relative to one another, as well as how often gestures and speech explicitly described the same information relative to when they supplemented one another (see Table 3). For temporal gesture–speech synchrony, we found that all adolescents produced more temporally asynchronous gestures when speaking to a visible listener than a non-visible listener [visible = .04 g/100 (.04 g/100); non-visible = .02 g/100 (.03 g/100); $F(1, 22) = 14.19, p = .001, \eta_p^2 = .27$], and that this did not differ by diagnosis. For gestures conveying

the same information as speech, there was an effect of listener visibility, but no effect of diagnosis, or listener visibility by diagnosis interaction. However, for gestures conveying information differing from that conveyed via speech (e.g., supplementary information), there was a diagnosis by visibility interaction (see Table 3). A simple main effect analysis indicated that TD adolescents produced more gestures conveying different information in the presence of a visible listener relative to a non-visible listener [visible = .02 g/100 (.02 g/100); non-visible = .01 g/100 (.01 g/100); $F(1, 22) = 12.08, p = .002, \eta_p^2 = .35$], and that adolescents with ASD showed a

nonsignificant trend towards doing so [visible = .05 g/100 (.06 g/100); non-visible = .02 g/100 (.03 g/100); $F(1, 17) = 3.47, p = .08, \eta_p^2 = .17$]. These findings indicate that both TD and ASD adolescents produce more gestures conveying supplementary information in the presence of a visible listener, but that ASD adolescents increase their production of supplementary gestures to a lesser degree when in the presence of a visible listener compared to TD adolescents.

For communicative quality, we found an interaction of listener visibility by diagnosis for both coherency [$F(1, 22) = 15.18, p < .001, \eta_p^2 = .30$] and engagement [$F(1, 22) = 14.93, p < .001, \eta_p^2 = .29$]. Simple main effects analyses indicated that, in the visible listener condition, the retellings of TD adolescents were judged to be more coherent [ASD = 4.20 (1.48); TD = 5.21 (1.07); $t(22) = 3.13, p = .003, d = .98$] and engaging [ASD = 3.92 (1.21); TD = 5.22 (.65); $t(22) = 3.61, p = .001, d = 1.20$] than the retellings of adolescents with ASD. These findings are consistent with previous work examining communication in the presence of a visible listener, which showed that narratives produced by adolescents with ASD are perceived as lower in coherency and engagement than narratives produced by TD adolescents (de Marchena and Eigsti 2010). In contrast, no differences between the retellings of ASD and TD adolescents were found in the non-visible listener condition for coherency [ASD = 3.10 (1.44); TD = 3.60 (.93); $t(39) = 1.35, p > .10, d = .41$] or engagement [ASD = 2.90 (1.17); TD = 3.43 (.75); $t(39) = 1.77, p = .09, d = .54$]. Retellings in the visible condition were judged to be more coherent and engaging than retellings in the non-visible condition for adolescents with ASD [coherency: $t(17) = 3.86, p = .002, d = .67$; engagement: $t(22) = 4.82, p < .001, d = .86$], as well as TD adolescents [coherency: $t(22) = 9.59, p < .001, d = 1.61$; engagement: $t(22) = 12.21, p < .001, d = 2.55$; see Fig. 3]. The coherency of narratives produced by adolescents with ASD in the visible listener condition was negatively correlated with social reciprocity and

communication scores on the ADI-R ($r = -.71, p < .001$; $r = -.77, p < .001$); however, no such relationships were found for the retellings of adolescents with ASD in the non-visible listener condition ($r = -.18, p = .47$; $r = -.11, p = .67$). When contrasted directly, the correlation between coherency and the ADI-R communication score was significantly greater in the visible than the non-visible condition [$z(18) = 2.49, p = .01$]. Taken together, these results demonstrate that both ASD and TD adolescents are sensitive to listener visibility, and that this sensitivity is reflected to some extent in their gesture. Moreover, they indicate that TD adolescents produce narratives that are significantly more coherent and engaging in the presence of a visible listener than adolescents with ASD, but that there is little difference between the communicative quality of narratives between the groups when the listener is not visible. Finally, they suggest that the failure of the visible listener to elicit increased communicative quality in ASD is related to social and communicative symptoms in ASD.

To understand which factors contribute to communicative quality in the presence of visible and non-visible interlocutors, we averaged several factors together within three categories: speech related features, gesture related features, and speech–gesture related features (see Table 4 for factors included in each category). We then regressed the aggregate scores for these categories onto coherency and engagement for TD and ASD adolescents in the visible and non-visible listener condition. As can be seen from Table 4, in the presence of a visible listener, speech related features predicted coherency and engagement for TD adolescents [$t(22) = 2.16, p = .04$; $t(22) = 3.73, p = .001$], and for adolescents with ASD, they showed a non-significant trend towards predicting coherency and predicted engagement [$t(15) = 1.97, p = .07$; $t(15) = 2.81, p = .01$]. Notably, in the presence of a visible listener, gesture and speech–gesture related features predicted engagement for TD adolescents [$t(22) = 2.60, p = .02$; $t(22) = -2.32, p = .03$], but not for adolescents with ASD ($t < 1$; $t < 1$). On the other hand, in the presence of a non-

Fig. 3 Average coherency ratings (a) and engagement ratings (b) for retellings of ASD and TD adolescents in the visible and non-visible listener conditions (error bars represent standard error)

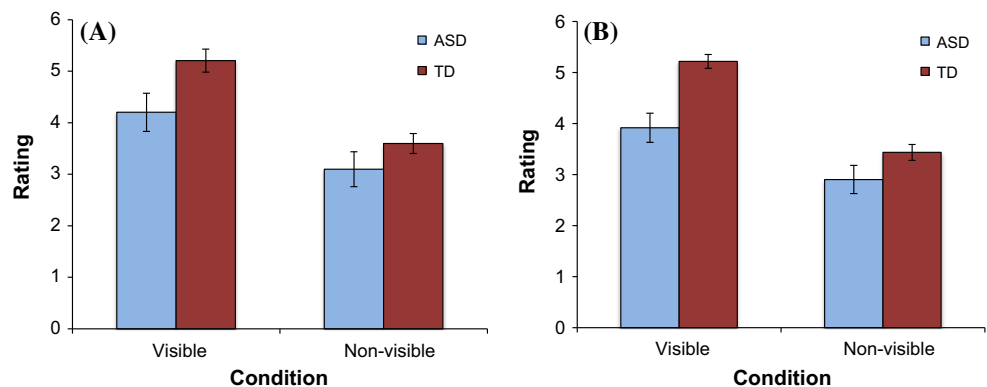


Table 4 Regression of speech, gesture, and speech–gesture related factors on coherency and engagement in TD and ASD adolescents in the presence of visible and non-visible listeners

| Variable | ASD, visible | | | ASD, non-visible | | | TD, visible | | | TD, non-visible | | |
|------------------------|--------------|-------------|-------------------|------------------|-------------|---------|-------------|-------------|---------|-----------------|-------------|---------|
| | <i>B</i> | SE <i>B</i> | β | <i>B</i> | SE <i>B</i> | β | <i>B</i> | SE <i>B</i> | β | <i>B</i> | SE <i>B</i> | β |
| <i>Coherency</i> | | | | | | | | | | | | |
| Speech related | .04 | .02 | 1.11 [†] | .01 | .01 | .40 | .01 | .01 | .51* | .001 | .01 | .03 |
| Gesture related | .77 | 1.15 | .46 | 1.56 | 4.36 | .23 | .29 | .64 | .28 | −1.18 | 1.00 | −.91 |
| Speech–gesture related | −1.08 | 1.10 | −.88 | −.39 | 1.53 | −.18 | −.03 | .28 | −.07 | .59 | .49 | .91 |
| <i>Engagement</i> | | | | | | | | | | | | |
| Speech related | .03 | .01 | 1.02** | .02 | .01 | .67* | .01 | .003 | .75*** | .001 | .01 | .07 |
| Gesture related | .52 | .63 | .37 | 3.33 | 2.86 | .60 | .88 | .34 | 1.37* | −.41 | .83 | −.38 |
| Speech–gesture related | −.52 | .61 | −.50 | −.96 | 1.00 | −.54 | .34 | .15 | −1.19* | .28 | .40 | .54 |

Speech related factors: number of words, number of syllables, speech duration, type:token ratio, mean length of utterance, F–K grade level, lexical density

Gesture related factors: iconic gestures, metaphorical gestures, beat gestures, deictic gestures

Speech–gesture related factors: number of gestures differing from words, number of gestures preceding speech, average amount of time of gesture preceding speech, number of gestures following speech, average amount of time of gesture following speech (note that positive values indicate greater gesture–speech discrepancies)

[†] $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

visible listener, speech related features predicted engagement for adolescents with ASD [$t(15) = 2.70$, $p = .02$], but not for TD adolescents ($t < 1$). These findings demonstrate that, in the presence of visible listeners, the speech, gesture, and speech–gesture relationships of TD adolescents play a key role in engagement, whereas only speech contributes to engagement for adolescents with ASD.

Discussion

The current study investigated communication impairments in ASD and their sources by manipulating social context during a story retelling task and examining the speech and gesture produced by adolescents with and without high functioning ASD. The results revealed that both ASD and TD adolescents produce more gestures in the presence of a visible than a non-visible listener, and more speech in the presence of a non-visible listener than a visible listener, indicating that adolescents with ASD are broadly aware of social context and adjust their gestures and speech in some similar ways as their TD peers. Nevertheless, communicative quality (i.e., coherency and engagement) increases for TD adolescents in the presence of a visible listener relative to a non-visible listener, whereas no such increase in communicative quality occurs for ASD adolescents. Notably, in the presence of a visible listener, these measures of communicative quality are related to the communicative and social symptoms in ASD. For TD adolescents, engagement is predicted by gesture

and gesture–speech relationships as well as speech in TD. In contrast, only speech related factors predicted engagement in adolescents with ASD, regardless of listener visibility. These findings suggest that the abnormalities in both the gesture and speech of high functioning individuals with ASD contribute to the communication deficits that they exhibit in face-to-face conversational settings, and that these abnormalities are reduced when the listener is not visible. Taken together, these results suggest that the communicative deficits observed in ASD adolescents are due to social context to a greater degree than impairments in core language function.

The findings concerning communicative quality and symptoms indicate that high functioning individuals with ASD are not able to enhance their communicative quality during face-to-face communication, and this impairment is particularly pronounced for individuals with more severe symptoms. Additionally, the results indicate that TD individuals use gesture and speech to enhance the engagement of their communication with visible listeners, whereas for individuals with ASD primarily only speech related to their communicative engagement in both the visible and non-visible conditions. These findings concerning communicative quality are consistent with research showing that individuals with ASD manage conversation topics more appropriately with non-human agents than with human agents (Tartaro and Cassell 2008). Furthermore, the findings suggest that, more so than speech, improper use of gesture during in-person interactions may be particularly relevant to communication deficits in high functioning ASD.

Regarding production of different gesture types, relative to their TD counterparts, adolescents with ASD produced similar amounts of iconic and deictic gestures, but fewer non-concrete (metaphorical and beat) gestures. This result complements the findings of deMarchena and Eigsti (2010), who found similar patterns of gesture production in adolescents with ASD, though they didn't reach significance in that previous study. When produced in conjunction with natural speech, iconic and deictic gestures illustrate concrete information conveyed via speech, whereas beat gestures supplement it by conveying prosodic cues. Metaphorical gestures can either illustrate or supplement information conveyed via speech, depending on the information that they convey (McNeill 2005). Thus, the gesture findings suggest that adolescents with ASD use gesture to illustrate concrete communication in a manner similar to their TD peers, but that they are less apt to use non-concrete (beat and metaphorical) gestures to supplement speech.

The findings concerning temporal and semantic relationships between gestures and their lexical affiliates provide evidence that major changes in the social context of communication, such as the presence or absence of a visible listener, affect the temporal and semantic links between gesture and speech in adolescents, regardless of whether they have been diagnosed with ASD. Nevertheless, the results revealed that adolescents with ASD produced a greater quantity of temporally asynchronous gestures than TD adolescents, regardless of listener visibility. This finding provides some support for previous work showing that the gestures of adolescents with ASD are more temporally asynchronous than the gestures of their TD counterparts (de Marchena and Eigsti 2010). Regarding production of different gesture types, adolescents with ASD produced gestures conveying the same information as accompanying speech at a rate similar to that of TD adolescents, but were less likely to produce gestures conveying different information than speech (see Table 3). Furthermore, ASD adolescents increased their production of gestures conveying different information in the presence of a visible listener to a lesser degree than TD adolescents. These results suggest that the improper use of gesture in ASD primarily reflects insufficient use of gesture to supplement meaning conveyed via speech. Taken together, the results concerning the temporal and semantic relationships between speech and gesture provide evidence demonstrating that gesture production is not as closely linked to speech production in ASD adolescents as it is in TD adolescents.

In terms of speech production, while adolescents with ASD were clearly responsive to social context, they produce less speech overall, as well as speech that is less complex, in the presence of visible but not non-visible listeners relative to TD adolescents. These findings

demonstrate that social context negatively impacts ASD adolescents' speech above and beyond pre-existing speech communication deficits in these areas. However, production of many dysfluencies, including discourse markers and silent pauses, did not differ as a function of listener visibility in ASD. Thus, only some speech-related deficits in ASD are exacerbated by the social context of communication. Although we did not examine whether the semantic content of speech differs between TD and ASD, this finding is notable when considered in conjunction with the finding that adolescents with ASD produced gestures that differed qualitatively in terms of their meaning relative to speech, particularly in the presence of visible listeners. These results provide further evidence that social context plays a greater role in the communication deficits characterizing ASD than language processing deficits in and of themselves.

Caveats of this study include that the results may be restricted to high-functioning individuals with ASD, who are capable of fluent communication via both speech and gesture, and may not apply to lower-functioning individuals with ASD. Furthermore, the results are based on an experimental narrative retelling task, rather than unstructured communication, which was necessary to manipulate the social context of communication in a controlled manner. Thus, future research should examine communication via speech and gesture in a more naturalistic context, and across a broader range of functioning levels in ASD to determine the robustness of the findings.

Taken together, the results of this study indicate that, unlike TD adolescents, adolescents with ASD encounter difficulty using speech and gesture to communicate effectively in the presence of visible listeners. Although not all factors reached significance, for the most part, differences in speech and gesture between ASD and TD were greater in scope and magnitude in the presence of a visible listener. Moreover, communicative coherence related to clinically relevant communication deficits in ASD to a significantly greater degree in the presence of a visible listener than a non-visible listener. These results suggest that social processing deficits contribute to communication deficits in fluent, high functioning ASD more than language deficits. This pattern of results contrasts with the new DSM-5 criteria for ASD, which now combine social and language deficits within a single diagnostic criterion. For high-functioning adolescents with ASD, the results presented here suggest that treatments addressing the impact of social context on speech and gesture production would particularly help to improve communication impairments.

Acknowledgments This research was supported by an institutional training fellowship "Training in Transformative Discovery in Psychiatry" (MH016804-31; Robert A. Sweet, PI) to L.M.M. and a

NARSAD Young Investigator Award from the Brain and Behavior Research Foundation to A.S.G. It was done at the University of Pittsburgh, supported by NIMH K01MH081191 (PI: O'Hearn), and NIH HD055748 (PI: Minshew) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The authors thank Andrew S. Lynn for assistance with data collection and Ashlie Caputo and Marco Pilotta for assistance with data coding, as well as the participants and their families.

Author Contributions L.M.M., K.O'H., B.L., & A.S.G. designed research; L.M.M. and K. O'H. performed research; L.M.M. and A.S.G. analyzed data; L.M.M. & A.S.G. wrote the paper.

References

- Alibali, M. W., & Don, L. S. (2001). Children's gestures are meant to be seen. *Gesture, 1*, 113–127. doi:10.1075/gest.1.2.02ali.
- Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169–188. doi:10.1006/jmla.2000.2752.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.
- Astington, J. W., & Baird, J. A. (2005). *Why language matters for theory of mind* (Vol. xii). New York: Oxford University Press.
- Attwood, A., Frith, U., & Hermelin, B. (1988). The understanding and use of interpersonal gestures by autistic and Down's syndrome children. *Journal of Autism and Developmental Disorders, 18*, 241–257. doi:10.1007/BF02211950.
- Baltaxe, C. A. (1977). Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology, 2*, 176–180. doi:10.1093/jpepsy/2.4.176.
- Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495–520. doi:10.1016/j.jml.2007.02.004.
- Camaioni, L., Perucchini, P., Muratori, F., & Milone, A. (1997). Brief report: A longitudinal examination of the communicative gestures deficit in young children with autism. *Journal of Autism and Developmental Disorders, 27*, 715–725. doi:10.1023/A:1025858917000.
- Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30*, 213–236. doi:10.1017/S0305000902005482.
- Colgan, S. E., Lanter, E., McComish, C., Watson, L. R., Crais, E. R., & Baranek, G. T. (2006). Analysis of social interaction gestures in infants with autism. *Child Neuropsychology, 12*, 307–319. doi:10.1080/09297040600701360.
- Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition, 113*, 98–104. doi:10.1016/j.cognition.2009.06.006.
- Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology, 40*, 271.
- de Marchena, A., & Eigsti, I. M. (2010). Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism Research, 3*, 311–322. doi:10.1002/aur.159.
- de Marchena, A., & Eigsti, I. M. (2015). The art of common ground: Emergence of a complex pragmatic language skill in adolescents with autism spectrum disorders. *Journal of Child Language, 42*, 1–15. doi:10.1017/S0305000915000070.
- Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American Journal of Psychology, 111*, 43–62. doi:10.2307/1423536.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*, 516–522. doi:10.1111/1467-9280.00395.
- Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*, 101–119. doi:10.1016/0010-0277(93)90026-R.
- Huerta, M., Bishop, S. L., Duncan, A., Hus, V., & Lord, C. (2014). Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. *American Journal of Psychiatry, 171*, 2020–2026. doi:10.1176/appi.ajp.2012.12020276.
- Kincaid, J. P., Fishburne, R. P., Jr, Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas for navy enlisted personnel (Research Branch Report 8-75)*. TN: Millington.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16–32. doi:10.1016/S0749-596X(02)00505-3.
- Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin & Review, 18*, 135–140. doi:10.3758/s13423-010-0037-x.
- Landa, R. (2000). Social language use in Asperger syndrome. In A. Klin, F. Volkmar, & S. Sparrow (Eds.), *Asperger syndrome* (pp. 125–155). New York: Guilford Press.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.
- Landry, S. H., & Loveland, K. A. (1989). The effect of social context on the functional communication skills of autistic children. *Journal of Autism and Developmental Disorders, 19*, 283–299. doi:10.1007/BF02211847.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr, Leventhal, B. L., DiLavore, P. C., et al. (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*, 205–223. doi:10.1023/A:1005592401947.
- Lord, C., Rutter, M., & Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685. doi:10.1007/BF02172145.
- Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of Autism and Developmental Disorders, 16*, 335–349. doi:10.1007/BF01531663.
- Loveland, K. A., Landry, S. H., Hughes, S. O., Hall, S. K., & McEvoy, R. E. (1988). Speech acts and the pragmatic deficits of autism. *Journal of Speech and Hearing Research, 31*, 593–604. doi:10.1044/jshr.3104.593.
- Martin, I., & McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with Asperger syndrome. *Journal of Autism and Developmental Disorders, 34*, 311–328. doi:10.1023/B:JADD.0000029553.52889.15.
- McCann, J., & Peppe, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language Disorders, 28*, 1–15. doi:10.1080/0308018021000165191.

- & *Communication Disorders*, 38, 325–350. doi:10.1080/1368282031000154204.
- McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language & Communication Disorders*, 42, 682–702. doi:10.1080/13682820601170102.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language*, 66, 249–264. doi:10.1016/j.jml.2011.07.004.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 615–622. doi:10.1037/0278-7393.18.3.615.
- Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, 20, 115–128. doi:10.1007/BF02206861.
- O'Hearn, K., Lakusta, L., Schroer, E., Minshew, N., & Luna, B. (2011). Deficits in adults with autism spectrum disorders when processing multiple objects in dynamic scenes. *Autism Research*, 4, 132–142. doi:10.1002/aur.179.
- O'Hearn, K., Schroer, E., Minshew, N., & Luna, B. (2010). Lack of developmental improvement on a face memory task during adolescence in autism. *Neuropsychologia*, 48, 3955–3960. doi:10.1016/j.neuropsychologia.2010.08.024.
- Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders*, 24, 247–257. doi:10.1007/BF02172225.
- Paul, R., Shriberg, L. D., McSweeney, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 35, 861–869. doi:10.1007/s10803-005-0031-8.
- Shriberg, L. D. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, 44, 1097–1115. doi:10.1044/1092-4388(2001/087).
- Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition*, 115, 380–393. doi:10.1016/j.cognition.2010.01.002.
- Stone, W. L., Ousley, O. Y., Yoder, P. J., Hogan, K. L., & Hepburn, S. L. (1997). Nonverbal communication in two- and three-year-old children with autism. *Journal of Autism and Developmental Disorders*, 27, 677–696.
- Swerts, M., Wichmann, A., & Beun, R.-J. (1996). Filled pauses as markers of discourse structure. In *Proceedings of the fourth international conference on spoken language, 1996. ICSLP 96*. (Vol. 2, pp. 1033–1036). IEEE.
- Tartaro, A., & Cassell, J. (2008). Playing with virtual peers: Bootstrapping contingent discourse in children with autism. In *Proceedings of the 8th international conference on international conference for the learning sciences* (Vol. 2, pp. 382–389). Retrieved from <http://dl.acm.org/citation.cfm?id=1599919>.
- Tomasello, M. (2009). *Constructing a language: A usage-based theory of language acquisition*. Harvard: Harvard University Press.
- Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443–452). Cambridge: Cambridge University Press.
- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. Psychological Corporation.