

Oral-Motor and Lexical Diversity During Naturalistic Conversations in Adults with Autism Spectrum Disorder

Julia Parish-Morris^{1,2}, Evangelos Sariyanidi¹, Casey Zampella¹, G. Keith Bartley¹, Emily Ferguson¹, Ashley A. Pallathra³, Leila Bateman¹, Samantha Plate¹, Meredith Cola¹, Juhi Pandey¹, Edward S. Brodtkin², Robert T. Schultz^{1,2,4}, Birkan Tunç^{1,2}

¹Center for Autism Research, Childrens Hospital of Philadelphia, Philadelphia PA, 19104, USA.

²Department of Psychiatry, University of Pennsylvania, Philadelphia PA, 19104, USA.

³Department of Psychology, The Catholic University of America, Washington DC, 20064, USA.

⁴Department of Pediatrics, University of Pennsylvania, Philadelphia PA 19104, USA.

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by impaired social communication and the presence of restricted, repetitive patterns of behaviors and interests. Prior research suggests that restricted patterns of behavior in ASD may be cross-domain phenomena that are evident in a variety of modalities. Computational studies of language in ASD provide support for the existence of an underlying dimension of restriction that emerges during a conversation. Similar evidence exists for restricted patterns of facial movement. Using tools from computational linguistics, computer vision, and information theory, this study tests whether cognitive-motor restriction can be detected across multiple behavioral domains in adults with ASD during a naturalistic conversation. Our methods identify restricted behavioral patterns, as measured by entropy in word use and mouth movement. Results suggest that adults with ASD produce significantly less diverse mouth movements and words than neurotypical adults, with an increased reliance on repeated patterns in both domains. The diversity values of the two domains are not significantly correlated, suggesting that they provide complementary information.

1 Introduction

Autism spectrum disorder (ASD) is a behaviorally-defined neurodevelopmental condition that affects approximately 1.5% of children in the U.S. (Christensen et al., 2016). Individuals with ASD are characterized by social communication impairments and the presence of restricted and repetitive patterns of interests and activities (APA, 2013). One of the most striking features of ASD is extreme heterogeneity in its clinical presentation. For example, verbal abilities in ASD range from minimally verbal (a few words

or sounds) to above average (Pickles et al., 2014). This heterogeneity makes it harder to diagnose ASD reliably, and indeed, expert clinicians may disagree about whether or not an individual meets criteria (Regier et al., 2013). Diagnostic challenges are compounded by shortcomings in current phenotyping approaches, which are either time-consuming and expensive, or provide limited information via questionnaires. Moreover, although ecologically valid stimuli have been shown to be superior for capturing ASD-related differences in behavior (Chevallier et al., 2015), most traditional ASD assessments continue to be conducted in highly controlled contexts. Taken together, these challenges highlight the need for a precision medicine approach to ASD (Beversdorf, 2016) that includes quantified and precise behavioral assessments in naturalistic settings.

Recent computational methodologies, including wearable technologies, computer vision, and natural language processing, have great potential to facilitate automated identification of novel phenotypic markers of behavior in ecologically valid settings, with exquisite precision, and in a highly scalable manner. Clinically, these technological advancements in “quantified behavior” could support diagnostic decision making, while providing critical information about intervention effectiveness.

In this study, we explore the applicability of computational behavioral assessments for identifying manifestations of the restricted/repetitive dimension in ASD. Building on existing knowledge about language production (Bone et al., 2013, 2014; Heeman et al., 2010; Tanaka et al., 2014; Van Santen et al., 2010; Goodkind et al., 2018; Parish-Morris et al., 2016b) and facial movements in ASD (Yirmiya et al., 1989; Borsos and Gyori, 2017; Guha et al., 2018; Owada et al., 2018), as well as the known interrelation between the

two domains (Busso and Narayanan, 2007), this study investigates patterns of word production and mouth movements during natural conversation. Our goal is to test whether an underlying dimension of cognitive-motor restriction can be detected across multiple behavioral domains in ASD.

Prior research suggests that restricted patterns of behavior may be cross-domain phenomena in autism, and are therefore evident in a variety of modalities. For example, computational studies of language in ASD provide support for the existence of multifaceted restricted language patterns that emerge during conversation. Children with ASD produce significantly more semantically overlapping turns than typically developing children during clinical evaluations (Rouhizadeh et al., 2015). They also engage in more echolalia (repetition of words or phrases) than typical children during semi-structured interviews (van Santen et al., 2013), and utilize a restricted range of narrative tools (Capps et al., 2000) and words (Baixauli et al., 2016) during storytelling. Less is known about linguistic diversity in adults with ASD, particularly during naturalistic conversations.

While similar evidence for atypical patterns of facial movement in ASD exists, most prior work has investigated facial expressions in the context of emotion recognition and imitation. Individuals with ASD produce flattened facial expressions (Yirmiya et al., 1989) that are hard to read (Brewer et al., 2016), and overt facial expression mimicry is impaired (Yoshimura et al., 2015). Reduced complexity in facial behavior, particularly in the eye region, while participants produced various facial expressions has been reported (Guha et al., 2018). Limited research, however, has examined facial expressions and oral-motor movement in dynamic social contexts such as conversations.

This study adds to the existing literature by combining tools from computational linguistics, computer vision, and information theory to characterize lexical and oral-motor diversity in adults with ASD. We demonstrate the utility of our approach in a young adult data set consisting of 44 conversational partners, 17 with ASD, in naturalistic social scenarios. Results showed that participants with ASD used fewer words than the typically developing (TD) control group during 3-minute “get to know you” conversations, and paused more. They also produced significantly

less diverse mouth movements and words, suggesting increased reliance on repeated patterns (i.e., restriction) in both domains. Notably, the correlation between the diversity values of the two domains was not significant, suggesting that they provide complementary information. The findings reported here suggest that reduced behavioral diversity, across domains, captures an underlying dimension of restriction and repetition in ASD that distinguishes individuals on the spectrum from typical controls. In the future, these methods could be utilized to identify and track highly quantifiable treatment targets, thus advancing the goal of precision medicine for autism.

2 Methods

2.1 Participants

Forty-four adults participated in the present study (ASD: N=17, TD: N=27, all native English speakers). Participant groups did not differ significantly on mean chronological age, full-scale IQ estimates (WASI-II) (Wechsler, 2011), verbal IQ estimates, or sex ratio (Table 1). There was a trend toward a difference in full-scale IQ, so this variable was considered in models comparing diagnostic groups. Participants were diagnosed using the Clinical Best Estimate process (Lord et al., 2012b), informed by the Autism Diagnostic Observation Schedule - 2nd Edition, Module 4 (ADOS-2) (Lord et al., 2012a) and adhering to DSM-V criteria for ASD (APA, 2013). All aspects of this study were approved by Institutional Review Boards of the University of Pennsylvania and the Children’s Hospital of Philadelphia.

2.2 Procedure

After providing written informed consent to participate in a novel social skills intervention (NIH R34MH104407, “Services to enhance social functioning in adults with autism spectrum disorders”, PI: Brodtkin) participants underwent a battery of tasks at three time points separated by approximately 6 months each. These tasks assessed social communication competence and included a slightly modified Contextual Assessment of Social Skills (CASS) (Ratto et al., 2011). The current analysis focuses on the third time point, after all participants with ASD received the social skills intervention. Typical participants did not receive intervention, and participated in the CASS once after providing informed consent.

Variable	ASD Mean (SD)	TD Mean (SD)	Statistics	<i>p</i> -value
Age (years)	26.9 (7.3)	28.1 (8.4)	W = 234	0.923
Sex (Male, Female)	15, 2	23, 4	χ^2 : 0.08	0.774
Full-Scale IQ	102.1 (19.8)	111.7 (9.5)	W = 157	0.080
Verbal IQ	112.6 (22.1)	112.4 (11.2)	W = 215	0.736
ADOS Total	13.1 (3.0)	1.1 (0.9)	W = 442	< 2e-8*
ADOS Social Affect	9.8 (2.3)	1.0 (0.9)	W = 442	< 1e-8*
ADOS RRB	3.3 (1.5)	0.1 (0.3)	W = 441	< 1e-9*

Table 1: Demographics of participants in our sample. Wilcoxon rank sum tests with continuity correction were used for statistical group comparisons, except for sex ratios where a Chi-squared was used. One TD participant had missing ADOS-2 scores. RRB=Repetitive Behaviors and Restricted Interests subscore of the ADOS-2. * Statistically significant difference between diagnostic groups, $p < 0.05$.

The CASS is a semi-structured assessment of conversational ability designed to mimic real-life first-time encounters. Participants engaged in two 3-minute face-to-face conversations with two different confederates (research staff, blind to participant diagnostic status and unaware of the dependent variables of interest). In the first conversation (interested condition), the confederate demonstrated social interest by engaging both verbally and non-verbally in the conversation. In the second conversation (bored condition), the confederate indicated boredom and disengagement both verbally (e.g., one-word answers, limited follow-up questions) and physically (e.g., neutral affect, limited eye-contact and gestures). The current analysis is based on the interested condition only. Prior to each conversation, study staff provided the following prompt to the participants and confederates before leaving the room: “Thank you both so much for coming in today. Right now, you will have 3 minutes to talk and get to know each other, and then I will come back into the room.”

CASS confederates included 10 undergraduate students or BA-level research assistants (3 males, 7 females, all native English speakers). Confederates were semi-randomly selected, based on availability and clinical judgment (4 confederates interacted with the ASD group, 8 with the TD group, 2 with both). In order to provide opportunities for participants to initiate and develop the conversation, confederates were trained to speak for no more than 50% of the time and to wait 10s to initiate the conversation. If conversational pauses occurred, confederates were trained to wait 5s before re-initiating the conversation. No specific prompts were provided to either speaker.

Audio and video of the CASS was recorded using a specialized “TreeCam”, built in-house (Fig-

ure 1), that was placed between the participant and confederate on a floor stand. This device has two HD video cameras pointing in opposite directions to allow simultaneous recording of the participant and the confederate as they sit facing each other, with a central microphone to record audio. For the face analysis, the first 10 seconds of the video were cropped to remove RA instructions (which may have also removed a few seconds of the CASS), and recordings continued for 3 minutes. For the lexical analysis, the sample began when the first word of the CASS was uttered, after study staff left the room, and ended when study staff re-entered.

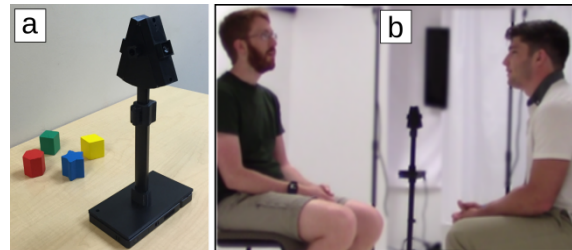


Figure 1: (a) The TreeCam video/audio capturing device. (b) Illustration of the task environment. Participants and confederates sat face-to-face while engaging in a “get to know each other” dialogue, with the TreeCam placed in between.

2.3 Processing of Language Data

Audio streams were extracted from video recordings, and saved in lossless .flac format. A team of reliable annotators produced time-aligned, verbatim, orthographic transcripts of the recordings in XTrans (Glenn et al., 2009). Each recording was processed by two junior annotators and one senior annotator, all of whom were undergraduate students and native English speakers. Before

becoming junior annotators for this cohort, each team member received at least 10 hours of training in Quick Transcription (Kimball et al., 2004) modified for use with clinical interviews of participants with ASD (Parish-Morris et al., 2016b,a, 2017). In addition, annotators were trained to reliability (defined as >90% in common with a Gold Standard transcript) on segmenting (marking speech start and stop times) and transcribing (writing down words and sounds produced, using the modified Quick Transcription specification). Training files included audio recordings of conversations between individuals with and without autism that were not used in this study. For the CASS, one reliable junior annotator segmented utterances into pause groups, while the second transcribed words produced by each speaker. A senior annotator then thoroughly reviewed and corrected each file (Figure 2). All senior annotators had at least 6 months of prior transcription experience. Final language data were exported from XTrans as tab-delimited files that were batch imported into R. Annotations marking non-speech sounds like laughter, indicators of language errors like stutters, and punctuation were removed, while other disfluencies (including filled pauses and whole-word repetitions) were left in. Total words, speech rate (total words/total length of speaking segments), sum of participant response latencies (Confederate-to-Participant inter-turn pauses or C2P; overlaps excluded), and number of conversational turns were calculated across each session.

2.4 Processing of Vision Data

CASS videos were processed by an image processing and feature extraction pipeline that included face detection, face registration, and facial movement quantification.

For face detection and localization of multiple facial landmarks (eyes, lip corners, nose etc.) within each face, we used a publicly available tool (OpenFace) (Baltrusaitis et al., 2016). The computation of facial movements requires image registration across frames, which we achieved via part-based registration (Sariyanidi et al., 2015). Using landmarks from the corners of the eyes and mouths at each frame, we subdivided the face into three overlapping parts covering the left eye region, the right eye region, and the mouth region (see Figure 3). Cropped sequences had visible jitter due to imprecise landmark localization at each frame, which

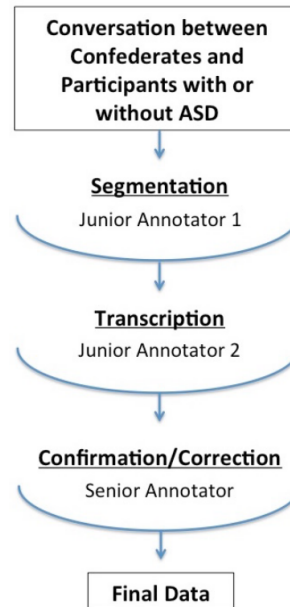


Figure 2: An illustration of the workflow for language processing.

is detrimental to the analysis of subtle face/head movements. We eliminated jitter using a video stabilization technique (Sariyanidi et al., 2017), which registers consecutive frames to one another.

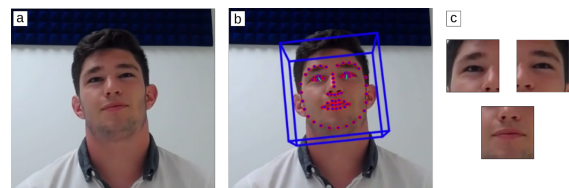


Figure 3: Illustration of the computer vision preprocessing pipeline. (a) Input video frames include the upper body of a participant/confederate. (b) Using OpenFace, faces are automatically detected and annotated with specific landmarks. (c) Faces are divided into three overlapping parts covering the left eye region, the right eye region, and the mouth region.

Quantification of facial movements was done using the Facial Bases method (Sariyanidi et al., 2017). This method uses 180 facial movement basis functions, 60 of which correspond to mouth movements. Each basis provides differential information (i.e. change of appearance) about a movement that occurs in a particular region of the face. Most bases are semantically interpretable; for example, one basis is activated when the lip corner of the subject moves upwards/downwards and another basis is activated when the subject’s lower lip moves, which typically occurs when the subject is talking (Figure 4). In this study, we used

the 60 bases corresponding to mouth movements. The entire video sequence of a participant was represented as a collection of 60 time series, where each time series quantified the activation level of one basis over time (Figure 4). In our analyses, we only used time points when participants were speaking.

Each time series underwent smoothing, peak detection, and normalization steps for reliability and comparability between participants and across 60 bases. We first smoothed each time series using a Gaussian filter with a filter width of 2 standard deviations. We then detected peaks by determining the time points of sign change in the first derivative (i.e. the point at which an increase in activation stops and a decrease begins).

Each facial basis may have a different maximum activation magnitude (Sariyanidi et al., 2017). We therefore normalized the heights of detected peaks via z-normalization, by using the time series from research confederates to calculate the mean activation and the standard deviation for each basis. Finally, we removed outlier peaks by setting activations whose absolute value is above 6 standard deviations to zero.

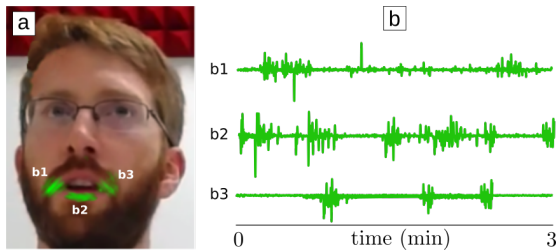


Figure 4: Quantification of mouth movements. (a) Example facial bases that explain mouth movements are highlighted. (b) Illustration of the bases as time series. Their activations quantify the mouth movements throughout the video.

2.5 Computation of Diversity

For both modalities (language and mouth movements) we quantified diversity using Shannon entropy (Cover and Thomas, 2006). From an information theoretical perspective, entropy can be described as the amount of information a data modality carries. Intuitively, one expects a higher entropy (diversity) when, for instance, a participant makes a rich set of facial expressions while speaking compared to a participant who generates only a restricted set of mouth movements. Similarly in the cognitive domain, higher lexical entropy (di-

versity) is expected when participants use a variety of words, and lower entropy is expected when participants produce repetitive speech. Shannon entropy (H) is calculated as

$$H = - \sum_{i=1}^n p(x_i) \log_b p(x_i)$$

where b is the base of the logarithm. In this work we used $b = 2$, yielding a measure of entropy in *bits*. The probability of generating a word x_i (or activation of a facial basis), $p(x_i)$, is calculated from the sample of generated words (or basis activations).

The ‘diversity’ function of the ‘qdap’ package in R (R Core Team, 2017) was used to calculate lexical (word-level) entropy for each participant. This function counts the number of different words produced by each participant, resulting in a vector of word counts. The probability of each word, $p(x_i)$, is then calculated by dividing its count by the total number of word counts. Note that the possible number of words and the exact words used by a participant can differ from one participant to other. Therefore, we also tested whether calculated entropy values were affected by total word counts (see Results).

For mouth movements, all participants were assessed using the same set of 60 bases. We calculated the number of times each facial basis was activated (similar to word counts), also taking into account the magnitude of activation, by calculating the sum of the entire time series. Note that the summation of positive and negative values in a time series should be zero, since a basis activation (i.e., a positive value) is followed by a deactivation (a negative value). For example, when a lip corner is stretched, it is then relaxed. Therefore, instead of summing the raw values of the time series, we summed the positive and the absolute of negative values separately, taking the average as our final count value. We repeated this procedure for all 60 bases, yielding a vector of movement counts.

Different facial bases may have different expected activation patterns, with some of them activated more frequently than others naturally. We therefore normalized the total activation count of each basis by the maximum count that was observed for the same basis of research confederates. Finally, entropy was calculated using the normalized counts.

2.6 Statistical Analysis

Our research design included repeated confederates across participants (i.e., the same 10 confederates joined multiple conversations with different ASD and TD participants). In order to account for this nested design when assessing group differences in diversity values (ASD vs TD), we began by using linear mixed effects models that included confederate ID as a random effect (function ‘lmer’ from package ‘lme4’ in R) (R Core Team, 2017; Bates et al., 2015).

We measured the contribution of random effects to the model by comparing the conditional and marginal coefficients of determination, using ‘MuMIn’ package and ‘r.squaredGLMM’ function. The conditional and marginal coefficients of determination correspond to variance explained by fixed effects alone and variance explained by both fixed and random effects, respectively. When there was no difference between the two models (i.e. random effects did not contribute to model fit), we also fit ordinary linear regression models using the ‘lm’ function. Due to our small sample size ($n = 44$), simpler models were used when possible, to preserve degrees of freedom.

The ASD and TD groups did not differ significantly on mean age, sex ratio, or verbal IQ estimates, but there was a trend toward a difference in full-scale IQ (Table 1). To gauge the robustness of diagnostic group differences and check for the utility of these variables as potential predictors, we also fit models that included sex, age, and IQ as covariates. For the analysis of mouth movements, we used speech length (the sum of participant speech segments) as a covariate; more movement is expected with longer talk times, which may impact diversity. The pipeline for mouth movements described above is sensitive to overall head movements since facial bases may be spuriously activated with head movement. Therefore, we quantified the average head movement of each participant (as provided by OpenFace), by measuring the total motion of the head center during the conversation, and used it as another covariate.

Effect sizes for group differences are reported using Cohen’s d . We calculated Cohen’s d by dividing the estimated coefficient of the diagnostic variable (0: TD, 1: ASD) in the fitted model (lmer or lm) by the pooled standard deviation of the diversity value (i.e. average standard deviation of ASD and TD groups). Following (Cohen, 1988),

d values between 0.20 and 0.50 reflect a small effect, between 0.50 and 0.80 a medium effect, and > 0.80 a large effect.

Agreement between lexical and mouth movement diversity was measured using Spearman’s rank-order correlation coefficient.

3 Results

3.1 Basic Conversational Differences

Preliminary analyses revealed that conversations differed on a variety of basic linguistic features, according to the diagnosis of the Participant (Table 2; t -values of the main effect of diagnosis are reported; the random effect of confederate ID contributed only to models for confederate word count and conversational turns; ordinary linear models are reported for all other variables). Conversational length did not differ for ASD and TD participants, which was expected given the controlled 3-minute task design. Confederates in each condition produced the same number of words regardless of the diagnosis of their conversational partners. However, participants with ASD produced fewer words than TD participants ($p = 0.002$), and conversational partners exchanged marginally fewer turns when the participant had ASD ($p = 0.10$). Participant groups did not differ on speech rate, but the ASD group had a significantly larger sum of Confederate-to-Participant (C2P) pauses than the TD group. These results demonstrate that participants with ASD produced fewer words and longer pauses than TD participants during the CASS, and trended toward engaging in fewer conversational turns despite comparable task duration.

3.2 Lexical Diversity

Preliminary analyses revealed that inclusion of confederate ID as a random effect did not significantly improve model fit for lexical diversity in any model that included diagnosis as a fixed effect; we therefore report ordinary linear models.

A simple linear model revealed significantly reduced lexical diversity in participants with ASD (Mean=4.50, SD=0.22) as compared to TD participants (Mean=4.64, SD=0.12; $t(42)=2.85$, $p=0.007$, Cohen’s $d=0.82$). The effect of diagnosis on diversity continued to be significant after accounting for age, IQ, and gender ($t(39)=3.25$, $p=0.002$). Diversity of confederate language did not differ by participant diagnosis ($t(35.26)=0.17$,

$p=0.86$), suggesting that the effect of diagnosis on diversity in participants is driven by internal participant-level variables and not by differences in confederate language.

Given the expected (neurotypical) association between word count and entropy (Witten and Bell, 1990; Shannon, 1951), a second model was constructed that included word count, diagnosis, and the interaction between word count and diagnosis as predictors of participant lexical diversity. A significant interaction was revealed (Table 3), such that the slope of the relationship between word count and diversity was greater in the TD group than the ASD group (Figure 5).

3.3 Diversity of Mouth Movements

The random effect of confederate ID did not contribute to model fit when predicting mouth movement diversity; therefore, we report the results of ordinary linear models.

Mirroring our language findings, we observed a significant decrease in mouth movement diversity in the ASD group as compared to the TD group (Cohen’s $d=1.0$, $t=-2.73$, $p=0.009$) in a model using head movement and speech length as covariates. This difference remained significant when age, sex, and IQ were included as covariates (Cohen’s $d=1.0$, $t=-2.52$, $p=0.016$). None of the covariates contributed significantly to the model. In contrast to the observed relationship between word count and word diversity (Table 3), there was no significant relationship between speech length and mouth movement diversity ($t=0.50$, $p=0.619$).

3.4 Correlations Between Language and Mouth Modalities

We also investigated whether the two modalities (mouth movement and words produced) provided redundant information when characterizing ASD-related restriction in oral-motor and linguistic diversity. The diversity values of the two modalities were not significantly correlated in the ASD group (Spearman’s $r=-0.08$, $p=0.758$), in the TD group (Spearman’s $r=-0.11$, $p=0.566$), or across the sample as a whole (Spearman’s $r=0.18$, $p=0.240$). This suggests that lexical and oral-motor diversity provide unique information, and could potentially account for independent variance in future models designed to predict restricted interests/repetitive behaviors in ASD.

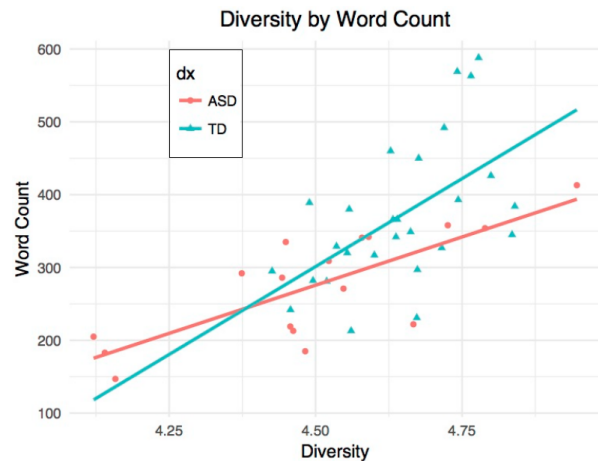


Figure 5: The relationship between word count and linguistic diversity differed by diagnostic status, with a steeper slope in the TD group than the ASD group.

4 Discussion

In this study, we identified medium-to-large group differences in behavioral entropy in adults with ASD vs. neurotypical adults, specifically in the areas of word production and mouth movement. This study is the first to use both computer vision and computational linguistics to show a “restricted” dimension in adult conversations with non-clinicians (most prior research used children’s interactions with psychologists during semi-structured clinical evaluations) (Rouhizadeh et al., 2015; van Santen et al., 2013).

In addition to basic group differences, our results revealed a novel interactive effect of word count and diagnosis on lexical diversity. As increasing numbers of words were produced by participants with ASD, they did not reach the same levels of linguistic diversity as their non-ASD peers. Indeed, this gap may widen over the course of longer conversations, and may differ by word category (e.g., function words vs. content words). We will explore these questions in future research with longer samples, wherein we evaluate the relationship between relatively deteriorated linguistic diversity and impressions of social communication ability by gathering post-conversational ratings of social communication quality from confederates.

Our finding that mouth movements are less diverse in ASD is also novel. One possible explanation for this finding is subtle oral-motor impairments in the ASD sample, as children with ASD have been reported to have oral-motor deficits

	ASD Mean (SD)	TD Mean (SD)	<i>t</i> -value	<i>p</i> -value
Duration (mins)	3.08 (0.12)	3.16 (0.23)	1.43	0.16
Word Count	Part: 275 (77)	370 (99)	3.37	0.002*
	Conf: 226 (67)	236 (55)	1.08	0.29
Conversational Turns	33.53 (8.90)	38.56 (8.63)	1.70	0.10
Speech Rate	190.31 (26.29)	193.53 (24.88)	0.41	0.69
Sum of C2P Pauses	12.60 (5.36)	7.33 (3.61)	3.91	0.001*

Table 2: Basic group differences between conversations that did or did not include participants with ASD.

* Statistically significant difference between diagnostic groups, $p < 0.05$.

Variable	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value
(Intercept)	3.882	0.11	35.27	<2.00E-16
WC	0.002	0	5.79	<.001*
Diagnosis	0.519	0.142	3.66	<.001*
WC:Diagnosis	-0.002	0	-3.51	0.001*

Table 3: Linear model to predict linguistic diversity. Model includes word count (WC), participant diagnosis (TD coded as 1) and the interaction between word count and diagnosis. * Statistically significant difference between diagnostic groups, $p < 0.05$.

(Adams, 1998), and oral-motor abilities in infancy and toddlerhood predict later speech fluency (Gernsbacher et al., 2008). However, all participants in this study were fluent English speakers without overt oral-motor impairments. Reduced phonological diversity could also result in restricted mouth movements, a hypothesis that will be explored in future analyses.

Reduced facial expressiveness (McIntosh et al., 2006), atypical expressiveness (Samad et al., 2018; Loveland et al., 1994), and limited integration of expressions and vocalizations (Lord et al., 2012a) have all been reported in ASD, which could lead to reduced diversity in mouth movements. Typically, when people take part in a conversation, vocalizations are accompanied by subtle changes in facial expressions (Busso and Narayanan, 2007). Integration across different modalities (e.g., language and facial expressions) is a critical aspect of social communication, and impairment in this area is assessed in common diagnostic instruments for ASD, such as the ADOS (Lord et al., 2012a). However, to the best of our knowledge, there are no objective methods for directly quantifying the degree to which such integration occurs during natural conversations. Development of novel computational tools to fill this gap is an especially promising future direction.

Of clinical note, adults with ASD who participated in our study had just completed an intensive intervention to improve social interaction skills.

It is striking that decreased entropy was evident across domains in this sample, despite the recent intervention that targeted social reciprocity and conversational skills. This suggests that our results may in fact underestimate the magnitude of differences that could be present in untreated individuals.

5 Conclusion

Adults with ASD exhibit restricted/repetitive patterns of behavior (APA, 2013), but computational efforts to quantify the restricted/repetitive dimension in real-world contexts are just beginning to emerge (Rouhizadeh et al., 2015; Bone et al., 2015; Goodwin et al., 2014). This knowledge gap makes adult impairments difficult to treat, and tracking the effectiveness of interventions that target RRBs is a significant challenge for clinicians and researchers. Our results suggest that cross-domain entropy during naturalistic conversations could serve as a quantitative behavioral marker of ASD.

This study advances the field by applying computational methods across oral-motor and lexical domains, to identify restricted patterns of behavior in ASD in real-world contexts. In future research, we will explore relationships between reduced behavioral diversity and clinical phenotype, with the goal of moving beyond group differences to predict individual variability, and establishing external validity with established measures. We

envision that future iterations of the methods described here will be utilized to identify and track highly quantifiable treatment targets in the area of restricted/repetitive behaviors, and will advance the goal of precision medicine for individuals with autism and their families.

References

- Lynn Adams. 1998. [Oral-Motor and Motor-Speech Characteristics of Children with Autism](#). *Focus on Autism and Other Developmental Disabilities*, 13(2):108–112.
- APA. 2013. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5*. American Psychiatric Association, Washington, D.C.
- Inmaculada Baixauli, Carla Colomer, Belén Roselló, and Ana Miranda. 2016. [Narratives of children with high-functioning autism spectrum disorder: A meta-analysis](#). *Research in Developmental Disabilities*, 59:234–254.
- Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency. 2016. [OpenFace: An open source facial behavior analysis toolkit](#). In *Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE.
- Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. [Fitting Linear Mixed-Effects Models Using {lme4}](#). *Journal of Statistical Software*, 67(1):1–48.
- David Q Beversdorf. 2016. [Phenotyping, Etiological Factors, and Biomarkers: Toward Precision Medicine in Autism Spectrum Disorders](#). *Journal of developmental and behavioral pediatrics : JDBP*, 37(8):659–73.
- Daniel Bone, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan. 2015. [Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises](#). *Journal of Autism and Developmental Disorders*, 45(5):1121–1136.
- Daniel Bone, Chi-Chun Lee, Matthew P Black, Marian E Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan. 2014. [The psychologist as an interlocutor in autism spectrum disorder assessment: insights from a study of spontaneous prosody](#). *Journal of speech, language, and hearing research : JSLHR*, 57(4):1162–77.
- Daniel Bone, Chi-Chun Lee, Theodora Chaspari, Matthew P Black, Marian E Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan. 2013. [Acoustic-prosodic, turn-taking, and language cues in child-psychologist interactions for varying social demand](#). In *INTERSPEECH*, pages 2400–2404.
- Zsófia Borsos and Miklos Gyori. 2017. [Can Automated Facial Expression Analysis Show Differences Between Autism and Typical Functioning?](#) *Studies in health technology and informatics*, 242:797–804.
- Rebecca Brewer, Federica Biotti, Caroline Catmur, Clare Press, Francesca Happé, Richard Cook, and Geoffrey Bird. 2016. [Can Neurotypical Individuals Read Autistic Facial Expressions? Atypical Production of Emotional Facial Expressions in Autism Spectrum Disorders](#). *Autism Research*, 9(2):262–271.
- Carlos Busso and Shrikanth S. Narayanan. 2007. [Interrelation Between Speech and Facial Gestures in Emotional Utterances: A Single Subject Study](#). *IEEE Transactions on Audio, Speech and Language Processing*, 15(8):2331–2347.
- L Capps, M Losh, and C Thurber. 2000. ["The frog ate the bug and made his mouth sad": narrative competence in children with autism](#). *Journal of abnormal child psychology*, 28(2):193–204.
- Coralie Chevallier, Julia Parish-Morris, Alana McVey, Keiran M Rump, Noah J Sasson, John D Herrington, and Robert T Schultz. 2015. [Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters](#). *Autism research : official journal of the International Society for Autism Research*, 8(5):620–8.
- Deborah L Christensen, Jon Baio, Kim Van Naarden Braun, Deborah Bilder, Jane Charles, John N Constantino, Julie Daniels, Maureen S Durkin, Robert T Fitzgerald, Margaret Kurzius-Spencer, Li-Ching Lee, Sydney Pettygrove, Cordelia Robinson, Eldon Schulz, Chris Wells, Martha S Wingate, Walter Zahorodny, Marshalyn Yeargin-Allsopp, and Centers for Disease Control and Prevention (CDC). 2016. [Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012](#). *Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002)*, 65(3):1–23.
- J. Cohen. 1988. *Statistical power analyses for the social sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- T. M. Cover and Joy A. Thomas. 2006. *Elements of information theory*. Wiley-Interscience.
- Morton Ann Gernsbacher, Eve A. Sauer, Heather M. Geye, Emily K. Schweigert, and H. Hill Goldsmith. 2008. [Infant and toddler oral- and manual-motor skills predict later speech fluency in autism](#). *Journal of Child Psychology and Psychiatry*, 49(1):43–50.
- Meghan Lammie Glenn, Stephanie Strassel, and Haejoong Lee. 2009. [XTrans: a speech annotation and transcription tool](#). In *INTERSPEECH*, pages 2855–2858.

- Adam Goodkind, Michelle Lee, Gary E Martin, Molly Losh, and Klinton Bicknell. 2018. [Detecting language impairments in autism: A computational analysis of semi-structured conversations with vector semantics](#). In *Society for Computation in Linguistics (SCiL)*, pages 12–22.
- Matthew S. Goodwin, Marzieh Haghighi, Qu Tang, Murat Akcakaya, Deniz Erdogmus, and Stephen Intille. 2014. [Moving towards a real-time system for automatically recognizing stereotypical motor movements in individuals on the autism spectrum using wireless accelerometry](#). In *International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pages 861–872, New York.
- Tanaya Guha, Zhaojun Yang, Ruth B. Grossman, and Shrikanth S. Narayanan. 2018. [A Computational Study of Expressive Facial Dynamics in Children with Autism](#). *IEEE Transactions on Affective Computing*, 9(1):14–20.
- Peter A. Heeman, Rebecca Lunsford, Ethan Selfridge, Lois Black, and Jan Van Santen. 2010. [Autism and interactional aspects of dialogue](#). In *Annual Meeting of the Special Interest Group on Discourse and Dialogue, Association for Computational Linguistics*, pages 249–252.
- Owen Kimball, Chia-Lin Kao, Rukmini Iyer, Teodoro Arvizo, and John Makhoul. 2004. [Using Quick Transcriptions to Improve Conversational Speech Models](#). In *International Conference on Spoken Language Processing*.
- C. Lord, M. Rutter, P. S. DiLavore, S. Risi, K. Gotham, and S. L. Bishop. 2012a. *Autism diagnostic observation schedule, second edition (ADOS-2)*. Western Psychological Services, Torrance, CA.
- Catherine Lord, Eva Petkova, Vanessa Hus, Weijin Gan, Feihan Lu, Donna M. Martin, Opal Ousley, Lisa Guy, Raphael Bernier, Jennifer Gerdtts, Molly Algermissen, Agnes Whitaker, James S. Sutcliffe, Zachary Warren, Ami Klin, Celine Saulnier, Ellen Hanson, Rachel Hundley, Judith Piggot, Eric Fombonne, Mandy Steiman, Judith Miles, Stephen M. Kanne, Robin P. Goin-Kochel, Sarika U. Peters, Edwin H. Cook, Stephen Guter, Jennifer Tjernagel, Lee Anne Green-Snyder, Somer Bishop, Amy Esler, Katherine Gotham, Rhiannon Luyster, Fiona Miller, Jennifer Olson, Jennifer Richler, and Susan Risi. 2012b. [A Multisite Study of the Clinical Diagnosis of Different Autism Spectrum Disorders](#). *Archives of General Psychiatry*, 69(3):306.
- Katherine A. Loveland, Belgin Tunali-Kotoski, Deborah A. Pearson, Kristin A. Brelsford, Juliana Ortegon, and Richard Chen. 1994. [Imitation and expression of facial affect in autism](#). *Development and Psychopathology*, 6(03):433.
- Daniel N. McIntosh, Aimee Reichmann-Decker, Piotr Winkielman, and Julia L. Wilbarger. 2006. [When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism](#). *Developmental Science*, 9(3):295–302.
- Keiho Owada, Masaki Kojima, Walid Yassin, Miho Kuroda, Yuki Kawakubo, Hitoshi Kuwabara, Yukiko Kano, and Hidenori Yamasue. 2018. [Computer-analyzed facial expression as a surrogate marker for autism spectrum social core symptoms](#). *PLOS ONE*, 13(1):e0190442.
- Julia Parish-Morris, Christopher Cieri, Mark Liberman, Leila Bateman, Emily Ferguson, and Robert T Schultz. 2016a. [Building Language Resources for Exploring Autism Spectrum Disorders](#). In *Language Resources and Evaluation Conference*, pages 2100–2107. European Language Resources Association (ELRA).
- Julia Parish-Morris, Mark Liberman, Neville Ryant, Christopher Cieri, Leila Bateman, Emily Ferguson, and Robert Schultz. 2016b. [Exploring Autism Spectrum Disorders Using HLT](#). In *Workshop on Computational Linguistics and Clinical Psychology*, pages 74–84. Association for Computational Linguistics.
- Julia Parish-Morris, Mark Y Liberman, Christopher Cieri, John D Herrington, Benjamin E Yerys, Leila Bateman, Joseph Donaher, Emily Ferguson, Juhi Pandey, and Robert T Schultz. 2017. [Linguistic camouflage in girls with autism spectrum disorder](#). *Molecular autism*, 8(1):48.
- Andrew Pickles, Deborah K Anderson, and Catherine Lord. 2014. [Heterogeneity and plasticity in the development of language: a 17-year follow-up of children referred early for possible autism](#). *Journal of child psychology and psychiatry, and allied disciplines*, 55(12):1354–62.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Allison B Ratto, Lauren Turner-Brown, Betty M Rupp, Gary B Mesibov, and David L Penn. 2011. [Development of the Contextual Assessment of Social Skills \(CASS\): a role play measure of social skill for individuals with high-functioning autism](#). *Journal of autism and developmental disorders*, 41(9):1277–86.
- Darrel A Regier, William E Narrow, Diana E Clarke, Helena C Kraemer, S Janet Kuramoto, Emily A Kuhl, and David J Kupfer. 2013. [DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses](#). *The American journal of psychiatry*, 170(1):59–70.
- Masoud Rouhizadeh, Richard Sproat, and Jan van Santen. 2015. [Similarity Measures for Quantifying Restrictive and Repetitive Behavior in Conversations of Autistic Children](#). *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, 2015:117–123.

- Manar D. Samad, Norou Diawara, Jonna L. Bobzien, John W. Harrington, Megan A. Witherow, and Khan M. Iftekharuddin. 2018. [A Feasibility Study of Autism Behavioral Markers in Spontaneous Facial, Visual, and Hand Movement Response Data](#). *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(2):353–361.
- Jan P H van Santen, Richard W Sproat, and Alison Presmanes Hill. 2013. [Quantifying repetitive speech in autism spectrum disorders and language impairment](#). *Autism research : official journal of the International Society for Autism Research*, 6(5):372–83.
- Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro. 2015. [Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition](#). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1113–1133.
- Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro. 2017. [Learning Bases of Activity for Facial Expression Recognition](#). *IEEE Transactions on Image Processing*, 26(4):1965–1978.
- C. E. Shannon. 1951. [Prediction and Entropy of Printed English](#). *Bell System Technical Journal*, 30(1):50–64.
- H. Tanaka, S. Sakti, G. Neubig, T. Toda, and S. Nakamura. 2014. Linguistic and acoustic features for automatic identification of autism spectrum disorders in children’s narrative. In *Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 88–96.
- J. P. H. Van Santen, E. T. Prud’hommeaux, L. M. Black, and M. Mitchell. 2010. [Computational prosodic markers for autism](#). *Autism*, 14(3):215–236.
- D. Wechsler. 2011. *Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II)*. Pearson Clinical, San Antonio, TX.
- Ian H. Witten and Timothy C. Bell. 1990. [Source models for natural language text](#). *International Journal of Man-Machine Studies*, 32(5):545–579.
- N Yirmiya, C Kasari, M Sigman, and P Mundy. 1989. [Facial expressions of affect in autistic, mentally retarded and normal children](#). *Journal of child psychology and psychiatry, and allied disciplines*, 30(5):725–35.
- Sayaka Yoshimura, Wataru Sato, Shota Uono, and Motomi Toichi. 2015. [Impaired overt facial mimicry in response to dynamic facial expressions in high-functioning autism spectrum disorders](#). *Journal of autism and developmental disorders*, 45(5):1318–28.