The Journal of Child Psychology and Psychiatry

The Association for Child and Adolescent Mental Health

*Journal of Child Psychology and Psychiatry* **:* (2022), pp **–** doi:10.1111/jcpp.13595

# Attention to audiovisual speech does not facilitate language acquisition in infants with familial history of autism

**Katarzyna Chawarska,[1,2] David Lewkowicz,[1,2] Hannah Feiner,[1] Suzanne Macari,[1] and Angelina Vernetti[1]**

[1]Child Study Center, Yale University School of Medicine, New Haven, CT, USA; [2]Haskins Laboratories, New Haven, CT, USA

**Background:** Due to familial liability, siblings of children with ASD exhibit elevated risk for language delays. The processes contributing to language delays in this population remain unclear. **Methods:** Considering well-established links between attention to dynamic audiovisual cues inherent in a speaker's face and speech processing, we investigated if attention to a speaker's face and mouth differs in 12-month-old infants at high familial risk for ASD but without ASD diagnosis (hr-sib; $n$ = 91) and in infants at low familial risk (lr-sib; $n$ = 62) for ASD and whether attention at 12 months predicts language outcomes at 18 months. **Results:** At 12 months, hr-sib and lr-sib infants did not differ in attention to face ($p$ = .14), mouth preference ($p$ = .30), or in receptive and expressive language scores ($p$ = .36, $p$ = .33). At 18 months, the hr-sib infants had lower receptive ($p$ = .01) but not expressive ($p$ = .84) language scores than the lr-sib infants. In the lr-sib infants, greater attention to the face ($p$ = .022) and a mouth preference ($p$ = .025) contributed to better language outcomes at 18 months. In the hr-sib infants, neither attention to the face nor a mouth preference was associated with language outcomes at 18 months. **Conclusions:** Unlike low-risk infants, high-risk infants do not appear to benefit from audiovisual prosodic and speech cues in the service of language acquisition despite intact attention to these cues. We propose that impaired processing of audiovisual cues may constitute the link between genetic risk factors and poor language outcomes observed across the autism risk spectrum and may represent a promising endophenotype in autism. **Keywords:** Infancy; autism; audiovisual speech; eye-tracking; attention.

## Introduction

Autism spectrum disorder (ASD) is an early-onset complex neurodevelopmental condition defined by atypical patterns of social interaction, repetitive interests, and motor mannerisms (American Psychiatric Association, 2013). In addition to social deficits, children with ASD exhibit early language delays (Garrido, Petrova, Watson, Garcia-Retamero, & Carballo, 2017; Longard et al., 2017; Weismer, Lord, & Esler, 2010). Language delays are also observed in unaffected siblings of children with ASD. These delays typically emerge in the second year of life and manifest more strongly in the receptive language domain (Garrido et al., 2017; Marrus et al., 2018; Weismer et al., 2010). Considering that language delays cosegregate or are shared amongst affected and unaffected siblings and that they occur more frequently in unaffected siblings than in the general population, it has been proposed that language delays are linked with genetic risk factors associated with autism (Frazier et al., 2015). The processes linking the genetic risk factors with language outcomes in ASD or endophenotypes (intermediate phenotypes) (Gottesman & Gould, 2003) have not been identified. The present study investigated whether language acquisition in unaffected siblings of children with ASD (high-risk siblings, hr-sib) and in siblings of infants without ASD (low-risk siblings, lr-sib) benefits to the same extent from the facilitatory effects of audiovisual speech cues inherent in face and mouth movements accompanying speech. By comparing the hr-sib and lr-sib groups, we aimed to capture the effects of autism risk on links between attention to audiovisual prosodic and speech cues and language development without confounds related to certain core symptoms of autism (e.g., joint-attention impairments) which are known to have detrimental effects on language development (Kasari, Paparella, Freeman, & Jahromi, 2008; Mundy, Sigman, & Kasari, 1990; Weismer et al., 2010). We hypothesized that language delays, often reported in high-risk infants in the second year of life, will be associated with atypical attention to a speaker's face, an area recognized as a source of rich prosodic and speech cues known to facilitate speech perception. To examine this hypothesis, we investigated attention to audiovisual speech cues at 12 months and then assessed receptive and expressive language at 18 months. This developmental window constitutes an important transition between an age when infants have completed their tuning to the phonology of their native language to an age when they begin acquiring expressive and receptive vocabulary.

Conflict of interest statement: No conflicts declared.

The current study was motivated by extensive evidence linking attention to audiovisual prosodic and speech cues and language processing in typically developing children and adults. Speech processing in adults, in typically developing infants, and in young children benefits from the complementary and temporally synchronized auditory and visual speech cues originating from a speaker's mouth area (Sumby & Pollack, 1954; Summerfield, 1979, 1992; Yehia, Rubin, & Vatikiotis-Bateson, 1998). In adults, attention to such cues increases speech comprehension (Grant & Seitz, 2000; MacLeod & Summerfield, 1987; Shahin & Miller, 2009; Sumby & Pollack, 1954; Summerfield, 1979) and speeds up the neural processing of speech (van Wassenhove, Grant, & Poeppel, 2005). Attention to audiovisual cues, exemplified by preference for a speaker's mouth, begins to emerge in infancy after 6 months of age (Hillairet de Boisferon, Hansen-Tift, Minar, & Lewkowicz, 2017; Lewkowicz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowicz, 2015), and there is evidence that infants can integrate the audio and visual information contained in a speaker's mouth into a single multisensory representation (Kuhl & Meltzoff, 1984; Lewkowicz, Minar, Tift, & Brandon, 2015; Patterson & Werker, 1999; Rosenblum, Schmuckler, & Johnson, 1997). Attention to audiovisual speech facilitates learning of native phonetic forms (Lewkowicz & Hansen-Tift, 2012) and phoneme boundaries (Teinonen, Aslin, Alku, & Csibra, 2008), as well as separation between different languages in bilingual infants (Birulés, Bosch, Brieke, Pons, & Lewkowicz, 2018; Pons et al., 2015). Attention to audiovisual speech has also been linked with lexical acquisition both concurrently and prospectively (Habayeb et al., 2021; Imafuku, Kawai, Niwa, Shinya, & Myowa, 2019; Tenenbaum et al., 2015; Young, Merin, Rogers, & Ozonoff, 2009).

There is extensive evidence that 2-year-olds with ASD attend less to interactive partners, particularly to a speaker's face and mouth regions, and that precursors of these deficits can be observed in infancy, before behavioral symptoms of autism become apparent (Asberg Johnels, Gillberg, Falck-Ytter, & Miniscalco, 2014; Chawarska, Macari, & Shic, 2012; Habayeb et al., 2021; Macari et al., 2020; Righi et al., 2018; Shic, Macari, & Chawarska, 2014; Shic, Wang, Macari, & Chawarska, 2020). These findings suggest that by 6 months, the affected infants do not avail themselves to the same extent of visual prosodic and speech cues that are known to facilitate speech processing and language acquisition in unaffected infants. Although poor attention to faces has been typically examined for its links with symptom severity, several studies report that lower attention to a speaker's face and mouth is associated with lower language skills concurrently (Habayeb et al., 2021; Shic et al., 2020) and prospectively (Shic et al., 2020) in ASD, although the mutual influences of symptom severity and attention on language outcomes remains to be elucidated (Arunachalam & Luyster, 2016).

In contrast to infants with ASD, hr-sib infants exhibit largely typical patterns of selective attention to faces and facial features in infancy. Six-month-old hr-sib infants attend to dynamic, speaking faces comparably to lr-sib infants both in response to video stimuli (Chawarska, Macari, & Shic, 2013; Shic et al., 2014) and live interaction (Macari et al., 2020). The evidence regarding gaze behaviors in hr-sib infants past the first months of life, however, is scarce, and it is not clear if their typical selective social attention patterns persist later on. Moreover, no studies to date have examined predictive links between attention to a speaker's face and mouth and subsequent language development in hr-sib infants at the critical time when typically developing infants begin to rely on access to redundant audiovisual speech cues in a speaker's mouth to tune to their native phonological forms in the first year of life (Lewkowicz & Hansen-Tift, 2012) and to acquire their native lexicon in the second year of life (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2018). Considering that hr-sib infants have elevated likelihood for language delays in the second year of life, it is important to assess the link between these infants' attention to a speaker's face and mouth and their subsequent language skills.

The present study investigated whether hr-sib infants show diminished ability to attend to and take advantage of audiovisual prosodic and speech cues in the service of language acquisition. The hr-sib ($n$ = 91) and lr-sib ($n$ = 62) infants were assessed prospectively at 12 and 18 months. At 12 months, they completed the free-viewing Selective Social Attention eye-tracking task consisting of a video depicting a woman looking at the camera and using child-directed speech to engage the infant's attention (Chawarska et al., 2012, 2013). Subsequently, we computed a proportion of valid looking time spent monitoring the social scene (%Scene), speaker's face (%Face), and the proportion of time spent monitoring the speaker's mouth (Mouth Ratio) (see Figure 1, also Methods section). In contrast to studies of audiovisual speech processing in TD infants, which usually present a single face and no distractors, we exposed infants to a speaker within a complex visual scene that compelled them to select the most socially salient regions (e.g., face and mouth) amidst other competing stimuli, as they would in a real-life setting (Chawarska et al., 2012). At 12 and 18 months, we also examined receptive language (RL) and expressive language (EL) skills with the Mullen Scales of Early Learning (Mullen, 1995) and assessed the severity of social vulnerabilities with the Autism Diagnostic Observation Schedule-2 (Lord et al., 2012). First, we examined whether hr-sib infants differ from lr-sib infants at 12 months in their preferences for a speaker's face and mouth region. Second, we examined if attention to a speaker's face

and mouth at 12 months predicts RL and EL skills at 18 months alongside language scores at 12 months and sex. Given the paucity of attentional data for unaffected siblings of children with ASD, it was not clear whether 12-month-old hr-sib infants continue to be aligned with lr-sib infants as seen at 6 months (Chawarska et al., 2013; Shic et al., 2014), or whether they begin to exhibit atypical attention patterns resembling those seen in toddlers with ASD (Chawarska et al., 2012; Shic et al., 2020) and whether, if observed, these vulnerabilities contribute to their language outcomes 6 months later. We found that unlike low-risk infants and despite that they exhibited intact attention to the social partner, high-risk infants did not appear to benefit from audiovisual prosodic and speech cues in service of language acquisition. We propose that impaired processing of audiovisual stimuli may be the link between distal genetic risk factors and proximal poor language outcomes observed across the autism risk spectrum.

## Methods and materials

### *Participants*

All infants participated in a prospective longitudinal study of social development. The study was approved by the Human Investigation Committee of the Yale School of Medicine, and informed written consent was obtained from all parents prior to testing their infants. The infants were recruited prior to 6 months of age through resources of the Yale Developmental Disabilities Clinic and Yale Early Social Cognition Lab, as well as through advertisement. Out of 168 participants who attended the visit at 12 months, 102 infants were younger siblings of children with ASD and thus at high familial risk (hr-sib) for ASD, while 66 infants had no history of ASD in 1st or 2nd degree relatives and were considered at low familial risk (lr-sib) for ASD. Families reported primary language being English. Exclusionary criteria were gestational age below 34 weeks, any hearing or visual impairment, nonfebrile seizure disorders, or known genetic syndromes. Presence of language or other developmental delays did not constitute an exclusion criterion in either group. Infants diagnosed with ASD were excluded from the present study.

Out of 168 infants seen at 12 months, 15 (9%) infants [4 (6%) lr-sib and 11 (11%) hr-sib] missed the 18-month visit and were excluded from the analysis. The two groups did not differ in the proportion of children who skipped the 18-month visit, $\chi^2(1) = 1.10$, $p = .294$. Those who missed the 18-month visit did differ from the retained sample in sample characteristics and in performance on the eye-tracking task (see Table S1). The final sample consisted of 91 of the hr-sib infants and 62 of the lr-sib infants ($N = 153$). Females constituted 41.83% (64 out of 152) of the sample, and the groups did not differ on the sex distribution (hr-sib: 35 out of 91, lr-sib: 29 out of 62) ($\chi^2(1) = 1.047$, $p = .306$). One family did not provide race information; 91% (82/90) of parents in the hr-sib group identified their child's race as Caucasian as compared to 81% in the lr-sib group (50/62), $\chi^2(1) = 3.519$, $p = .061$. Blacks represented 5.3% of the sample, Asians 4.58%, and 3.27% were more than once race. Hispanics represented 9.68% of the hr-sib group as compared to 9.68% of the lr-sib group, $\chi^2(1) = 1.057$, $p = .304$. 83.17% of the hr-sib infants had mothers who completed college education as compared to 80.36% in the lr-sib group, $\chi^2(1) = 0.194$, $p = .659$. All infants underwent a comprehensive diagnostic assessment capturing their developmental and medical history, verbal and nonverbal skills (Mullen Scales of Early Learning; MSEL) (Mullen, 1995), adaptive skills (Vineland Adaptive Behavior Scales; VABS) (Sparrow, Balla, & Cicchetti, 1984), and severity of autism symptoms (Autism Diagnostic Observation Schedule-2 Toddler Module) (Luyster et al., 2009). The clinical best estimate (CBE) diagnosis was assigned by a team of expert clinicians based on a review of all available records. One hundred fourteen (75%) children received their final diagnostic assessment at 36 months; the remaining children received it at 24 months ($n = 33$, 22%) or at 18 months ($n = 6$, 4%). Consistent with prior reports (Charman et al., 2017), toddlers in the hr-sib group were more likely to trigger clinical concerns either due to the presence of developmental delays or subthreshold social difficulties than toddlers in the lr-sib group (hr-sib: 40% (36/91) versus lr-sib: 8% (5/62), $\chi^2(1) = 18.65$, $p < .001$).

### *Stimuli*

The stimulus video depicted a woman positioned at the center of the screen, with four distractor toys presented in the four corners of the visual scene (Chawarska et al., 2012). The video contained 11 episodes (total duration 69s) during which the woman intended to engage the viewer by looking at the camera, smiling, slight nods and eyebrow movements, and using child-directed speech while addressing the viewer (Figure 1, left). The speech episodes were interspersed with episodes absent of speech, during which the woman made a sandwich, or looked at moving or stationary toys. Please see Chawarska et al., 2012 for detailed description. There were no artificial breaks in the video to re-engage or re-center the viewer's attention, thus requiring the infants to adjust their gaze patterns depending on context as they would in real life. The scene subtended 27 × 21 degrees of visual angle, the Face 3.9 × 5.6 degrees, the

**Figure 1** Screenshot from the SSA task. The stimulus video depicted a woman positioned at the center of the screen, with distractors presented in the four corners. The video contained 11 episodes (total duration 69 s) during which the woman looked at the camera and spoke using child-directed speech. The proportion of the looking time (%Scene) was standardized by the total duration of the speech episodes and signify the overall attention to the task; %Face was standardized by the total looking time at the scene; and the Mouth Ratio represents a proportion of looking time to the mouth over total looking time at the face region consisting of the upper (eyes) and lower (mouth) ROIs

Mouth 3.5 × 2.0 degrees, and each of the Toys 5.8 × 6.4 degrees.

### Apparatus

Gaze behaviors were recorded at a sampling rate of 60Hz using a SensoMotoric Instruments IView X™ RED eye-tracking system. Eye-tracking data were processed using custom software written in MATLAB. The software accommodated standard techniques for processing eye-tracking data, including blink detection, data calibration, recalibration, and region of interest (ROI) analysis (Duchowski, 2003; Shic, 2008).

### Procedure

During the free-viewing task, toddlers were seated in a car seat in a dark and soundproof room 75 cm in front of a 24" widescreen LCD monitor. Each session began with a cartoon video to help the infant get settled. A five-point calibration procedure was then initiated with calibration points consisting of dynamic targets (e.g., a meowing, walking cartoon tiger). Subsequently, each participant was presented with the video described in the Stimulus section.

### Data reduction

The visual scene was divided into ROIs (see Figure 1, right). Variables of interest were proportion of total looking (dwell) time at the entire scene (%Scene), the proportion of looking time at the woman's face consisting of the eye and mouth regions (% Face), and a proportion of looking at the mouth (Mouth Ratio). The proportion of the looking time (%Scene) was standardized by the total duration of the speech episodes and signify the overall attention to the task; %Face was standardized by the total looking time at the scene during the speech episodes and stands for the ability to attend selectively to the face of the speaker over other elements of the scene; and the Mouth Ratio represents a proportion of looking time to the mouth over total looking time at the face region consisting of the upper (eyes) and lower (mouth) ROIs. Calibration error was on average $M = .67$ ($SD = .39$) degrees in the hr-sib and $M = .67$ ($SD = .40$) degrees in the lr-sib groups ($p = .922$). Sessions in which infants contributed less than 20% of valid eye-tracking data were excluded from the analysis of %Face and Mouth Ratio ($n = 2$).

### Statistical analysis

Preliminary analysis of RL and EL scores was conducted using one-way analysis of variance (ANOVA). Associations between continuous variables were examined using Pearson's $r$ correlation coefficient analysis, and associations between a binary variable (sex) and continuous variables were examined using Point Biserial correlation coefficient analysis. Multivariate regression analysis was used to examine 12-month predictors of the 18-month RL and EL skills in the lr-sib and hr-sib samples. Predictors included sex, the eye-tracking variables (% Face and Mouth Ratio), and the RL and EL scores at 12 months. Diagnostics to evaluate for potential collinearity amongst predictor variables were performed using COLLIN option in the SAS REG procedure and none were identified. The analyses were conducted using SAS 9.4.

## Results

### Preliminary analyses

At 12 months, there were no statistically significant differences between the hr-sib and lr-sib groups in their MSEL RL and EL t-scores ($d = 0.15$ and $d = 0.16$; see Table 1). In contrast, by 18 months, the hr-sib group had lower RL scores than the lr-sib group ($p = .01$, $d = 0.42$) even though the EL scores for these two groups were comparable ($d = 0.03$). When only the proportion of children whose RL t-scores fell more than 1.5 *SD* below the mean were considered (i.e., had *t*-scores below 35), the 18-month-old hr-sib group was twice more likely to fall into this category than the 18-month-old lr-sib group (hr-sib: 34% (31 out of 91) versus lr-sib: 18% (11 out of 62), $\chi^2(1) = 4.93$, $p = .026$). The analogous analysis of the EL scores revealed no significant differences between the groups (hr-sib: 24% (22 out of 91) versus lr-sib: 18% (11 out of 62), $\chi^2(1) = 0.902$, $p = .342$). In contrast, at 12 months, there were no differences between the groups in the proportion of children with t-scores below 35 in the RL or EL domains. In the RL domain, 12% of HR and 10% of LR infants had t-scores under 35, chisq (1) = .217, $p = .641$ and in the EL domain in the HR group 19% infants and in the LR group, 27% of infants had t-scores under 35, chisq (1) = 1.62, $p = .202$. Nonverbal t-score (averaged across the Fine Motor and Visual Reception domains) was comparable across the two groups at 12 and 18 months ($d = 0.08$, $d = 0.04$). At both 12 and 18 months, the hr-sib and lr-sib groups had comparable ADOS Total calibrated comparison scores ($d = 0.16$ and $d = 0.19$).

### Selective attention scores

At 12 months, there were no statistically significant differences between the groups in the %Scene, % Face, and Mouth Ratio variables (all $p$-values > .14; see Table 1, Figure 2). The Mouth Ratio was significantly greater than chance (.50) in both groups (hr-sib: 66% ($SD = 22$), $t(89) = 6.91$, $p < .001$, lr-sib: 62% ($SD = 22$), $t(60) = 4.16$, $p < .001$), suggesting that, regardless of risk status, infants favored the mouth over the eye region when they attended to the speaker's face. Moreover, there was a significant correlation between overall attention to the scene and the speaker's face in both groups (hr-sib: $r(90) = .370$, $p < .001$, lr-sib: $r(61) = .390$, $p = .002$) suggesting that infants who looked more at the screen tended to also focus more on the speaker's face rather than on the other elements of the scene and this pattern was consistent in both risk groups.

### 12-month predictors of 18-month Receptive Language scores

We performed a series of multiple linear regression analyses to examine the association between 12-month selective attention and 18-month RL scores. Results from the lr-sib group indicated that there was a collective significant effect of the predictor eye-tracking indices, language scores at 12 months, and sex on the *18-month RL scores*, $F(5,55) = 7.11$, $p <$

**Table 1** Sample characteristics at 12 and 18 months

| | hr-sib *M* (*SD*) | lr-sib *M* (*SD*) | *p*-value | Cohen's *d* |
|---|---|---|---|---|
| *N* | 91 | 62 | | |
| Sex (% female) | 38.46 | 46.77 | .31 | |
| 12 mo | | | | |
| % Scene | 79.55 (19.20) | 76.43 (18.77) | .320 | 0.16 |
| % Face | 67.08 (14.97) | 63.08 (17.01) | .130 | 0.25 |
| Mouth Ratio | 0.66 (0.22) | 0.62 (0.22) | .295 | 0.18 |
| MSEL EL *t*-score | 43.02 (10.04) | 44.92 (12.93) | .309 | 0.16 |
| MSEL RL *t*-score | 44.30 (10.50) | 45.76 (9.20) | .376 | 0.15 |
| MSEL Nonverbal DQ | 115.32 (14.13) | 116.47 (11.77) | .259 | 0.08 |
| ADOS-2 Total Severity | 3.40 (1.74) | 3.15 (1.37) | .368 | 0.16 |
| 18 mo | | | | |
| MSEL EL *t*-score | 46.76 (13.73) | 47.19 (12.24) | .841 | 0.03 |
| MSEL RL *t*-score | 45.95 (16.48) | 52.87 (16.76) | .012 | 0.42 |
| MSEL Nonverbal DQ | 104.22 (11.32) | 107.09 (14.27) | .168 | 0.04 |
| ADOS-2 Total Severity | 2.71 (1.40) | 2.42 (1.61) | .232 | 0.19 |

MSEL, Mullen Scales of Early Learning; EL, Expressive Language; RL, Receptive Language; DQ, Developmental Quotient; ADOS-2 Autism Diagnostic Observation Scale–2.

.001, $R^2$ = .39 (see Table 2). Individual predictor analysis indicated that Mouth Ratio ($t$ = 2.06, $p$ = .044, $\beta$ = .22), %Face ($t$ = 2.10, $p$ = .040, $\beta$ = .22), and 12-month RL $t$-score ($t$ = 3.74, $p$ < .001, $\beta$ = .46) contributed significantly to the model. The analogous analysis in the hr-sib sample indicated that the full model was also significant, $F(5,84) = 10.16$, $p$ < .001, $R^2$ = .38, and that the significant predictors included the 12-month RL scores ($t$ = .4.41, $p$ < .001, $\beta$ = .42) and sex ($t$ = 3.15, $p$ < .002, $\beta$ = .29), but not the attentional indices (see Table 3). Thus, in the lr-sib infants, greater attention to the speaker's face and a higher proportion of time spent monitoring the speaker's mouth contributed significantly to better language comprehension scores at 18 months, above and beyond the contribution of language levels at 12 months and sex. In contrast, for hr-sib infants, neither attention to the face nor the mouth contributed significantly to the model; only higher RL skills at 12 months and female sex were predictive of better verbal comprehension at 18 months in the high-risk sample.

### *12-month predictors of 18-month Expressive Language scores*

Results of the multiple linear regression in the lr-sib infants yielded a significant effect of the 12-month predictors on the 18-month EL scores, $F(5,55)$ = 5.93, $p$ < .001, $R^2$ = .35 (see Table 4). Two variables contributed significantly to the model: 12-month % Face ($t$ = 2.99, $p$ = .004, $\beta$ = .33) and 12-month EL t-scores ($t$ = 2.54, $p$ = .014, $\beta$ = .33). The analogous linear regression analysis in the hr-sib infants yielded a similarly collective significant effect of the 12-month predictors, $F(5,84) = 9.35$, $p$ < .001, $R^2$ = .36. In contrast, however, only 12-month RL $t$-scores ($t$ = 4.12, $p$ < .001, $\beta$ = .40) and EL t-scores ($t$ = 3.37, $p$ = .001, $\beta$ = .32) contributed significantly to the model (see Table 5). Thus, lr-sib infants who spent more time monitoring the speaker's face, and those with stronger 12-month EL scores exhibited higher 18-month EL t-scores. In contrast, in the hr-sib infants, none of the 12-month eye-tracking measures contributed significantly to 18-month EL scores. Instead, higher 12-month RL and EL t-scores were individually predictive of stronger EL in hr-sib infants 6 months later.

Considering that hr-sib infants often exhibit subtle vulnerabilities in the areas of social interaction and communication, often referred to as broader autism phenotype (Macari et al., 2012; Rowberry et al., 2015), we conducted an additional regression analysis to evaluate if severity of autism features, as measured at 12 months by the ADOS-2, would help explain language delays observed at 18 months. Results indicated that autism symptom severity did not contribute significantly to the models predicting EL and RL of the hr-sib group and that inclusion of symptom severity did not change the overall pattern of results (see Table S2). Thus, the observed receptive language delay in hr-sib infants was not due to elevated autism symptoms.

## Discussion

The present study examined the contribution of selective attention to dynamic speaking faces at 12 months to language outcomes at 18 months in infants with and without genetic risk for ASD. We demonstrate, for the first time, that in low-risk infants, individual differences in attention to a speaker's face at 12 months contribute to receptive and expressive language outcomes 6 months later. The finding suggests that kinetic cues including head nods and facial gestures that typically accompany child-directed speech provide important visual prosodic cues that complement tonal and temporal

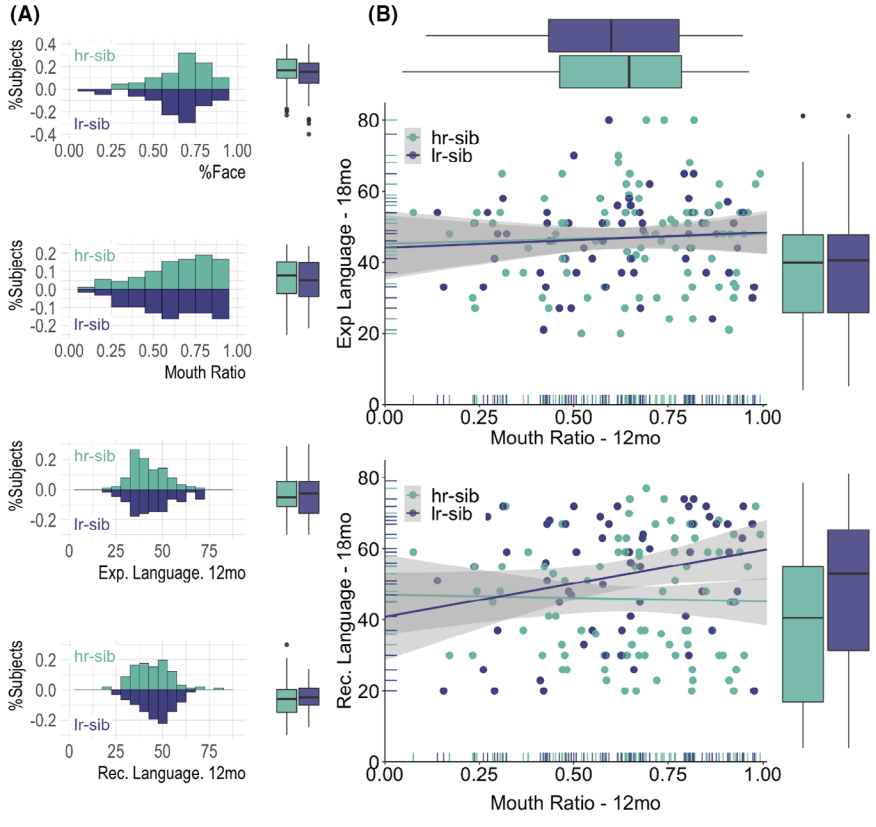


**Figure 2** (A) Relative frequency distributions of the proportion of looking time at the speaker's face (%Face), and mouth preference (Mouth Ratio), as well as Mullen Scales of Early Learning Expressive and Receptive Language *t*-scores at 12 months of age of hr-sib (green) and lr-sib (blue) groups. (B) Correlational plots with regression lines of Mouth Ratio at 12 months with Expressive Language (top) and Receptive Language (bottom) at 18 months in hr-sib (green) and lr-sib (blue) groups. Gray area represents 95% confidence interval. Boxplots in panels (A) and (B) represent the median, 1st and 3rd quartiles of the eye-tracking, and language measures for both hr-sib and lr-sib groups

**Table 2** Multiple regression of Mouth Ratio (MR), proportion of looking at the face (%Face), MSEL Receptive and Expressive Language (RL, EL) *t*-scores, and sex variables at 12 months on MSEL Receptive Language *t*-scores at 18 months in low-risk infants

| Predictors | 18 m RL (DV) | 12 m MR | 12 m %Face | 12 m RL | 12 m EL | *B* | *SE B* | *β* | $r^2$ (semi-partial) |
|---|---|---|---|---|---|---|---|---|---|
| MR | .25 | | | | | 16.17 | 7.84* | .22 | .047 |
| %Face | .24 | .04 | | | | 21.52 | 10.25* | .22 | .049 |
| 12 m RL | .50 | .06 | .01 | | | 0.82 | 0.22*** | .46 | .154 |
| 12 m EL | .38 | .01 | .05 | .47 | | 0.15 | 0.16 | .12 | .010 |
| Female | .07 | −.02 | .01 | −.06 | .20 | 1.49 | 3.36 | .05 | .002 |
| Intercept | | | | | | −16.11 | 11.88 | 0 | – |
| Mean | 52.87 | .62 | 63.08 | 45.76 | 44.91 | | | $R^2$ = .39 | |
| *SD* | 16.76 | .22 | 17.01 | 9.20 | 12.93 | | | Adjusted $R^2$ = .34 | |

MR, Mouth Ratio; EL, Expressive Language; RL, Receptive Language.
*$p$ < .05, **$p$ < .01, ***$p$ < .001.

prosodic cues in auditory speech (Esteve-Gibert & Guellaï, 2018). Our findings complement extant, albeit still limited, evidence from older children suggesting that presence of kinetic speech cues facilitates attention to a speaker, helps parse concatenated speech (Kitamura, Guellaï, & Kim, 2014), and facilitates word learning and recall (Booth, McGregor, & Rohlfing, 2008; Igualada, Esteve-Gibert, & Prieto, 2017). Moreover, consistent with prior work in premature and in typically developing low-risk 12-month-olds (Imafuku et al., 2019; Tenenbaum et al., 2015), our study suggests that preferential attention to the speaker's mouth at 12 months predicts better language skills at 18 months. Taken together, the findings in the low-risk group highlight the beneficial effects of attention to the audiovisual cues inherent in faces of interactive partners for early language development. While early attention to speaker's face contributes to better expressive *and* receptive language skills at 18 months, greater focus on the mouth is associated with better receptive language skills.

**Table 3** Multiple regression of Mouth Ratio (MR), proportion of looking at the face (%Face), MSEL Receptive and Expressive Language (RL, EL) *t*-scores, and sex variables at 12 months on MSEL Receptive Language *t*-scores at 18 months in high-risk infants

| Predictors | 18 m RL (DV) | 12 m MR | 12 m %Face | 12 m RL | 12 m EL | *B* | *SE B* | *β* | $r^2$ (semi-partial) |
|---|---|---|---|---|---|---|---|---|---|
| MR | −.02 | | | | | −4.35 | 6.74 | −.06 | .003 |
| %Face | .08 | .08 | | | | 3.38 | 9.61 | .03 | .001 |
| 12 m RL | .52 | .02 | .05 | | | 0.66 | 0.16*** | .42 | .144 |
| 12 m EL | .29 | −.10 | .09 | .34 | | 0.13 | 0.15 | .08 | .005 |
| Female | .43 | .10 | .07 | .31 | .21 | 9.78 | 3.10** | .29 | .074 |
| Intercept | | | | | | 7.67 | 10.59 | 0 | – |
| Mean | 45.95 | .66 | 67.08 | 44.30 | 43.02 | | | $R^2$ = .38 Adjusted $R^2$ = .34 | |
| *SD* | 16.47 | .22 | 14.97 | 10.50 | 10.04 | | | | |

MR, Mouth Ratio; EL, Expressive Language; RL, Receptive Language.
*$p$ < .05, **$p$ < .01, ***$p$ < .001.

**Table 4** Multiple regression of Mouth Ratio (MR), proportion of looking at the face (%Face), MSEL Receptive and Expressive Language (RL, EL) *t*-scores, and sex variables at 12 months on MSEL Expressive Language *t*-scores at 18 months in low-risk infants

| Predictors | 18 m EL (DV) | 12 m MR | 12 m %Face | 12 m RL | 12 m EL | *B* | *SE B* | *β* | $r^2$ (semi-partial) |
|---|---|---|---|---|---|---|---|---|---|
| MR | .08 | | | | | 2.91 | 5.76 | .06 | .003 |
| %Face | .35 | .04 | | | | 22.50 | 7.53** | .33 | .105 |
| 12 m RL | .33 | .06 | .01 | | | 0.27 | 0.16 | .21 | .033 |
| 12 m EL | .50 | .01 | .05 | .47 | | 0.30 | 0.12* | .33 | .076 |
| Female | .14 | −.02 | .01 | −.06 | .20 | 1.33 | 2.61 | .06 | .003 |
| Intercept | | | | | | 4.50 | 8.73 | 0 | – |
| Mean | 47.19 | .62 | 63.08 | 45.76 | 44.91 | | | $R^2$ = .35 Adjusted $R^2$ = .29 | |
| *SD* | 12.24 | .22 | 17.01 | 9.20 | 12.93 | | | | |

MR, Mouth Ratio; EL, Expressive Language; RL, Receptive Language.
*$p$ < .05, **$p$ < .01, ***$p$ < .001.

**Table 5** Multiple regression of Mouth Ratio (MR), proportion of looking at the face (%Face), MSEL Receptive and Expressive Language (RL, EL) *t*-scores, and sex variables at 12 months on MSEL Expressive Language *t*-scores at 18 months in high-risk infants

| Predictors | 18 m RL (DV) | 12 m MR | 12 m %Face | 12 m RL | 12 m EL | *B* | *SE B* | *β* | $r^2$ (semi-partial) |
|---|---|---|---|---|---|---|---|---|---|
| MR | .04 | | | | | 4.74 | 5.65 | .075 | .005 |
| %Face | −.06 | .08 | | | | −10.35 | 8.05 | −.114 | .13 |
| 12m RL | .52 | .02 | .05 | | | 0.52 | 0.13*** | .398 | .130 |
| 12m EL | .44 | −.10 | .09 | .34 | | 0.43 | 0.13** | .320 | .087 |
| Female | .22 | .10 | .07 | .31 | .21 | 0.46 | 2.60 | .017 | 0 |
| Intercept | | | | | | 8.94 | 8.88 | 0 | – |
| Mean | 46.76 | .66 | 67.08 | 44.30 | 43.02 | | | $R^2$ = .36 Adjusted $R^2$ = .32 | |
| *SD* | 13.73 | .22 | 14.97 | 10.50 | 10.04 | | | | |

MR, Mouth Ratio; EL, Expressive Language; RL, Receptive Language.
*$p$ < .05, **$p$ < .01, ***$p$ < .001.

Although at 12 months both groups had comparable expressive and receptive language scores, by 18 months the high-risk group had lower receptive language scores than the control group and this delay was more pronounced in male high-risk siblings. This finding suggests that previously reported language delays in high-risk toddlers (Marrus et al., 2018) emerge as infants transition from tuning to their native language in the first year of life to rapid acquisition of the lexicon and grammar in the second year of life and that male siblings are more vulnerable during this transition than female siblings. The specific vulnerability observed here in the language comprehension versus expression domain is similar to that observed in toddlers with ASD (Marrus et al., 2018), suggesting consistency in the language

development pattern across the autism risk spectrum. While language levels at 12 months were significant predictors of later language outcomes, the autism symptom severity scores were not. This suggests that the observed language delays are not related to subthreshold social vulnerabilities often observed in unaffected siblings. Most importantly, although the 12-month-old high-risk infants deployed a comparable proportion of attention to the speaker's face and mouth regions as did the low-risk infants, individual differences in their gaze behaviors at 12 months did not predict later language outcomes. This suggests that intact social attention in high-risk infants does not guarantee that the infants extract, process, and utilize the audiovisual cues to the same extent as the low-risk infants.

We propose that the absence of a relationship between attention to the speaker's face and mouth regions at 12 months and language at 18 months in high-risk infants may be due to a disruption in the processing and integration of dynamic facial and speech cues and that, unlike impaired social attention, which appears to be specific to ASD, this deficit may be shared amongst children across the autism risk spectrum. This idea is consistent with findings suggesting that 9-month-old high-risk infants exhibit reduced audiovisual speech integration ability (Guiraud et al., 2012) and that preschoolers (Newman, Kirby, Von Holzen, & Redcay, 2021) and school-aged children (Smith & Bennetto, 2007) with ASD benefit less from lip movements during speech decoding, with the effect driven largely by lower audiovisual integration skills (Smith & Bennetto, 2007). Findings also show that preschoolers with ASD do not prefer synchronous over asynchronous audiovisual events, suggesting altered multisensory speech processing and that diminished preference for synchronous speech was linked with lower language skills (Righi et al., 2018). Interestingly, manipulations that enhance attention to both a speaker's mouth and a referenced object enhance word comprehension in verbal preschoolers with autism (Tenenbaum et al., 2015). The audiovisual integration challenges are more pronounced at younger ages in ASD, and the differences between affected and unaffected samples are most pronounced when integration skills are tested with linguistic and social (e.g., speech and/or faces) rather than nonsocial and nonlinguistic stimuli (Feldman et al., 2018). Notably, the perception of speech and dynamic facial gestures and expressions as well as the integration of visual and auditory speech cues are subserved by the superior temporal sulcus (STS) (Allison, Puce, & McCarthy, 2000; Rennig & Beauchamp, 2018; Stevenson & James, 2009) and atypical activation of STS has been found in individuals with ASD and in their unaffected siblings (Ahmed & Vander Wyk, 2013; Alaerts et al., 2014; Redcay, 2008; Spencer et al., 2011; von dem Hagen et al., 2011). Thus, even though unaffected siblings of children with ASD exhibit intact attention to social partners whereas the affected siblings typically do not, the two groups may share disrupted processing and integration of audiovisual prosodic and speech cues and this may have negative cascading effects on their acquisition of early language skills.

If our interpretation is correct, then impaired processing and integration of audiovisual prosodic and speech cues may represent a promising endophenotype candidate (Gottesman & Gould, 2003). Mechanistically, impaired audiovisual prosody and speech processing may lay between the distal genetic risk factors for autism and proximal overt expressions of these factors in behavior (i.e., atypical language development). Currently, empirical evidence on the processing of audiovisual cues along with their neural correlates across the autism risk spectrum is extremely scant despite its importance for understanding mechanisms driving language delays so common and disabling in this population. Importantly, processing of audiovisual speech cues can be measured reliably in children with a wide range of developmental skills enabling researchers to examine if deficits in audiovisual speech processing cosegregates and aggregates in the affected and unaffected family members, and to what extent it is functionally associated with language outcomes. The present findings motivate a full-scale investigation into audiovisual speech processing in infants and toddlers with ASD and their unaffected siblings with the goal of mapping their developmental dynamics and establishing their role in the development of core and co-occurring features. Future studies should also address whether improving audiovisual speech processing and integration in infancy can ameliorate subsequent language delays and thus, provide evidence for causal links between audiovisual processing and language outcomes in ASD.

Finally, our results have important clinical implications. They suggest that in some contexts involving social interaction and verbal communication and during early development when language acquisition is occurring, close monitoring of a speaker's mouth is highly adaptive. Thus, any intervention aimed at altering social attentional patterns in ASD or in infants at risk for ASD need to consider both the child's developmental level and the context in which the child is to make attentional choices in complex everyday environments. Furthermore, it suggests that altering the attentional patterns alone (i.e., making children look less or more at certain social stimuli) may not be effective unless the ability to integrate the multisensory attributes of such stimuli is also targeted for intervention.

### *Limitations and future directions*

The free-viewing study design reliably captured how long the participants dwelled on specific ROIs (e.g.,

face, mouth). Nonetheless, our prior work (Chawarska & Shic, 2009; Wang, Chang, & Chawarska, 2020) has shown that intact attention does not always guarantee that the key information contained in the displays is learned and remembered if the mechanisms responsible for the processing of such information are not functioning properly. Thus, to fully understand factors responsible for language delays in infants with ASD and in the unaffected high-risk siblings, future studies should examine processing of audiovisual prosodic and speech cues across the autism risk spectrum at the behavioral and neurophysiological levels.

## Conclusions

Receptive language delays in unaffected siblings of children with ASD emerge between 12 and 18 months and are potentially linked with impaired ability to exploit audiovisual prosodic and speech cues in the service of language acquisition. The current findings suggest a novel endophenotype linked with language outcomes in infants carrying familial risk for autism and a possible avenue for identifying novel developmentally informed treatment targets.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

**Table S1**. A comparison of the sample retained for the analysis ($n = 153$) and the sample dropped ($n = 15$) due to the missing follow-up data at 18 months.

**Table S2**. Multiple regression of Mouth Ratio (MR), proportion of looking at the face (%Face), Receptive and Expressive Language (RL, EL) t-scores, and sex variables at 12 months on Receptive Language *t*-scores at 18 months in high-risk sibling group.

## Acknowledgments

The study was supported by the National Institute of Mental Health R01 MH087554 and P50 MH115716 grants awarded to K.C. D.J.L. was supported by Grant BCS 1749507 from the National Science Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank the children and their families for participating in the study. The authors acknowledge Dr. Richard Aslin for his comments on an earlier draft of the manuscript and the clinical team of the Yale Social and Affective Neuroscience of Autism Program for their contribution to sample characterization and data collection. The authors have declared that they have no competing or potential conflicts of interest.

## Correspondence

Katarzyna Chawarska, Yale Child Study Center, 300 George Street, Suite 900, New Haven CT 06511; Email: Katarzyna.chawarska@yale.edu

### Key points

- It has been proposed that in ASD, language delays which are present in both affected and unaffected family members and which occur more frequently than in the general population are linked to shared genetic risk factors.
- Here we demonstrate that unlike low-risk infants, high-risk infants fail to benefit from audiovisual prosodic and speech cues in the service of language acquisition despite intact attention to these cues.
- We propose that impaired processing of such cues may constitute the link between genetic risk factors and poor language outcomes observed across the autism risk spectrum and may represent a promising endophenotype in autism.

## References

Ahmed, A.A., & Vander Wyk, B.C. (2013). Neural processing of intentional biological motion in unaffected siblings of children with autism spectrum disorder: An fMRI study. *Brain and Cognition*, *83*(3), 297–306.

Alaerts, K., Woolley, D.G., Steyaert, J., Di Martino, A., Swinnen, S.P., & Wenderoth, N. (2014). Underconnectivity of the superior temporal sulcus predicts emotion recognition deficits in autism. *Social Cognitive and Affective Neuroscience*, *9*(10), 1589–1600.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*(7), 267–278.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM 5*, 5th ed. Washington DC: American Psychiatric Association.

Arunachalam, S., & Luyster, R.J. (2016). The integrity of lexical acquisition mechanisms in autism spectrum disorders: A research review. *Autism Research*, *9*(8), 810–828.

Asberg Johnels, J., Gillberg, C., Falck-Ytter, T., & Miniscalco, C. (2014). Face-viewing patterns in young children with autism spectrum disorders: Speaking up for the role of language comprehension. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2246–2252. https://doi.org/10.1044/2014_JSLHR-L-13-0268

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D.J. (2018). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental Science*, *22*(3), e12755. https://doi.org/10.1111/desc.12755.

Booth, A.E., McGregor, K.K., & Rohlfing, K.J. (2008). Socio-pragmatics and attention: Contributions to gesturally

guided word learning in toddlers. *Language Learning and Development*, *4*(3), 179–202.
Charman, T., Young, G.S., Brian, J., Carter, A., Carver, L.J., Chawarska, K., … Zwaigenbaum, L. (2017). Non-ASD outcomes at 36 months in siblings at familial risk for autism spectrum disorder (ASD): A baby siblings research consortium (BSRC) study. *Autism Research*, *10*(1), 169–178. https://doi.org/10.1002/aur.1669.
Chawarska, K., Macari, S., & Shic, F. (2012). Context modulates attention to social scenes in toddlers with autism. *Journal of Child Psychology and Psychiatry*, *53*(8), 903–913.
Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, *74*(3), 195–203.
Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*(12), 1663.
Duchowski, A.T. (2003). *Eye tracking methodology: Theory and practice*. New York: Springer.
Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the auditory and visual domains: A developmental perspective. *Frontiers in Psychology*, *9*, 338.
Feldman, J.I., Dunham, K., Cassidy, M., Wallace, M.T., Liu, Y., & Woynaroski, T.G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *95*, 220–234.
Frazier, T.W., Youngstrom, E.A., Hardan, A.Y., Georgiades, S., Constantino, J.N., & Eng, C. (2015). Quantitative autism symptom patterns recapitulate differential mechanisms of genetic transmission in single and multiple incidence families. *Molecular Autism*, *6*(1), 1–12.
Garrido, D., Petrova, D., Watson, L.R., Garcia-Retamero, R., & Carballo, G. (2017). Language and motor skills in siblings of children with autism spectrum disorder: A meta-analytic review. *Autism Research*, *10*(11), 1737–1750. https://doi.org/10.1002/aur.1829.
Gottesman, I.I., & Gould, T.D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry*, *160*(4), 636–645.
Grant, K.W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3), 1197–1208.
Guiraud, J.A., Tomalski, P., Kushnerenko, E., Ribeiro, H., Davies, K., Charman, T., … Johnson, M.H. (2012). Atypical audiovisual speech integration in infants at risk for autism. *PLoS One*, *7*(5), e36428. https://doi.org/10.1371/journal.pone.0036428
Habayeb, S., Tsang, T., Saulnier, C., Klaiman, C., Jones, W., Klin, A., & Edwards, L.A. (2021). Visual traces of language acquisition in toddlers with autism spectrum disorder during the second year of life. *Journal of Autism and Developmental Disorders*, *51*(7), 2519–2530.
Hillairet de Boisferon, A., Hansen-Tift, A., Minar, N.J., & Lewkowicz, D.J. (2017). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*(3), https://doi.org/10.1111/desc.12381
Hillairet de Boisferon, A., Tift, A.H., Minar, N.J., & Lewkowicz, D.J. (2018). The redeployment of attention to the mouth of a talking face during the second year of life. *Journal of Experimental Child Psychology*, *172*, 189–200. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022096517303028
Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3-to 5-year-old children. *Journal of Experimental Child Psychology*, *156*, 99–112.
Imafuku, M., Kawai, M., Niwa, F., Shinya, Y., & Myowa, M. (2019). Audiovisual speech perception and language acquisition in preterm infants: A longitudinal study. *Early Human Development*, *128*, 93–100.
Kasari, C., Paparella, T., Freeman, S., & Jahromi, L.B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology*, *76*(1), 125.
Kitamura, C., Guellaï, B., & Kim, J. (2014). Motherese by eye and ear: Infants perceive visual prosody in point-line displays of talking heads. *PLoS One*, *9*(10), e111467.
Kuhl, P.K., & Meltzoff, A.N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development*, *7*(3), 361–381.
Lewkowicz, D.J., & Hansen-Tift, A.M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436.
Lewkowicz, D.J., Minar, N.J., Tift, A.H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, *130*, 147–162. https://doi.org/10.1016/j.jecp.2014.10.006
Longard, J., Brian, J., Zwaigenbaum, L., Duku, E., Moore, C., Smith, I.M., … Bryson, S. (2017). Early expressive and receptive language trajectories in high-risk infant siblings of children with autism spectrum disorder. *Autism & Developmental Language Impairments*, *2*, 239694151773741.
Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule*, 2nd edition (ADOS-2). Torrance, CA: Western Psychological Services.
Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., … Lord, C. (2009). The Autism Diagnostic Observation Schedule-Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(9), 1305–1320. https://doi.org/10.1007/s10803-009-0746-z
Macari, S., Campbell, D., Gengoux, G.W., Saulnier, C.A., Klin, A.J., & Chawarska, K. (2012). Predicting developmental status from 12 to 24 months in infants at risk for autism spectrum disorder: A preliminary report. *Journal of Autism and Developmental Disorders*, *42*(12), 2636–2647.
Macari, S., Milgramm, A., Reed, J., Shic, F., Powell, K.K., Macris, D., & Chawarska, K. (2020). Context-specific dyadic attention vulnerabilities during the first year in infants later developing autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(1), 166–175.
MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131–141.
Marrus, N., Hall, L.P., Paterson, S.J., Elison, J.T., Wolff, J.J., Swanson, M.R., … Constantino, J.N. (2018). Language delay aggregates in toddler siblings of children with autism spectrum disorder. *Journal of Neurodevelopmental Disorders*, *10*(1), 29.
Mullen, E.M. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service.
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*(1), 115–128.
Newman, R.S., Kirby, L.A., Von Holzen, K., & Redcay, E. (2021). Read my lips! Perception of speech in noise by preschool children with autism and the impact of watching the speaker's face. *Journal of Neurodevelopmental Disorders*, *13*(1), 1–20.
Patterson, M.L., & Werker, J.F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior & Development*, *22*(2), 237–247.

Pons, F., Bosch, L., & Lewkowicz, D.J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, *26*(4), 490–498.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *32*(1), 123–142.

Rennig, J., & Beauchamp, M.S. (2018). Free viewing of talking faces reveals mouth and eye preferring regions of the human superior temporal sulcus. *NeuroImage*, *183*, 25–36.

Righi, G., Tenenbaum, E.J., McCormick, C., Blossom, M., Amso, D., & Sheinkopf, S.J. (2018). Sensitivity to audiovisual synchrony and its relation to language abilities in children with and without ASD. *Autism Research*, *11*(4), 645–653. https://doi.org/10.1002/aur.1918

Rosenblum, L.D., Schmuckler, M.A., & Johnson, J.A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*(3), 347–357.

Rowberry, J., Macari, S., Chen, G., Campbell, D., Leventhal, J.M., Weitzman, C., & Chawarska, K. (2015). Screening for autism spectrum disorders in 12-month-old high-risk siblings by parental report. *Journal of Autism and Developmental Disorders*, *45*(1), 221–229.

Shahin, A.J., & Miller, L.M. (2009). Multisensory integration enhances phonemic restoration. *The Journal of the Acoustical Society of America*, *125*(3), 1744–1750.

Shic, F. (2008). *Computational methods for eye-tracking analysis: Applications to autism*. New Haven, CT: Yale University.

Shic, F., Macari, S., & Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. *Biological Psychiatry*, *75*(3), 231–237.

Shic, F., Wang, Q., Macari, S.L., & Chawarska, K. (2020). The role of limited salience of speech in selective attention to faces in toddlers with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *61*(4), 459–469.

Smith, E.G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, *48*(8), 813–821.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service.

Spencer, M.D., Holt, R.J., Chura, L.R., Suckling, J., Calder, A.J., Bullmore, E.T., & Baron-Cohen, S. (2011). A novel functional brain imaging endophenotype of autism: The neural response to facial expression of emotion. *Translational Psychiatry*, *1*(7), e19.

Stevenson, R.A., & James, T.W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223. https://doi.org/10.1016/j.neuroimage.2008.09.034

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.

Summerfield, Q. (1979). Use of visual information in phonetic perception. *Phonetica*, *36*, 314–331.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 71–78.

Teinonen, T., Aslin, R.N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855.

Tenenbaum, E.J., Sobel, D.M., Sheinkopf, S.J., Shah, R.J., Malle, B.F., & Morgan, J.L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190.

van Wassenhove, V., Grant, K.W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC545853/pdf/pnas-0408949102.pdf

von dem Hagen, E.A., Nummenmaa, L., Yu, R., Engell, A.D., Ewbank, M.P., & Calder, A.J. (2011). Autism spectrum traits in the typical population predict structure and function in the posterior superior temporal sulcus. *Cerebral Cortex*, *21*(3), 493–500.

Wang, Q., Chang, J., & Chawarska, K. (2020). Atypical value-driven selective attention in young children with autism spectrum disorder. *JAMA Network Open*, *3*(5), e204928.

Weismer, S.E., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers with developmental delay. *Journal of Autism and Developmental Disorders*, *40*(10), 1259–1273.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, *26*(1–2), 23–43.

Young, G.S., Merin, N., Rogers, S.J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, *12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x

Accepted for publication: 21 January 2022