J Autism Dev Disord (2017) 47:1791–1805
DOI 10.1007/s10803-017-3098-0

ORIGINAL PAPER

Attempting to "Increase Intake from the Input": Attention and Word Learning in Children with Autism

Elena J. Tenenbaum[1,4] · Dima Amso[2] · Giulia Righi[3,4] · Stephen J. Sheinkopf[1,4]

Published online: 24 March 2017

Abstract Previous work has demonstrated that social attention is related to early language abilities. We explored whether we can facilitate word learning among children with autism by directing attention to areas of the scene that have been demonstrated as relevant for successful word learning. We tracked eye movements to faces and objects while children watched videos of a woman teaching them new words. Test trials measured participants' recognition of these novel word-object pairings. Results indicate that for children with autism and typically developing children, pointing to the speaker's mouth while labeling a novel object impaired performance, likely because it distracted participants from the target object. In contrast, for children with autism, holding the object close to the speaker's mouth improved performance.

Keywords Autism · Eye-tracking · Word-learning · Attention to faces · Joint attention

✉ Elena J. Tenenbaum
etenenbaum@wihri.org

[1] Brown Center for the Study of Children at Risk, Women and Infants Hospital, 101 Dudley Street, Providence, RI 02905, USA

[2] Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA

[3] Emma Pendleton Bradley Hospital, East Providence, RI, USA

[4] Department of Psychiatry & Human Behavior, Brown University, Providence, RI, USA

Introduction

Successful word learning requires matching a novel word to its referent in the world. The process involves gathering information from the speech stream (Rapin and Dunn 2003; Tager-Flusberg et al. 2005), recognizing joint attention cues (Baldwin 1991, 1993; Baldwin et al. 1996), monitoring the co-occurrence of label and object (Smith et al. 2011; Smith and Yu 2008; Trueswell et al. 2013) and finally, integrating information across those domains (Carter et al. 1998; Tager-Flusberg 2006). Autism is a developmental disorder that is often associated with language delay. These delays may result from atypical attention to information in the social environment (Arunachalam and Luyster 2015). Indeed, for some children with autism spectrum disorder (ASD), there is evidence of limited orienting to speech sounds (Čeponienė et al. 2003; Dawson et al. 1998), diminished response to joint attention cues (Carpenter et al. 2002; Dawson et al. 2004; Leekam et al. 1998), and limited understanding of intention in word learning tasks (Parish-Morris et al. 2007). With atypical performance in many of the skills thought to be relevant for early language acquisition, it is not surprising that many children with autism show delays in word learning (Charman et al. 2003; Luyster et al. 2007). This study explores whether we can facilitate word learning among children with autism by modifying the environment to draw attention to the most relevant sources of information.

Two main lines of research have demonstrated that attention to the social environment is related to early word learning in both typically developing children and children with autism. These are attention to faces (and in particular, the mouth of a speaker), and attention to objects (also referred to as gaze following or response to joint attention). Among typically developing infants, research has demonstrated that attention to faces and objects shifts as a function of the information most relevant to the child's current stage of language development.

In the first months of life, infants tend to fixate the eyes of a speaker or to distribute their attention evenly between the mouth and eyes (Hunnius 2007; Merin et al. 2007). As infants enter the babbling stage of language development between 6 and 8-months, attention seems to shift down to the mouth of the speaker (Lewkowicz and Hansen-Tift 2012; Rutherford et al. 2015; Tenenbaum et al. 2013). Primary focus seems to remain on the mouth until 12–18 months when attention shifts back to the eyes as children become more adept at receiving the auditory and visual input from their native languages (Lewkowicz and Hansen-Tift 2012; Rutherford et al. 2015). It is important to note that while there do seem to be trends in the patterns of attention to faces over the first years of life, at least within a given stimulus set, there is also significant variability even among typically developing infants (Tenenbaum et al. 2013). This variability has emerged as a meaningful predictor of later language abilities. Larger vocabulary size in toddlerhood has been observed among typically developing infants and infants at-risk for ASD who demonstrated increased attention to the mouth at 6 months (Young et al. 2009). A similar advantage in vocabulary growth during toddlerhood has also been observed among typically developing infants who focused their attention on the mouth at 12 months (Tenenbaum et al. 2014b).

Because children with ASD are not typically diagnosed until their 2nd or 3rd year, we know much less about the early face scanning patterns within this population. Prospective studies of infants at-risk for developing ASD that have explored face scanning show mixed patterns depending on the stimuli used. Relative to typically developing infants, infants who go on to develop ASD show a decline in attention to the eyes in the first six months when watching videos of a woman speaking (Jones and Klin 2013) while infants who went on to receive a diagnosis of ASD showed the reverse pattern (increased attention to the eyes) between 3 and 9 months when looking at still images of faces (Rutherford et al. 2015). Similar discrepancies have emerged in face scanning patterns of toddlers and children with ASD. Chawarska et al. (2016) found diminished attention to the mouth among children with ASD during the second year and Nakano and colleagues (Nakano et al. 2010) found similar results into early childhood. In contrast, von Hofsten et al. (2009) found comparable attention to the mouth in typically developing toddlers and those with ASD while Jones et al. (2008) showed that 2-year-olds with ASD looked *more* at the mouth than typically developing peers. These discrepancies are once again likely due to the specific set of stimuli used in each study (e.g. direction of gaze, presence of distractors). Face scanning patterns among infants at risk for ASD (Young et al. 2009) and toddlers with ASD (Chawarska et al. 2016) have both been linked with later language ability. Increased attention to the mouth has been shown to correlate with languge ability in teenagers with ASD (Norbury et al. 2009) and with social adaptation and lower levels of autism severity in high-functioning adults with ASD (Klin et al. 2002). Rice et al. (2012) suggest that the association between mouth fixation and social disability is moderated by IQ profiles. These authors posit that for highly verbal adults with ASD, attention to the mouth may offer a window into the interpretation of social cues.

In addition to the relevance of attention to the mouth and face for successful word learning, to learn the names for objects, a child must also orient to the object in question. Finding the object that is being labeled is often achieved by looking where someone else is looking. Typically developing children begin to display this gaze following behavior, also called responding to joint attention, between 6 and 12 months (Carpenter et al. 1998). Though results differ by experimental design, participant age, and language level (Akechi et al. 2011, 2013; Luyster and Lord 2009), a number of studies have found delays in response to joint attention among children with ASD (Carpenter et al. 2002; Dawson et al. 2004; Leekam et al. 1998). The tendency to follow a speaker's gaze to a target referent has been shown to predict language ability among typically developing children (Brooks and Meltzoff 2005, 2008; Carpenter et al. 1998; Morales et al. 2000; Mundy et al. 2007) and children with ASD (Bono et al. 2004; Dawson et al. 2004).

Two possible paths may lead children who are able to respond appropriately to joint attention to the correct matching between the label they are hearing and the object they are seeing. One is that by fixating that referent or possibly holding it up and moving it around, the speaker is drawing attention to that object. The co-occurrence of the increased salience of the object with the label is then sufficient to yield successful mapping of the label onto the target (Smith 2000). An alternative account suggests that children understand that the speaker intends to name that object and thus they draw the connection between the label and its referent (Baldwin and Moses 2001; Bloom and Markson 1998). Some have posited that the relative weight of these perceptual and social cues shifts over the course of development in favor of higher level social understanding (Hollich et al. 2000). Though recent evidence calls the developmental reweighting of these cues into question (Yurovsky and Frank 2015), the extant literature remains consistent with the view that both perceptual salience and social understanding contribute to successful word learning.

Children with autism have been shown to struggle with social cues in word learning tasks (Baron-Cohen et al. 1997) and demonstrate deficits in their understanding of intention in word learning paradigms (Parish-Morris et al. 2007). Gaze following among children with autism is necessary but not sufficient for successful language learning and recognition of the communicative relevance of gaze shifts are related to level of social impairment in children with autism (Gliga et al. 2012). Though there is evidence that higher functioning verbal children with autism are able to succeed in recognizing the social cues relevant for word learning (Luyster and Lord 2009), for those with lower verbal ability, recognizing the meaning of these social cues remains a struggle (Baron-Cohen et al. 1997; Preissler and Carey 2005).

In a recent review of lexical acquisition mechanisms in ASD, Arunachalam and Luyster (2015) suggested that while the underlying mechanisms guiding word learning in ASD may be in tact, a disconnect between the language information available in the environment and the information that the child with ASD takes from that input may be hindering their success. Arunachalam and Luyster further suggest that future research should address "How we can increase children's intake from the input" (p. 11). Given what we know about the limited success of some children with autism in interpreting social cues for word learning (e.g. gaze following and intention understanding), it seems that ratcheting up the perceptual cues (e.g. perceptual salience) might allow us to increase the intake from the input.

In line with this, Akechi and colleagues showed that increasing the salience of a target referent with movement (Akechi et al. 2011) or with use of multiple cues (e.g. gaze and pointing) (Akechi et al. 2013) can improve performance in a word learning task for children with ASD. In their experiments, Akechi and colleagues used simplified schematic drawings of faces and objects. Here, we were interested in exploring how increased salience of both the linguistic information available in the mouth and the perceptual salience of the target object might increase performance using more naturalistic stimuli of a person speaking. In two experiments, we attempted to improve word learning among children with autism by directing their attention to the areas of the scene we know to be most relevant for word learning. In Experiment 1 we focused on the mouth region as a target for attention and in Experiment 2 we directed their attention to both the mouth and the target object.

Experiment 1

In the first experiment, we explored whether we could facilitate word learning among typically developing children and those with ASD by shifting attention to the mouth of a speaker with a pointing cue. To do this, while teaching the participants the label for the novel object, the speaker pointed to her mouth in an attempt to direct attention to what has been demonstrated to be an informative area of the scene for language development (Chawarska et al. 2016; Tenenbaum et al. 2014a, b; Young et al. 2009).

Method

All procedures were approved by the institutional research ethics committee and have been performed in accordance with the ethical standards as laid down in the 1965 declaration of Helsinki and its later amendments. Informed consent was obtained from all participants included in the study.

Participants

Nineteen participants ($M=47.09$ months, $SD=12.62$; 14 male, 5 female) with diagnoses of ASD were recruited from early intervention centers, medical and mental health clinics and by referral from local providers. Diagnoses were confirmed with the Toddler Module ($n=1$), Module 1 ($n=14$), or Module 2 ($n=3$) of the Autism Diagnostic Observation Schedule 2nd Edition (Lord et al. 2012). One additional child with a community diagnosis of autism was included in the analysis though the family moved before the ADOS and other testing could be administered. Three additional children with ASD participated in the experiment, but were excluded from analysis because their scores on the Preschool Language Scales (PLS) (Zimmerman et al. 2011) were equivalent to 12 months or below.

Nineteen typically developing (TD) participants ($M=19.95$ months, $SD=9.23$; 14 male, 5 female) were recruited from an established research database of children born in the state as language matched controls. An additional four TD infants were recruited but excluded from analysis because calibration to the eye tracker could not be established ($n=2$), PLS scores were 12 months or below ($n=1$) or they had a sibling diagnosed with ASD ($n=1$). Cognitive and language test scores are provided in Table 1.

Parental Surveys

All parents completed a demographic form. Typical controls were screened for social impairment with parent report measures. Depending on the age of the child, we used the Communication and Symbolic Behavior Scales (CSBS) (Wetherby et al. 2002), the modified checklist for autism in toddlers (M-CHAT) (Robins et al. 2001), or the Social Communication Questionnaire (Rutter et al. 2003).

Table 1 Testing results for ASD and TD participants in experiments 1 and 2

	Experiment 1			Experiment 2		
	ASD ($n=19$)	TD ($n=19$)	F	ASD ($n=24$)	TD ($n=24$)	F
# Male	14	14	–	20	17	–
Age in months	47.09 (12.62)	19.95 (9.23)	64.57**	69.76 (20.92)	30.93 (20.05)	43.10**
PLS composite age equivalent	22.44 (9.24)	24.68 (10.73)	0.46	31.71 (20.04)	34.04 (21.33)	0.15
PLS composite scaled	62.83 (7.09)	104.47 (1.11)	110.74**	69.08 (28.23)	105.58 (10.47)	35.26**
Cognitive	67.06 (17.83)	102.95 (10.06)	57.69**	64.75 (20.30)	102.58 (13.51)	57.76**
ADOS						
Social affective	15.35 (2.60)	–	–	13.83 (3.55)	–	–
Repetitive behaviors	4.47 (2.43)	–	–	4.79 (2.17)	–	–
Severity score	8.06 (1.61)	–	–	7.58 (1.50)	–	–

Experiment 1: cognitive scores are standard scores from the Bayley Scales of Infant Development (BSID) (ASD: $n=7$; TD: $n=17$) or the Wechsler Preschool and Primary Scale of Intelligence-III (WPPSI-III) (ASD: $n=11$; TD: $n=2$). Experiment 2: BSID (ASD: $n=2$; TD: $n=15$), WPPSI-III (ASD: $n=20$; TD: $n=9$) or the Wechsler Intelligence Scale for Children-V (ASD: $n=2$).

PLS Preschool Language Scales, *ADOS* Autism Diagnostic Observation Schedule

**$p<.01$

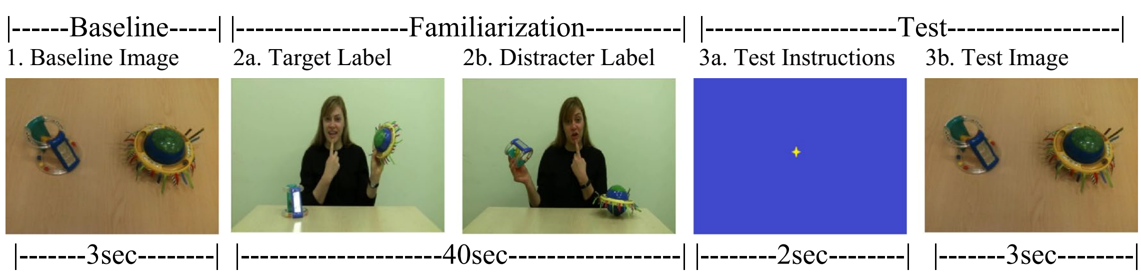


Fig. 1 Screen shots from a single trial in Experiment 1. **1** Baseline image of target and distractor objects during which baseline preferences are assessed, **2a** Familiarization screen shot from a Point trial during the target labeling segment (e.g. "Do you see the dobi? What a great dobi. Do you like the dobi?" **2b** Familiarization screen shot during the distracter labeling segment "Look at this! Do you see it? What a great thing!" **3a** Central fixation to direct attention to midline while instructions are played (e.g. "Look at the dobi"), **3b** Test segment when images reappear and preferential looking to the "dobi" relative to baseline is measured

Cognitive and Language Tests

All participants underwent cognitive and language testing administered by trained staff and supervised by a licensed clinical psychologist. Cognitive tests were either the Bayley Scales of Infant Development, 3rd edition (Bayley 2005), or the Wechsler Preschool and Primary Scale of Intelligence-III (Wechsler 2002), depending on the age and verbal ability of the child. All participants were tested using the Preschool Language Scales (PLS)—5th edition (Zimmerman et al. 2011). If cognitive or language testing had been completed by a licensed professional within the previous 6 months, the child's existing score was used.

Stimuli

Four familiar objects and 12 novel objects, matched on size, were used to create the stimuli. Each child saw eight trials: two training trials with the familiar objects (ball, car, duck, bear) followed by six test trials (3 Point, 3 No Point). Both the objects and the labels for the six test trials were selected to be novel. The labels provided were all disyllabic nonce words that are phototactically valid in English (*dobi, relo, ganer, pibo, blicket, flantu*). Two versions of the experiment were created such that across subjects, each object was presented in both a Point trial and a No Point trial. Each trial consisted of baseline preferential looking, familiarization, and a test phase (see Fig. 1). During the baseline

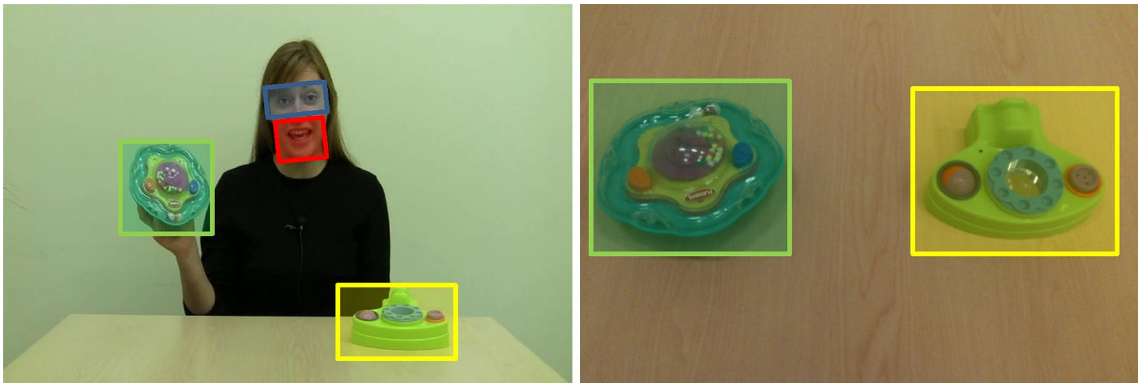


Fig. 2 Areas of interest (AOIs) for the familiarization and baseline/test phases. Familiarization AOIs were eyes, mouth, target, and distractor. Baseline/test AOIs were target and distractor

preferential looking (Fig. 1, 1), infants saw an image of two objects on screen for 3 s. This was used to determine baseline preferences for one object over the other.

The baseline phase was followed immediately by the familiarization phase (Figs. 1, 2). During familiarization, children saw videos of a woman sitting at a table with two objects in front of her. Following an animated greeting to signal communicative relevance (Senju and Csibra 2008), the woman picked up one of the objects (hereafter, the target object) and began labeling it. The speaker held the object steady while labeling it three times (e.g. "Look, it's a dobi! Wow, what a great dobi. Do you see the dobi?") (Fig. 1, 2a). Between each labeling phrase, the speaker shifted her gaze to the object. While labeling the object, she maintained her fixation on the camera. She then picked up the other object (hereafter, the distracter object) and described it but did not provide a label (e.g. "Wow, look at this. Do you see this thing? It's really something.") (Fig. 1, 2b). On Point trials, the speaker pointed to her mouth when she provided the labeling phrase for both the target and the distractor. On No Point trials, she kept her hand at her side. The familiarization portion of the trial lasted approximately 40 s. Target location and order of description (target or distractor first) were counterbalanced across trials. Order of trials was randomized across subjects.

After familiarization, a yellow plus sign was displayed in the center of a blue screen to draw the attention of the participants back to midline while the test instructions (e.g. "Look at the dobi!") were presented over the computer speakers (Fig. 1, 3a). The same image of the target and distracter objects that had been used for baseline then reappeared on the screen (the speaker was once again out of view during this phase) (Fig. 1, 3b). This final Segment of the test phase lasted 3 s. On training trials (the first two trials) children received positive feedback in the form of a happy face on screen while they heard "Yes!" or "Great!" once they had successfully oriented to the target familiar object during the test phase. On half the trials, the location of the objects was reversed between familiarization and test to ensure that the children were identifying the object rather than the location. Images used at test were identical to those used in the baseline phase

Eye Tracking

Eye tracking was completed on a SensoMotoric Instruments (SMI) RED500 Remote Eye Tracking system (SensoMotoric Instruments, Inc., Boston, MA). This system includes a remote-controlled infrared eye camera with automatic eye and head tracker. Tracking relies on a binocular image of the pupil and corneal reflection collected at a rate of 60 Hz with spatial resolution of 0.03° and gaze position accuracy of 0.4°. Blink recovery time is at maximum 4 ms and tracking recovery time for excessive movement is at maximum 90 ms.

An experimenter was seated at a computer adjacent to the display monitor, but hidden from view with a dark curtain. Calibration and stimulus presentation were displayed using the Experiment Center software provided by SMI. A two point calibration was used. This involved presentation of an animated image designed to attract the infant's attention to fixation points at the top left and lower right corners of the screen. Calibration was then verified with a four point display of animated objects. Average deviation from calibration in the X plane was 1.23° ($SD = 1.08$) and in Y, 1.47° ($SD = 1.11$). Deviation from calibration did not differ significantly between groups, X: $F(2,36) = 1.69$, $p = .20$; Y: $F(2,36) = 0.00$, $p = .97$.

Stimuli were presented on a 19″ (48.26 cm) computer monitor. Children were seated approximately 70 cm from the display monitor. At this distance, the woman's face was 7 cm × 6 cm (eyes: 5 cm × 4 cm, 4° × 3° visual angle; mouth: 4 cm × 5 cm, 3° × 4° visual angle; separated by a gap of approximately 2° visual angle). The objects were approximately 35 cm² during familiarization and 130 cm² during baseline and test.

Procedure

Children sat on a caretaker's lap or on a chair by themselves in front of the eye tracker and display screen. Each session began with the calibration described above. Once the child was successfully calibrated, testing began. At the start of each trial, an animated image was used to capture the child's attention. Once the child had oriented to the screen, presentation of the first trial began. The procedure lasted between 5 and 10 min.

Analysis

Eye tracking data was preprocessed using SMI BeGaze native software (SensoMotoric Instruments, Inc., Boston, MA). Fixations were defined as at least 100 ms spent fixating a 100 pixel area. For the baseline (Fig. 1) and test images (Fig. 1, 3b), there were two areas of interest (AOIs): target and distracter. For the familiarization phase (Fig. 1, 2a, b), we defined four dynamic AOIs, frame by frame: eyes, mouth, target, and distracter. Figure 2 shows a screen capture from the familiarization phase and the baseline/test images including the defined AOIs. Overall attention to the familiarization trials was calculated as the percentage of tracked trial time (maximum 40 s) that each participant spent looking at the screen. Time spent fixating the eyes, mouth, target, and distracter during familiarization focused on the portion of the familiarization trial during which the speaker was labeling the target object. The dependent measure of whether a child had encoded the label for the target object was the difference score between proportion of looks to the target (target/target + distracter) at test minus proportion of looks to the target (target/target + distracter) at baseline.

Results

Attentional Distribution Across Groups

Analyses explored group differences in time spent fixating the familiarization stimuli (Fig. 1, 2a, b), proportion of tracked time spent fixating the speaker's mouth, eyes, target, and distracter objects during target labeling (Fig. 1, 2a) and increases in attention to the target from baseline (Fig. 1) to test (Fig. 1, 3b). On average, participants were successfully tracked for 23.06 s ($SD = 7.84$) of the 40 s long familiarization phase (AD: $M = 19.18$ s $SD = 6.97$; TD: $M = 26.93$ s, $SD = 6.78$). One-way ANOVA revealed that children with autism looked significantly less at the stimuli than TD controls, $F(1,36) = 12.06$, $p < .01$. Given the variability in overall looking time, attention to the AOIs was calculated as proportion of tracked time spent fixating each AOI. A mixed model repeated measures ANOVA with between subjects factor of Group (ASD vs. TD) and within subjects factors of Condition (Point vs. No Point) and AOI (eyes, mouth, target, and distracter) was used to explore differences in attention to the specific regions of the screen by trial type. Analysis focused specifically on the target labeling segment of the familiarization trial (Fig. 1, 2a) during which the speaker pointed to her mouth and labeled the target object three times (or for the No Point trials, simply held the target object and labeled it three times). Analysis revealed significant effects of Condition, $F(1,36) = 5.36$, $p < .05$, partial $\eta^2 = 0.13$, and AOI, $F(3,34) = 22.59$, $p < .01$, partial $\eta^2 = 0.73$, as well as a significant Condition x AOI interaction, $F(3,34) = 6.06$, $p < .01$, partial $\eta^2 = 0.25$. There was not a significant effect of Group or any significant interactions involving the Group variable.

Planned contrasts revealed that across groups while the speaker was labeling the target object, participants looked more to the target than the distracter, $F(1,35) = 40.12$, $p < .01$, partial $\eta^2 = 0.53$, and more at the mouth than the eyes, $F(1,35) = 25.82$, $p < .01$, partial $\eta^2 = 0.43$. The interaction between Condition and AOI was driven by a decrease in attention to the target object, but not the distracter, on Point trials relative to No Point trials, $F(1,35) = 9.72$, $p < .01$, partial $\eta^2 = 0.22$, and by a marginally significant trend for participants to look more at the mouth and less at the eyes on Point trials relative to No Point Trials, $F(1,35) = 3.99$, $p = .05$, partial $\eta^2 = 0.10$. This suggests that pointing to the speaker's mouth served to decrease attention to the target object, but with only marginally significant increases in attention to the mouth. For proportion of looks by Group and Condition during this segment of the familiarization phase, see Fig. 3.

Performance at Test

Word learning within the task was quantified as the difference score between attention to the target object after familiarization with the label (Fig. 1, 3b) and attention to the target at baseline (Fig. 1). Mixed model repeated measures ANOVA with a between subjects factor of Group and a within subjects factor of Condition revealed a marginally significant effect of Condition, $F(1,36) = 3.19$, $p = .08$, partial $\eta^2 = 0.08$. The effect of Group was not significant, nor was the interaction between Condition and Group. For both TD and ASD participants, attention to the target increased from baseline more in the No Point condition than in the Point condition. Contrary to our predictions, this result suggests that participants tended to show enhanced recognition of the target referent when the speaker did not point

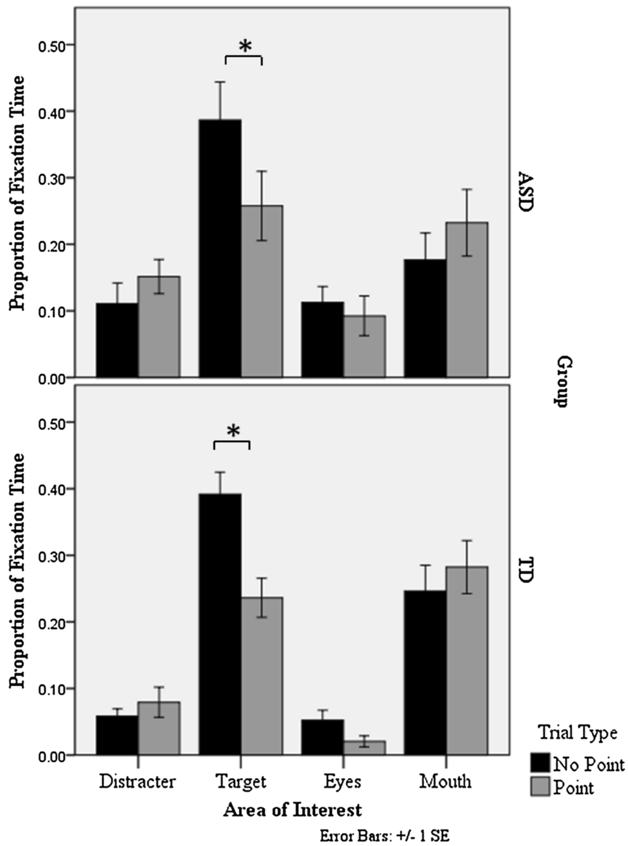


Fig. 3 Proportion of looks to the AOIs by condition and group during the labeling of the target object for Experiment 1. On Point Trials, the speaker was pointing to her mouth during this time. On No Point Trials, her arm was by her side. In both cases, she was holding the target object while describing it, *$p < .05$

to her mouth while labeling the target object (though the effect was only marginally significant). To further explore whether participants had successfully encoded the label for the target object between baseline and familiarization, we ran *t*-tests within groups comparing attention to the target at test (Fig. 1, 3b) and baseline (Fig. 1, 1) within each condition. In fact, neither group looked reliably more at the target after familiarization in either condition (ASD: No Point, $t(18) = -0.84$, $p = .41$, Point, $t(18) = -0.26$, $p = .80$; TD: No Point, $t(18) = 0.84$, $p = .41$, Point, $t(18) = 0.26$, $p = .80$). Figure 4 shows performance at test relative to baseline by Group and Condition.

Summary

In Experiment 1, we explored whether we might be able to enhance word learning in a laboratory context by directing attention to the mouth of a speaker while she provided the novel label for a target object. Though the manipulation did have the marginally significant effect of increasing attention to the mouth, it also had the significant effect of decreasing attention to the target object. Reduced attention to the target object likely contributed to the unpredicted result that performance at test was marginally worse in the Point condition relative to the No Point condition across groups. It seems that by directing attention to the mouth, we distracted participants from the target object and impaired their ability to encode the novel label. This does not necessarily suggest that attention to the mouth itself was deleterious, rather reduced attention to the target object in the word learning context may have impaired performance.

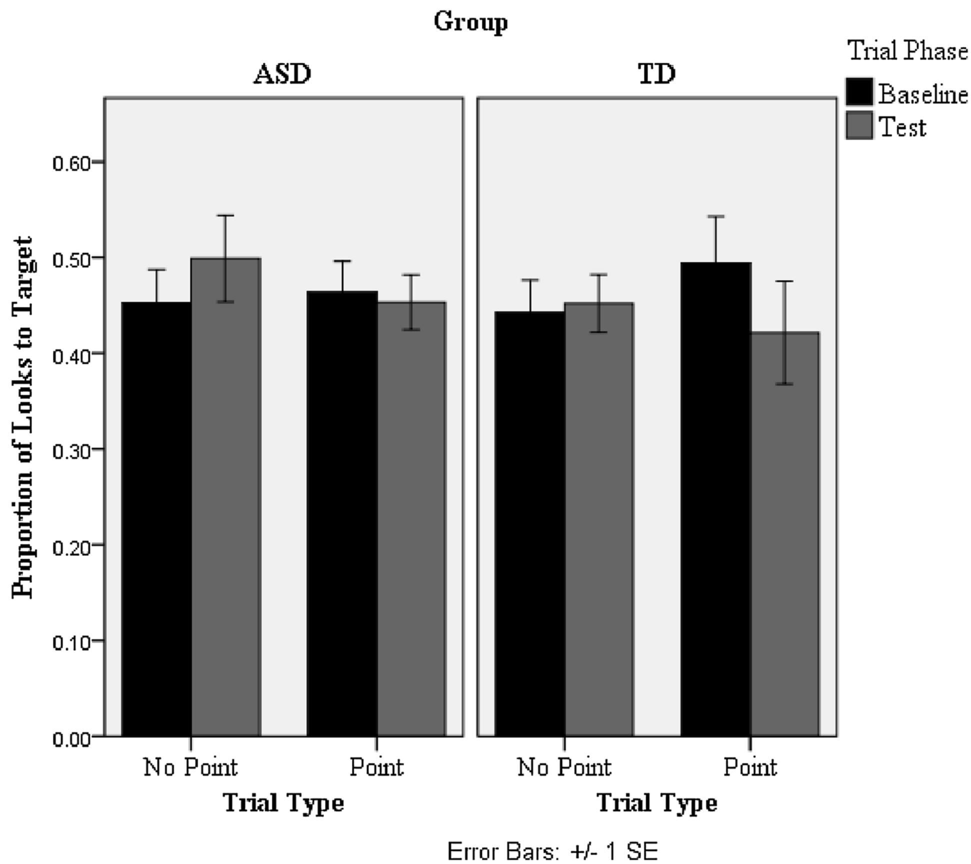
An alternative account for the unexpected results in Experiment 1 is that the participants interpreted the point to the mouth as referential and thus assigned the novel label to the mouth rather than the novel object. Recent work by Grassmann and Tomasello (2010) showed that when confronted with conflicting pointing and labeling cues, nearly 100% of 2 and 4-year-olds followed the pointing cue. For example, when presented with a car and a novel object, children responded by giving the examiner the car if she was pointing to it, even though the instruction was to "Give me the *modi*". This suggests that the pointing cue overrode the label in that context and may by extension mean that participants in this experiment interpreted the speaker's point as intended to label the mouth rather than the target object as the *dax*. While we cannot rule out this possibility, we believe that the pointing cue in the current study was more akin to that of Jaswal and Hansen (2006) in that it necessarily lacked gaze alternation between the mouth and the child (as one cannot shift attention to one's own mouth). In the Jaswal and Hansen study, the point cue was similarly non-ostensive (i.e. did not include a gaze shift) and 3–4-year-olds relied on the label (*modi*) rather than the point (car) to identify the referent.

It is important to note that neither group in Experiment 1 demonstrated successful word learning in this task (i.e. attention to the target object did not reliably increase from baseline to test following familiarization with the target label). This was likely due to the difficulty of the task and the relatively low language levels of the participants (approximately 2 years, on average). Work by Yurovsky and Frank (2015) suggest that typically developing children younger than 2 years struggled with novel word learning in a similar paradigm.

Experiment 2

Experiment 2 was designed to address two issues from Experiment 1: (1) Drawing attention to the mouth resulted in distraction from the target object, and (2) The task was too difficult. To resolve the first issue, we attempted to draw attention to both the speaker's mouth and the target

Fig. 4 Experiment 1 proportion of looks to the target at test relative to baseline. Proportion of looks to target at baseline and at test (proportion = target/target + distractor) across No Point and Point conditions by group. Neither group showed reliable increases in attention to the target at test relative to baseline



object simultaneously during familiarization. To address the task difficulty, a number of changes were made to the experimental design. First, the target label was repeated four times instead of three. Second, the target object was selected to be less salient than the distracter to allow for increases in attention to the target between baseline and test (if the child was focused primarily on the target at baseline in Experiment 1, there was little room to increase their attention to that object at test). Finally, participants with higher language abilities were recruited for Experiment 2.

Method

Participants

Twenty-four participants ($M = 69.76$ months, $SD = 20.92$; 20 male, 4 female) with a diagnosis of ASD who were producing at least single words were recruited from early intervention centers, medical and mental health clinics and by referral from local providers. Diagnoses were confirmed with Module 1 ($n = 16$), or Module 2 ($n = 8$) of the Autism Diagnostic Observation Schedule 2nd Edition (Lord et al. 2012). Two additional children with ASD participated but were excluded from the analysis because they were unable to sit for the experiment ($n = 1$) or because their PLS scores were 12 months or below ($n = 1$).

Twenty-four TD participants ($M = 30.93$ months, $SD = 20.05$; 17 male, 7 female) were recruited from an established research database of children born in the state as language matched controls. Four additional TD children were recruited but excluded from analysis because they were unable to sit for the experiment ($n = 2$), or because English was not their native language ($n = 2$). Testing results are provided in Table 1.

Procedure

Screening and experimental visit procedures were identical to the section "Experiment 1."

Stimuli

Stimuli were comparable to those used in Experiment 1 with the following modifications. (1) Instead of Point and No Point trials, there were three conditions: Far, Near, and Cover. On Far trials the speaker held an object off to the side while labeling it as she had in No Point trials in Experiment 1. In this case, the label was provided four times (e.g., "Look, it's a dobi! Wow, what a great dobi. Do

Fig. 5 Screen captures from the familiarization phases of the three conditions in Experiment 2: Cover, Far, and Near

you see the dobi? What a great dobi."). On Near trials, the speaker held the object next to her mouth while labeling it. On Cover trials, the speaker held the object in front of her mouth. This condition was added to explore what effect removing information from the mouth has on word learning. In all three trial types, the speaker then manipulated the distracter as she had the target while drawing attention to but not labeling the object (e.g. Look at this! Do you see it? What a great thing. It's nice.). Order of presentation of the target and distracter were again counterbalanced across trials. (2) Four familiar objects and 16 novel objects were used to create the stimuli. Objects were paired such that the target object would be less salient (smaller, less colorful, or less interesting) than the distracter object. The relative salience of the objects was confirmed with pilot testing in which typically developing participants looked more at the distracter objects than the target objects when shown the baseline images.

Following calibration and validation as described in the section "Experiment 1," each participant saw ten trials: two training trials with the familiar objects (ball, car, duck, bear) followed by eight test trials distributed across conditions such that each participant saw all three trial types.[1] Familiarization lasted approximately 50 s. Fig. 5 shows a captured image from familiarization trials in each of the conditions. As in Experiment 1, each trial consisted of baseline, familiarization (target and distracter labeling), and test phases (reorientation to midline while test instructions played and finally, test image).

[1] Two "filler" test trials were included in the experimental design in the event that preferences for the target versus distracter objects reversed between pilot testing (which was based on typically developing participants only) and data collection. In the end, experimental data showed the same patterns of preference as pilot testing (see results), so these filler trials were retained in the analysis. These "bonus trials" were evenly distributed across the three conditions and three experimental versions (e.g. one version of the experiment had one extra Far and one extra Near trial, another version of the experiment had one extra Far and one extra Cover trial and the third version of the experiment had one extra Near and one extra Cover trial).

Eye Tracking

Eye tracking procedures were identical to those used in Experiment 1. Average deviation from calibration in the X plane was 1.69° (*SD* = 1.89) and in Y, 1.73° (*SD* = 1.44). Deviation from calibration did not differ significantly between groups, X: $F(2,44) = 0.99$, $p = .33$; Y: $F(2,44) = 1.38$, $p = .25$.

Stimuli were presented on a 19″ (48.26 cm) computer monitor. Children were seated approximately 70 cm from the display monitor. At this distance, the woman's face was 5 cm × 7 cm (eyes: 4 cm × 3 cm, 3° × 2.5° visual angle; mouth: 3 cm × 4 cm, 2.5° × 3° visual angle; separated by a gap of approximately 2° visual angle). The objects were approximately 20 cm² during familiarization and 50 cm² during test.

Results

Attentional Distribution Across Groups

We first explored group differences in time spent fixating the familiarization stimuli overall and proportion of tracked time spent fixating the speaker's mouth, eyes, target, and distracter objects during the object labeling segments of the familiarization phase. Here, we analyzed proportions of looking time during the target and distracter labeling segments of the familiarization trials separately to explore the effects of our salience manipulation. Of note, because on Cover trials the object was largely obstructing the mouth, looks to the object were also counted as mouth fixations, and vice versa while the two were overlapping. On average, participants were successfully tracked for 22.73 s ($SD = 7.58$) of the 50 s long familiarization stimuli (AD: $M = 19.29$ s $SD = 5.42$; TD: $M = 26.34$ s, $SD = 7.96$). One-way ANOVA revealed that children with autism looked significantly less at the stimuli than TD controls, $F(1,39) = 11.10$, $p < .01$. Given the variability in overall looking time, attention to the AOIs was calculated as proportion of tracked time during the target and distracter labeling in which the participant fixated each AOI within a given trial.

To confirm that our salience manipulation had worked, we ran paired *t*-tests on looks to the target versus distracter objects during the baseline presentation of the image of the objects. These comparisons reached significance in 7 of the 8 stimulus pairs with *p*-values ranging from 0.000 to 0.03. For one stimulus pair, though as expected, mean time spent fixating the distracter at baseline was greater than that of the target, the difference did not reach significance, $p = .54$. This object was retained for analysis because no preference for the target object was observed.

A mixed model repeated measures ANOVA with between subjects factor of Group (ASD vs. TD) and within subjects factors of Familiarization Segment (Target vs. Distracter Labeling), Condition (Cover vs. Far vs. Near) and AOI (eyes, mouth, target, and distracter) was used to explore differences in attention to the specific regions of the screen by Condition and Trial Segment. Analysis revealed significant effects of Condition, $F(2,42) = 53.94$, $p < .001$, partial $\eta^2 = 0.72$, and AOI, $F(3,41) = 17.05$, $p < .001$, partial $\eta^2 = 0.56$, and significant interactions between Condition and AOI, $F(6,38) = 5.38$, $p < .001$, partial $\eta^2 = 0.46$, and Segment and AOI, $F(3,41) = 118.97$, $p < .001$, partial $\eta^2 = 0.90$. There was also a significant three-way interaction between Segment, AOI, and Condition, $F(6,38) = 4.07$, $p = .003$, partial $\eta^2 = 0.39$. No other main effects or interactions were significant.

Planned contrasts revealed that across groups, participants looked more at the mouth than the eyes of the speaker, $F(1,43) = 12.47$, $p < .001$, partial $\eta^2 = 0.23$ and more at the distracter (the more salient object) than the target over the combined target and distracter labeling segments of the familiarization phase, $F(1,43) = 35.82$, $p < .001$, partial $\eta^2 = 0.45$. Participants across groups looked less at the stimuli in general on Far trials than on Cover trials, $F(1,43) = 108.80$, $p < .001$, partial $\eta^2 = 0.72$, or Near trials, $F(1,43) = 6.63$, $p = .01$, partial $\eta^2 = 0.13$. The interaction between Segment and AOI was driven by greater attention to the target during the target labeling segment and greater attention to the distracter during distracter labeling, $F(1,43) = 313.85$, $p < .001$, partial $\eta^2 = 0.88$. The interaction between AOI and Condition was driven by significantly greater attention to the mouth relative to the eyes on Cover trials relative to Far trials, $F(1,43) = 24.08$, $p < .001$, partial $\eta^2 = 0.36$. This is likely due to the overlap between the objects and the mouth on Cover trials. The three-way interaction between Segment, AOI, and Condition was driven by greater difference between attention to the target and distracter (more attention to the distracter) during distracter labeling on Cover relative to Far trials in contrast with greater difference between target and distracter (more attention on the target) on Cover relative to Far trials during target labeling, $F(1,43) = 20.67$, $p < .001$, partial $\eta^2 = 0.33$. For proportion of looks by Group and Condition during the two segments of the familiarization trials, see Fig. 6.

Performance at Test

Word learning within the task was quantified as the difference score between attention to the target object after familiarization with the label and attention to the target at baseline. Mixed model repeated measures ANOVA with a between subjects factor of Group and a within subjects factor of Condition revealed no significant effects of Condition, $F(2,34) = 0.47$, $p = .63$, partial $\eta^2 = 0.03$, or Group, $F(1,35) = 0.25$, $p = .62$, partial $\eta^2 = 0.01$, nor any significant interactions, $F(2,34) = 1.09$, $p = .35$, partial $\eta^2 = 0.06$. To explore whether participants had successfully encoded the label for the target object between baseline and familiarization, we ran paired comparison *t*-tests within groups for each condition. TD participants showed a marginal increase in looks to the target at test relative to baseline in the Far condition, $t(22) = -1.78$, $p = .09$, and reliable increases in both the Cover, $t(23) = -2.42$, $p = .02$, and Near conditions, $t(21) = -2.58$, $p = .02$ In contrast, ASD participants did not show any increase in looks to the target on the Cover, $t(21) = 0.03$, $p = .98$, or Far conditions, $t(20) = -1.00$, $p = .33$, but did show a reliable increase in the Near condition, $t(21) = 2.48$, $p = .02$. Figure 7 shows performance at test relative to baseline by Group and Condition.

Summary

Experiment 2 allowed us to address two issues that emerged in Experiment 1. First, the task of identifying a novel target (i.e. increasing looks to the target from baseline) after a brief familiarization proved quite challenging for ASD and TD participants alike in Experiment 1. To resolve this, we made three changes to the experimental design: (1) the target label was repeated four times instead of three, (2) the novel object was selected to be less salient than the target to allow for increased attention to the target between baseline and test, and (3) children who were not yet producing any

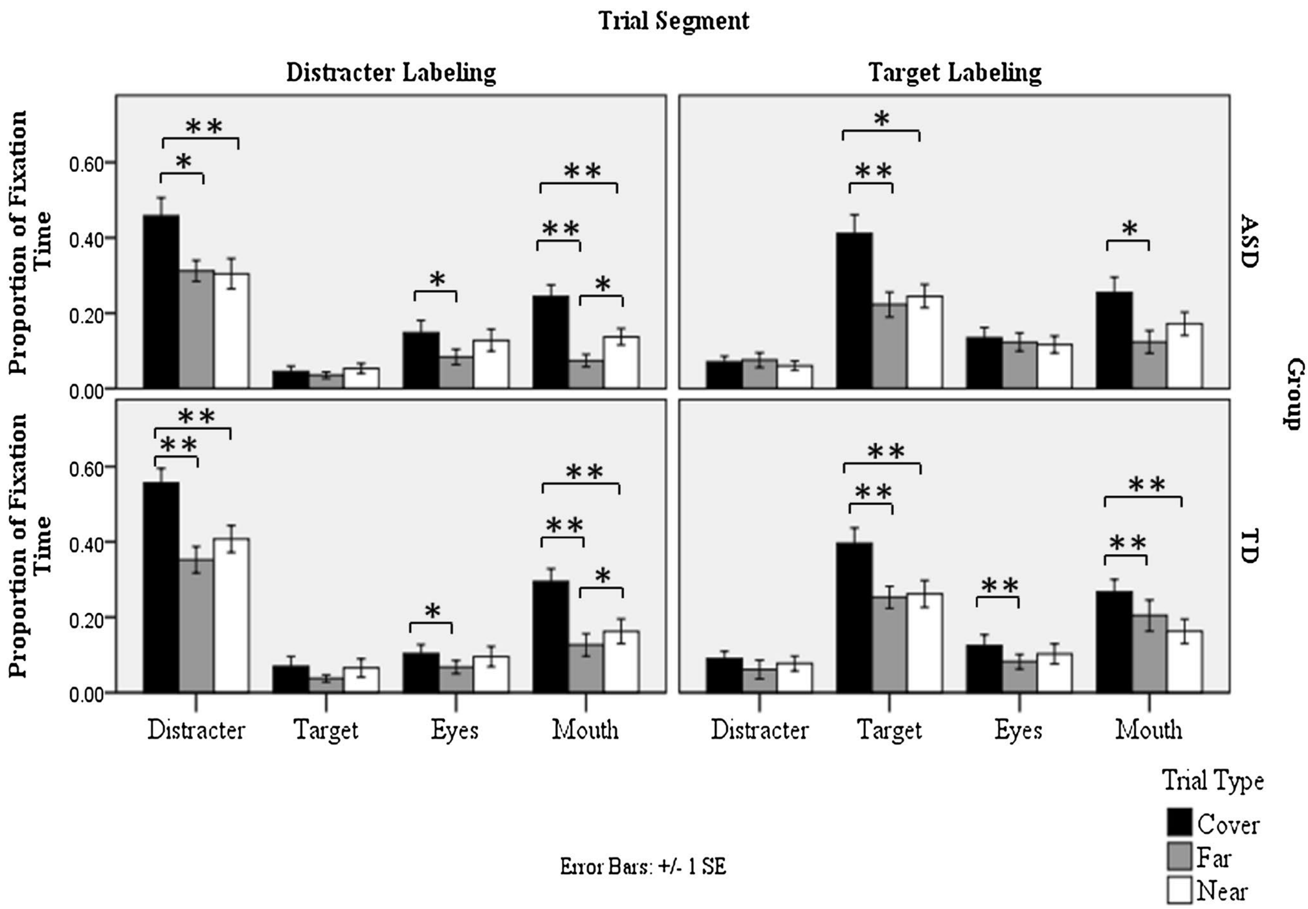


Fig. 6 Proportion of looks to the AOIs by Condition and Group in Experiment 2. On Cover trials, the object was held in front of the speaker's mouth. On Far trials, the object was held off to the side and on Near trials, the object was near the speaker's mouth, **p* <.05, ***p* <.01

words were excluded from participation. These modifications resulted in successful recognition of the target object for the language matched TD participants in all conditions (Far, Near, and Cover).

The second issue we targeted in Experiment 2 was the distraction from the target object that had resulted from the pointing manipulation in Experiment 1. To address this, the speaker in Experiment 2 held the target object near her mouth while providing the labels so as to draw attention to the mouth and the object simultaneously. In contrast to the pointing manipulation in Experiment 1, this manipulation served to improve performance among the participants with ASD (while the TD participants were successful with or without this manipulation).

General Discussion

The way a child looks at the world around her has been shown to impact her word learning and language development. Both attention to the face (Chawarska et al. 2016; Tenenbaum et al. 2014a, b; Young et al. 2009) and response to joint attention (Bono et al. 2004; Brooks and Meltzoff 2005, 2008; Carpenter et al. 1998; Dawson et al. 2004; Morales et al. 1998; Mundy et al. 2007) have been associated with superior language skills in typically developing children and children with ASD. This study attempted to boost word learning by directing attention to the areas of the screen most relevant for language learning. In Experiment 1, the speaker pointed to her mouth while labeling an object and in Experiment 2, she held the object close to her mouth while labeling it. Results show that while the pointing cue did increase attention to the speaker's mouth somewhat, it also pulled attention away from the target object. This distraction resulted in diminished recognition of the target object at test. In contrast, when the speaker held the object near her mouth, children with ASD demonstrated success in recognizing the target object at test where they failed to do so if she held the object further away or covered her mouth with the object. TD participants succeeded in all three conditions in Experiment 2 (Cover, Far, and Near), though

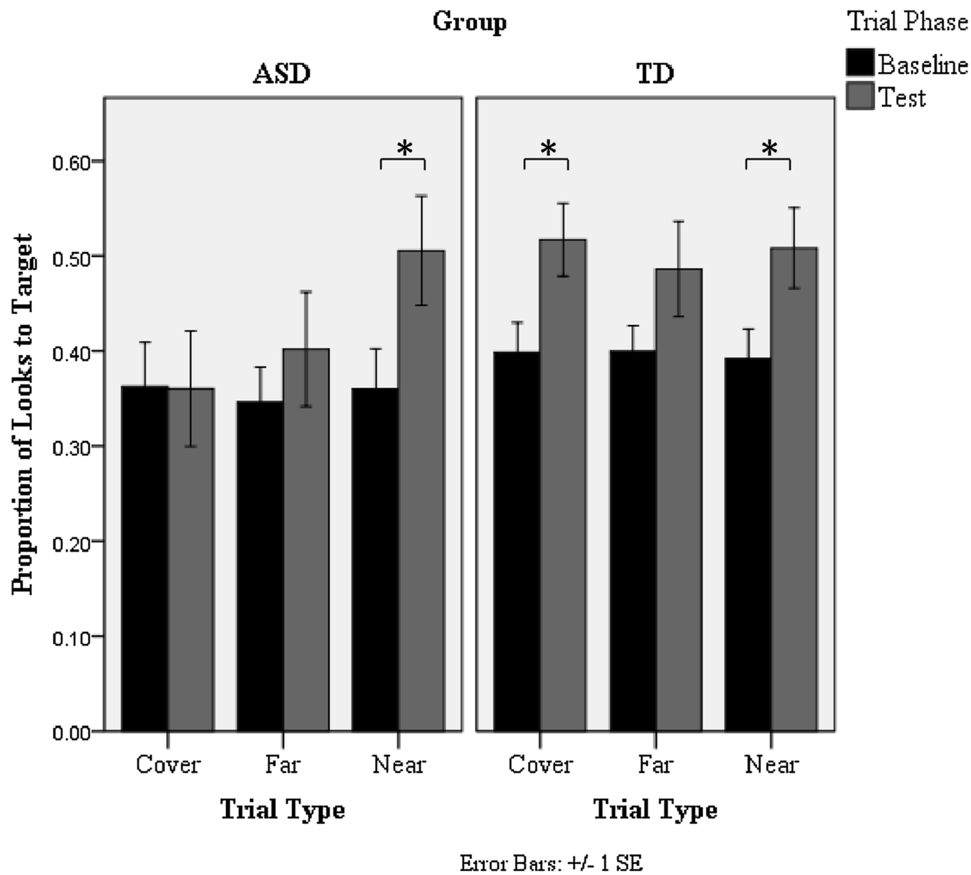


Fig. 7 Experiment 2 proportion of looks to the target at test relative to baseline. Proportion of looks to target at baseline and at test (proportion = target/ target + distractor) across the three conditions by group for Experiment 2. The TD participants showed reliable increases in attention to the target at test relative to baseline in the Cover and Near conditions, and a marginally significant effect in the Far condition. ASD Participants only showed reliable increase in attention to the target in the Near condition, *$p < .05$

the effect was only marginally significant in the Far condition.

Results from Experiment 1 were unexpected. We believe the most likely account for these results is that pointing to the mouth of the speaker distracted the participants from the target object and resulted in failure to connect the object with the target label. As noted above, it is also possible that failure to successfully encode the target referent in the Point condition may have been related to the participants' understanding of intentions. Interpreting a speaker's intent to label is known to be challenging for children with ASD (Parish-Morris et al. 2007) and may have been further compromised by the point to the mouth. As this experiment was not designed to assess intentional understanding, we cannot say whether this in fact played a part in the results.

Results from Experiment 2 are consistent with findings by Akechi and colleagues (Akechi et al. 2011, 2013) who showed that increasing the salience of a target object with movement (Akechi et al. 2011) and adding the social cue of pointing to the object (Akechi et al. 2013) facilitated word learning among older children with autism watching a schematic of a face and pointing gesture. Here we have extended these results to children in the early word learning stages of development using video stimuli of a person and using the simple modification of shifting the object a few inches closer to the speaker's face. Though the Far condition in this experiment provided equivalently redundant cues as the look and pointing condition used by Akechi et al. (2013), the younger age range of the participants and lack of simplified schematics may have prevented the children with ASD from succeeding in this condition in the current study.

Yurovsky and Frank (2015) suggest that a significant feature of development in word-learning paradigms such as this is the ability to flexibly shift attention between the face and objects. Specifically, over time, children develop the capacity to successfully disengage from the speaker's face in order to see the object. In this experiment, by holding the object very close to the speaker's face, we have eliminated the need for children to disengage from one source of information to gain access to the other. It is not entirely clear whether the children with ASD would have the same difficulty disengaging from the face that the typically developing population in Yurovsky and Frank's study had. It is possible that for the children with ASD, the difficulty in the Far condition would have been disengaging from the object in favor of the face (Chawarska et al. 2010). In either case, keeping the two areas adjacent to each other seems to have allowed participants to gain information from both regions simultaneously.

For TD participants, the Cover condition may also have served to maintain engagement on the face and object resulting in successful recognition of the target. This is interesting in light of the fact that the speaker's utterances were somewhat muffled in the Cover condition. It is possible that covering the mouth of the speaker made the Cover trials more salient to the TD participants as this was an odd thing for the speaker to have done while speaking. If the salience of this somewhat odd condition was driving this effect, however, one might have expected the Cover condition to have resulted in improved performance for the ASD participants as well. It is also possible that the TD participants in this condition were attempting to locate the source of the sound and this search led them to focus on the target object.

In line with Arunachalem and Luyster's suggestions that the mechanisms for word learning remain in tact among children with autism (but see Norbury et al. 2010), these results, in combination with previous findings (Akechi et al. 2011, 2013; Parish-Morris et al. 2007), suggest that word learning can proceed effectively if we can increase the intake from the information available in the environment (Arunachalam and Luyster 2015). If we can capture the attention of the child with autism, either by making the target more interesting (Akechi et al. 2011; Parish-Morris et al. 2007), by providing redundant social and salient cues (Akechi et al. 2013; Luyster and Lord 2009), or by implementing child specific orienting approaches (Koegel et al. 2009), we may be able to improve language acquisition in this population. Further, the importance of attracting attention to the intended referent may continue beyond childhood as even adults with ASD have been shown to struggle in word learning paradigms when perceptual salience cues diverted from social cues (Aldaqre et al. 2015).

It is interesting to note that the participants with ASD in this experiment succeeded in recognizing the target referent at test if the speaker held the object near her mouth during familiarization even though the target object was less salient than the distracter object. The fact that the children with ASD were able to overcome the lack of salience of the target object, which is often a barrier in word learning (Akechi et al. 2011) speaks to the strength of the proximity cues used in this experiment.

In sum, these results suggest that attention to the mouth alone (Point trials in Experiment 1) or to the object alone (Cover trials in Experiment 2) are not sufficient to support word learning among children with autism. Rather, attention to both the mouth and the target (Near trials in Experiment 2) were required for the children with autism to succeed in this word learning task. This has clear implications for interventions targeting word learning among children with ASD. While some speech pathologists report drawing attention to the mouth in interventions with children with ASD, they are more likely to do so when attempting to elicit expressive language from the child (A. Weiss, Personal communication, August 9, 2016). This study suggests that drawing attention to the mouth during receptive tasks may also be worthwhile, but that it is important to direct attention to the target referent simultaneously.

Acknowledgments Special thanks to Madeline Mahowald and Emily Davis for their assistance with the stimuli and to the participant families.

Funding This research was supported by grants from the Autism Science Foundation (Grant No.11-1013), Autism Speaks (Grant No. 7897), and the National Institute of Mental Health (NIMH) (Grant No. 5T32MH019927-24).

Author contribution ET helped in conception of the study, participated in its design, coordinated data collection, performed the statistical analysis, and drafted the manuscript; DA participated in the design and interpretation of the data; GR helped in conception and design of the study, and participated in interpretation of the data; SS participated in the design of the study, data collection, and interpretation of the data. All authors read and approved the final manuscript.

References

Akechi, H., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2013). Brief report: pointing cues facilitate word learning in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(1), 230–235. doi:10.1007/s10803-012-1555-3.

Akechi, H., Senju, A., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2011). Do children with ASD use referential gaze to learn the name of an object? An eye-tracking study. *Research in Autism Spectrum Disorders, 5*(3), 1230–1242.

Aldaqre, I., Paulus, M., & Sodian, B. (2015). Referential gaze and word learning in adults with autism. *Autism: The International Journal of Research and Practice, 19*(8), 944–955.

Arunachalam, S., & Luyster, R. J. (2015). The integrity of lexical acquisition mechanisms in autism spectrum disorders: A research review. *Autism Research, 9*(8), 810–828.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*(5), 875–890.

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*(5), 832–843.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, 67*(6), 3135–3153.

Baldwin, D. A., & Moses, L. J. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*(3), 309–329 doi:10.1111/1467-9507.00168.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development, 68*(1), 48–57.

Bayley, N. (2005). *Bayley scales of infant development, 3rd ed.*. San Antonio: PsychCorp.

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences, 2*(2), 67–73.

Bono, M. A., Daley, L. T., & Sigman, M. (2004). Relations among joint attention, amount of intervention and language gain in autism. *Journal of Autism and Developmental Disorders, 34*(5), 495–505.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8*(6), 535–543. doi:10.1111/j.1467-7687.2005.00445.x.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language, 35*(1), 207–220.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society of Research in Child Development, 63*, i–vi, 1–174.

Carpenter, M., Pennington, B. F., & Rogers, S. J. (2002). Interrelations among social-cognitive skills in young children with autism. *Journal of Autism and Developmental Disorders, 32*(2), 91–106.

Carter, A. S., Volkmar, F. R., Sparrow, S. S., Wang, J.-J., Lord, C., Dawson, G., ... Schopler, E. (1998). The Vineland Adaptive Behavior Scales: Supplementary norms for individuals with autism. *Journal of Autism and Developmental Disorders, 28*(4), 287–302.

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech–sound-selective auditory impairment in children with autism: They can perceive but do not attend. *Proceedings of the National Academy of Sciences, 100*(9), 5567–5572. doi:10.1073/pnas.0835631100.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Drew, A., & Cox, A. (2003). Predicting language outcome in infants with autism and pervasive developmental disorder. *International Journal of Language & Communication Disorders, 38*(3), 265–285. doi:10.1080/136820310000104830XDYKN6D5P3YP2M6H.

Chawarska, K., Volkmar, F., & Klin, A. (2010). Limited attentional bias for faces in toddlers with autism spectrum disorders. *Archives of General Psychiatry, 67*(2), 178–185.

Chawarska, K., Ye, S., Shic, F., & Chen, L. (2016). Multilevel differences in spontaneous social attention in toddlers with autism spectrum disorder. *Child Development, 87*(2), 543–557.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders, 28*(6), 479–485. doi:10.1023/a:1026043926488.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A. & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology, 40*(2), 271–283. doi:10.1037/0012-1649.40.2.2712004-11032-012.

Gliga, T., Elsabbagh, M., Hudry, K., Charman, T., Johnson, M. H., & The BASIS Team. (2012). Gaze following, gaze reading, and word learning in children at risk for autism. *Child Development, 83*, 926–938. doi:10.1111/j.1467-8624.2012.01750.x.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The autism diagnostic observation schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders, 37*(4), 613–627.

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science, 13*(1), 252–263.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., ... Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, i–135.

Hunnius, S. (2007). The early development of visual attention and its implications for social and cognitive development. In C. VonHofsten & K. Rosander (Eds.), *From action to cognition* (Vol. 164, pp. 187–209). Oxford: Elseiver.

Jaswal, V. K., & Hansen, M. B. (2006). Learning words: Children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental Science, 9*(2), 158–165.

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry, 65*(8), 946–954. doi:10.1001/archpsyc.65.8.946.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature, 504*(7480), 427–431.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*(9), 809–816. doi:10.1001/archpsyc.59.9.809.

Koegel, R. L., Shirotova, L., & Koegel, L. K. (2009). Brief report: Using individualized orienting cues to facilitate first-word acquisition in non-responders with autism. *Journal of Autism and Developmental Disorders, 39*(11), 1587–1592.

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and cues: Gaze-following in children with autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 39*(7), 951–962.

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*, 1431–1436. doi:10.1073/pnas.1114783109.

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

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule-2nd edition (ADOS-2)*. Los Angeles: Western Psychological Corporation.

Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur-Bates communicative development inventory (CDI). *Journal of Child Language, 34*(03), 623–654. doi:10.1017/S0305000907008094.

Luyster, R., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology, 45*(6), 1774.

Merin, N., Young, G., Ozonoff, S., & Rogers, S. (2007). Visual fixation patterns during reciprocal social interaction distinguish a subgroup of 6-month-old infants at-risk for autism from comparison infants. *Journal of Autism and Developmental Disorders, 37*(1), 108–121.

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development, 21*, 373–377.

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention Across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology, 21*, 283–298.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development, 78*(3), 938–954. doi:10.1111/j.1467-8624.2007.01042.x.

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., ... Kitazawa, S. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20100587.

Norbury, C. F., Brock, J., Cragg, L., Einav, S., Griffiths, H., & Nation, K. (2009). Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry, 50*(7), 834–842. doi:10.1111/j.1469-7610.2009.02073.x.

Norbury, C. F., Griffiths, H., & Nation, K. (2010). Sound before meaning: word learning in autistic disorders. *Neuropsychologia, 48*(14), 4012–4019.

Parish-Morris, J., Hennon, E. A., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child development, 78*, 1265–1287.

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: evidence from autism. *Cognition, 97*(1), B13–B23.

Rapin, I., & Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain and Development, 25*(3), 166–172.

Rice, K., Moriuchi, J. M., Jones, W., & Klin, A. (2012). Parsing heterogeneity in autism spectrum disorders: Visual scanning of dynamic social scenes in school-aged children. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(3), 238–248.

Robins, D., Fein, D., Barton, M., & Green, J. (2001). The Modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*(2), 131–144. doi:10.1023/a:1010738829569.

Rutherford, M. D., Walsh, J. A., & Lee, V. (2015). Brief report: Infants developing with ASD show a unique developmental pattern of facial feature scanning. *Journal of Autism and Developmental Disorders, 45*(8), 2618–2623 (2015). doi:10.1007/s10803-015-2396-7.

Rutter, M., Bailey, A., & Lord, C. (2003). *SCQ: Social Communication Questionnaire*. Los Angeles: Western Psychological Services.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology, 18*(9), 668–671. doi:10.1016/j.cub.2008.03.059.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-Learning mechanisms. *Cognitive Science, 35*(3), 480–498. doi:10.1111/j.1551-6709.2010.01158.x.

Smith, L. B. (2000). Learning how to learn words: An associative crane. *Becoming a Word Learner: A Debate on Lexical Acquisition*, 51–80.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*(3), 1558–1568. doi:10.1016/j.cognition.2007.06.010.

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience Research, 6*(3), 219–224.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. *Handbook of Autism and Pervasive Developmental Disorders, 1*, 335–364.

Tenenbaum, E., Amso, D., Abar, B., & Sheinkopf, S. J. (2014a). Attention and word learning in autistic, language delayed and typically developing children. *Frontiers in Psychology, 5*, 490.

Tenenbaum, E., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2013). Increased Focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study. *Infancy, 18*(4), 534–553. doi:10.1111/j.1532-7078.2012.00135.x.

Tenenbaum, E., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2014b). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, 1–18.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, 66*(1), 126–156.

von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders, 3*(2), 556–569. doi:10.1016/j.rasd.2008.12.003.

Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence third edition (WPPSI-III)*. San Antonio: Pearson.

Wetherby, A. M., Allen, L., Cleary, J., Kublin, K., & Goldstein, H. (2002). Validity and reliability of the communication and symbolic behavior scales developmental profile With very young children. *Journal of Speech, Language & Hearing Research, 45*(6), 1202.

Young, G., Merin, S., Rogers, N., S, J., &amp, 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.

Yurovsky, D., & Frank, M. C. (2015). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*. doi:10.1111/desc.12349.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool language scales fifth edition (PLS-5)*. Bloomington, MN: Pearson.