



Published in final edited form as:

J Child Lang. 2015 November ; 42(6): 1173–1190. doi:10.1017/S0305000914000725.

Attention to the mouth and gaze following in infancy predict language development

ELENA J. TENENBAUM,

Brown Center for the Study of Children at Risk at Women and Infants Hospital Providence, RI

DAVID M. SOBEL,

Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI

STEPHEN J. SHEINKOPF,

Brown Center for the Study of Children at Risk at Women and Infants Hospital Providence, RI and Department of Psychiatry, Brown University, Providence, RI

BERTRAM F. MALLE,

Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI

JAMES L. MORGAN

Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI

Abstract

We investigated longitudinal relations among gaze following and face scanning in infancy and later language development. At 12 months, infants watched videos of a woman describing an object while their passive viewing was measured with an eye-tracker. We examined the relation between infants' face scanning behavior and their tendency to follow the speaker's attentional shift to the object she was describing. We also collected language outcome measures on the same infants at 18 and 24 months. Attention to the mouth and gaze following at 12 months both predicted later productive vocabulary. The results are discussed in terms of SOCIAL ENGAGEMENT, which may account for both attentional distribution and language onset. We argue that an infant's inherent interest in engaging with others (in addition to creating more opportunities for communication) leads infants to attend to the most relevant information in a social scene and that this information facilitates language learning.

INTRODUCTION

Language is a multimodal form of communication that develops primarily in the context of face-to-face interactions between infants and their caregivers. Although speech is the primary channel through which the linguistic signal is transmitted, visible social and contextual cues provide information critical for interpreting speech (e.g., Csibra & Gergeley, 2006). There is now substantial evidence that selective attention to informative areas in the social scene (e.g., attention to the face or objects) is positively correlated with language

development (Brooks & Meltzoff, 2005, 2008; Carpenter, Nagell & Tomasello, 1998; Morales et al., 2000b; Morales, Mundy & Rojas, 1998; Mundy & Gomes, 1998; Mundy, Kasari, Sigman & Ruskin, 1995; Young, Merin, Rogers & Ozonoff, 2009). An open question is what aspects of selective attention contribute to language development, and how they do so. In this paper, we explore the relative contributions of two forms of attention that are likely to play an important role in word learning: gaze following and face scanning. We consider the contribution of each for language development as well as the relation between the two behaviors. It is possible that gaze following, an indicator of social cognition and face scanning, an indicator of active searching for linguistically relevant information, represent different functions. If so, they might be independently related to language outcomes. Alternatively, they could both be indicators of social engagement and thus the relation between them and language may not be specific but rather an epiphenomenon of social motivation (Locke, 1996).

Gaze following, defined as the ability to follow the attentional focus of another person, emerges during the second half of the first year of life (Butterworth & Jarrett, 1991; Carpenter et al., 1998; Corkum & Moore, 1995, 1998; D'Entremont, 2000; D'Entremont, Hains & Muir, 1997; Gredebäck, Fikke & Melinder, 2010; Perra & Gattis, 2010; Scaife & Bruner, 1975). Various researchers have shown that gaze following reliably predicts aspects of language development (Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998; Morales et al., 2000b; Morales et al., 1998; Mundy & Gomes, 1998; Mundy et al., 1995). Typically, correlations between gaze following and language development have been explained as a function of the importance of resolving a speaker's intent to label objects or events (Baldwin, 1995; Bloom, 2002; Bruner, 1983; Tomasello & Farrar, 1986). Infants must coordinate their attention with an interlocutor's gaze to recognize the objects to which he/she is referring. For example, to understand that *ball* refers to the round object to which the speaker is attending, the infant must follow the speaker's gaze to that object. Infants who do not reliably orient to objects in a speaker's direction of gaze may fail to develop vocabulary at the same rate as children who exploit these visual cues more successfully.

Face scanning during communication can also provide infants with important information for language learning. One study demonstrated that infants who attend to their mother's mouths during communication (via live video feed) at 6 months have larger productive vocabulary sizes in toddlerhood (Young et al., 2009). In contrast to the studies cited above which demonstrated that attention to a speaker's eyes and direction of gaze predicted later language outcomes, Young and colleagues showed that attention to the mouth may also play a role in successful language learning. Why should attention to the mouth be important for infants' language acquisition? Information in and around the mouth can foster the infant's associations between mouth shape (e.g., a horizontal vs. vertically opened mouth or labial vs. oral closure) and speech sounds (e.g., [i] vs. [a] or [b] vs. [d]) while also providing consistent visual cues for the ephemeral speech stream. Consistent with the notion that attention to the mouth may be relevant for language learning, Lewkowicz and Hansen-Tift (2012) found that attention to a speaker's mouth increased among 12-month-olds hearing a foreign language relative to those hearing their native language.

Although gaze following and attention to the mouth during face scanning have both been found to correlate individually with increased vocabulary size, relations between these two behaviors and language development have not previously been explored within the same group of infants. It is interesting that both of these patterns of attention have been found to correlate with later language skills, given that they are mutually exclusive behaviors; when looking at the mouth, gaze following is not happening, and when a child is following a speaker's gaze to an object, he is not looking at the speaker's mouth. The present study examined how these two behaviors relate to each other and how the two combine to predict word learning. Both attention to the mouth and gaze following may reflect contextually dependent indicators of individual differences in social attention and thus may reflect a common developmental tendency that supports language development. This view would be consistent with theories that emphasize the evolutionary benefit of social engagement and its relation to language learning (see Locke, 1996). Alternatively, face scanning and gaze following could represent different developmental functions. Face scanning could reflect an infant's sensitivity to linguistically relevant information during speaking, whereas gaze following could reflect aspects of social cognition that also relate to language, as noted above. These two alternative models would predict either (i) that the functions are related to a common underlying factor and may thus predict language, but not uniquely from one another, or (ii) that they reflect different processes and would therefore have unique relations to language outcomes.

In addition to providing useful information for language acquisition, gaze following and focus on an interlocutor's mouth while hearing speech are both indications that the infant is scanning the scene for communicatively relevant information. We hypothesize that infants who actively scan a scene for communicative information are more likely to succeed at language development. On this hypothesis, we would expect these behaviors to be related both to language success and to each other.

To test this, we examined gaze patterns of 12-month-olds who took part in a longitudinal eye-tracking study of face scanning and gaze following (Tenenbaum, Shah, Sobel, Malle & Morgan, 2013) in relation to parental surveys of vocabulary size at 18 and 24 months. The longitudinal study explored the effects of gaze direction and mouth movement on infants' distribution of attention. To that end, four trial types were counterbalanced within subjects for presence or absence of information in the eyes (the speaker was either gazing into the camera or looking at an object) and mouth (the speaker was either speaking or smiling). The four trial types were thus: (i) speaking and looking at object; (ii) speaking and looking into camera; (iii) smiling and looking at object; and (iv) smiling and looking at camera. These stimuli were further counterbalanced across subjects such that the speaker looked at the object either by shifting both her head and gaze toward the object (head-turn stimuli) or by shifting her gaze alone while keeping her head facing the camera (gaze-shift stimuli). This between-subjects manipulation allowed us to explore patterns of attention and gaze following without the added cues of head position (as well as chin and nose position). Because the current study is focused on how face scanning interacts with gaze following behaviors, and how these behaviors relate to language outcome measures, we examined only the subset of trials from the original study in which the speaker spoke and looked at the object (from both the head-turn and gaze-shift stimuli). The three remaining trial types were

useful for the longitudinal study, but were uninformative for questions of how attention to a speaker's mouth and gaze following combine as cues for language learning, and were thus omitted from this analysis.

Given the results from previous studies, we expected gaze following (Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998; Morales et al., 2000b; Mundy & Gomes, 1998; Mundy et al., 1995) and attention to the mouth (Young et al., 2009) to predict later vocabulary size. We were also interested in exploring the extent to which attention to the mouth and gaze following behaviors might both be driven by the infant's interest in scanning the scene for available information. To this end, we considered the relation between these factors and explored whether combining these measures acts as a significantly more powerful predictor of vocabulary size than either factor alone, or whether they independently predict later language abilities.

METHODS

Participants

Sixty-one infants completed the eye-tracking experiment at 12 months (33 male, 28 female, $M = 396$ days, $SD = 12$ days). Infants were recruited from public birth records. All infants were born full-term to monolingual English-speaking families and had no known developmental or hearing disabilities. Additional infants were tested but excluded from analysis for the following reasons: could not obtain eye tracker calibration ($n = 6$), fewer than half the trials successfully tracked ($n = 8$), or equipment failure ($n = 3$). Of the 61 infants tested, parents of 56 completed surveys at 18 months (30 male, 26 female, $M = 548$ days, $SD = 13$ days) and parents of 42 completed surveys at 24 months (24 male, 18 female, $M = 741$ days, $SD = 14$ days).

Eye-tracking apparatus

Infants were seated on a caregiver's lap in a sound-treated testing room, approximately 80 centimeters from the computer screen on which the video stimuli were displayed. The display at this distance subtended approximately $30^\circ \times 20^\circ$ visual angle. Below the computer screen was an Applied Science Laboratories (ASL) Pan-Tilt 5000 eye-tracker (sampling rate of 60 hertz), which was connected to a control unit running updated 6000-level software in the adjacent room. The control unit was connected to a PC in the testing booth which ran ASL software for eye tracking. Three video capture boards allowed the experimenter to view (i) the wide angle view of infant and caregiver, (ii) the infant's eye and (iii) a simultaneous view of the screen infants were watching overlain with crosshairs indicating the running average of the infant's current fixation point and the three previous fixation points. Videos were captured at a rate of 30 frames per second.

Stimuli

Our previous work (Tenenbaum et al., 2013) analyzed how infants differed in their patterns of attention to four types of trial, but did not consider any later outcome measure. Here, we were interested in the relation between face scanning and gaze following as predictors of later vocabulary size, so we restricted analysis to trials in which the speaker spoke about the

object and shifted her attention to that object. The description below is specific to the trials used in the current analysis. For descriptions of all four trial types presented in the longitudinal study, see Tenenbaum et al. (2013).

Infants saw video trials in which a speaker communicated information about one of two objects in front of her. On these trials, the speaker spoke about one of the two objects and then shifted her attention to the object being described. For approximately half of the infants, that attentional shift was in the form of a head turn towards the object; for the other half of the infants, the shift was a gaze shift to the object (i.e. the speaker's head remained focused on the camera). Infants observed four such trials, each with a different set of objects.

In addition to the manner of attentional shift, there was one small difference between the gaze-shift and head-turn stimuli. In the gaze-shift stimulus set, in order for the speaker to shift only her gaze to the object without turning her head, she needed to be positioned closer to the objects. This resulted in a zooming in on her face for the gaze-shift stimuli such that the size of the regions of interest on the screen differed slightly between the two stimulus sets. The woman's face on the video screen covered approximately $7^\circ \times 10^\circ$ for the head-turn stimuli and $10^\circ \times 14^\circ$ for the gaze-shift stimuli. The objects covered an average visual angle of $3.5^\circ \times 3.5^\circ$ for the head-turn stimuli and $5^\circ \times 5^\circ$ for the gaze-shift stimuli (see Figure 1).

At the start of each 8-second trial, the speaker smiled into the camera to demonstrate engagement with the infant (following Senju & Csibra, 2008). She then described one of the objects in front of her (hereafter, the target object). The object descriptions began with a phrase such as "Would you like to hear about a modi?" and were followed by a descriptive sentence that was not intended to identify the target object: "It's one little piece of a bigger garden we can make." The lack of uniquely descriptive information meant that speaker gaze was the only cue to target location. The onset of the target word occurred 2 seconds from the start of the trial. At that point, the speaker shifted her gaze from the camera toward one of the objects. She maintained fixation on that object for the remainder of the trial (see Figure 2). Target object and position were counterbalanced across infants and trials. Presentation order was randomized for each infant.

Procedure

A testing session began with a two-point (upper left and lower right) calibration sequence. Calibration was followed by a video of a moving ball. Experimenters determined on-line whether the infant was successfully calibrated, based on the apparent tracking of the motion of the ball (within an estimated $1\text{--}2^\circ$ visual angle).

Subsequently, the trials of interest started. At the onset of each trial, an animated figure appeared in one of five positions around the screen (thus randomizing the initial fixation point for each trial). This fixation also allowed the experimenter to confirm calibration before each trial. If calibration was deemed inaccurate (greater than $1\text{--}2^\circ$ visual angle), the infant could be recalibrated at any point. Once fixation was verified, the experimenter initiated the trial. The video of the infant's view overlain with crosshairs indicating fixation was then retained for off-line coding of the infant's focus across the trial.

Coding

For the eye-tracking procedure, we used custom-designed software to define four dynamic regions of interest (ROIs) (eyes, mouth, target, and distracter) for each stimulus video (see Figure 3). The remaining areas of the screen were designated as ‘away’ or, if crosshairs were not successfully recorded, ‘no track’. Care was taken to equate the eyes and mouth regions as the speaker shifted her position during trials. The ROIs surrounding the mouth and eyes each accounted for approximately 2% of the scene (4% for the gaze-shift stimuli). The objects each covered approximately 3% of the scene (5% for gaze-shift stimuli).

Crosshairs were extracted frame-by-frame at a rate of 30 frames per second (fps). A series of image filters was used to identify the location of the horizontal and vertical lines on the screen. For the 640×480 pixel scene, the estimated error for the X position was less than 3 pixels, and for the Y position less than 1 pixel. Because the sampling rate of the ASL camera (60 Hz) did not match that of the video capture board (30 fps), some frames had what appeared to be two fixation points. In these cases, we calculated the mean of the horizontal and vertical coordinates. Dependent measures were calculated based on the raw cumulative tracked time spent fixating each region (because these were calculated as proportions of looking time, no post-processing or fixation analysis was completed (Tenenbaum et al., 2013).

A subset of approximately 10% of all collected data (24 of the 225 test sessions that included the longitudinal dataset from which these were drawn) was hand coded, frame-by-frame by a naive rater. Inter-rater reliability between the automated scoring and that completed by hand was 90.5%, Cohen’s kappa = .90.

Surveys

Parents completed the MacArthur-Bates Short Form Vocabulary Checklist: Level I at 18 months and the MacArthur-Bates Communicative Development Inventory: Toddlers Form at 24 months (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994). The Short Form provides measures of receptive and productive vocabulary while the Toddlers Form provides measures of productive vocabulary only. See ‘Participants’ above for the response rates.

RESULTS

Preprocessing

Each infant contributed at least two usable trials to the analysis (with usable defined as having at least 50% of trial time successfully tracked, $M = 3.67$ of the 4 trials successfully tracked, $SD = .63$).

Patterns of eye gaze were comparable across the head-turn and gaze-shift stimulus sets. The proportion of tracked time spent fixating the speaker’s eyes, and the target and distracter objects were not significantly different across stimulus sets (all $F(1,60)$ -values < 0.68 , all p -values $> .40$). Moreover, neither the gaze-following measure nor the mouth-to-eyes index (both described below) differed between the two stimulus sets (both $F(1,60)$ -values < 1.74 , both p -values $> .18$). The only observed difference was on the proportion of tracked time

spent fixating the mouth, which was marginally greater in the gaze-shift stimulus set than the head-turn stimulus set ($F(1,60) = 3.92, p = .054$). As a result, we collapsed across stimulus sets for the remaining analyses.

Gaze following

To assess gaze following – i.e. whether infants looked in the direction of the object that the speaker indicated – we used the proportion of tracked time spent fixating the target object minus proportion of tracked time spent fixating the distracter object for the 2 seconds following the speaker's attentional shift. This window of time was chosen to allow sufficient time for 12-month-olds to orient in response to the speaker's attentional shift (based on results from Gredebäck et al., 2010) and was similar to a measure used by Senju and Csibra (2008). During this 2-second window, infants focused their attention on the target 3.64% of the time ($SD = 4.40\%$), and on the distracter 2.92% of the time ($SD = 3.87\%$). Though this was not significantly different ($t(60) = 1.36, p = .18$), there was considerable variation between infants.

Face scanning

To assess face scanning, we examined the percentage of tracked time infants fixated the face, dividing it into three roughly equivalently sized regions (mouth, nose, and eyes). Infants fixated these facial features for an average of 61.86% of tracked trial time ($SD = 15.95\%$). Attention to the objects accounted for 13.97% of the infants' tracked fixations ($SD = 10.78\%$), and the infants spent approximately 22.15% of the tracked trial time looking elsewhere on the screen ($SD = 13.40\%$). Within the face, infants focused primarily on the mouth ($M = 40.35\%, SD = 19.49\%$) and eyes ($M = 11.97\%, SD = 13.17\%$). For comparison with previous findings indicating a relation between attention to the mouth and vocabulary size (Young et al., 2009), we calculated a mouth to eyes index (proportion of attention to the mouth/proportion of attention to the mouth + eyes) as our measure of attention to the mouth (though Young *et al.*, used an eyes to mouth index, we chose to reverse this because our focus was on the effects of attention to the mouth). This measure of attention to the mouth was significantly correlated with the gaze following measure described above ($r(59) = 0.26, P < .05$), suggesting that infants who look more at the speaker's moving mouth when scanning the face are also more likely to follow the speaker's gaze to the target object.

Language outcome measures

We calculated percentile scores for receptive ($M = 68.86, SD = 34.53$) and productive ($M = 42.88, SD = 28.54$) vocabulary at 18 months and productive vocabulary at 24 months ($M = 45.67, SD = 27.67$). These measures were all significantly correlated with each other: receptive and productive vocabulary at 18 months ($r(56) = 0.70, p < .01$); receptive vocabulary at 18 months and productive vocabulary at 24 months ($r(41) = 0.71, p < .01$); and productive vocabulary at 18 and 24 months ($r(41) = 0.79, p < .01$).

Eye tracking in relation to language outcome measures

To explore the relations among face scanning, gaze following, and language outcome measures, we ran three hierarchical regressions with receptive and productive vocabulary at

18 months and productive vocabulary at 24 months as outcome variables. For each regression, we first constructed a model with overall tracked time (a proxy for general attention to the stimuli) as the only predictor measure. We then considered models that contained attention to the mouth and gaze following, entered stepwise. Consistent with results from Young et al. (2009), attention to the mouth and gaze following were not significant predictors of receptive vocabulary scores. The regressions for productive vocabulary at both 18 and 24 months were significant. Results of these regressions are shown in Tables 1 to 3. Even when the effects of overall attention were factored out, both gaze following and attention to the mouth at 12 months accounted for unique and significant portions of the variance in productive vocabulary size at 18 and 24 months.

DISCUSSION

The goal of this research was to explore the longitudinal relations between face scanning and gaze following in infancy and subsequent language development in toddlerhood. Based on previous findings, we had expected attention to the mouth and gaze following individually to predict vocabulary size. Indeed, we found that both attention to the mouth and gaze-following behaviors at 12 months correlated with productive vocabulary outcomes at 18 and 24 months.

The finding that attention to the mouth predicts later vocabulary size replicates a similar finding with typically developing infants and infants at risk for autism at 6 months of age (Young et al., 2009). Young and colleagues had used attention to faces during a live video feed of the still-face paradigm with 6-month-olds and their caregivers. Here we extended these results to a larger sample of typically developing infants, using prerecorded videos of a stranger talking and referencing objects. Importantly, we also showed that, even as late as 12 months, attention to the mouth is an adaptive pattern for language learning. Though Lewkowicz and Hansen-Tift (2012) found that infants hearing their native language begin shifting their attention from the mouth towards the eyes at 12 months, the presence of novel words in the current stimuli likely increased the relevance of the mouth just as hearing a non-native language had among infants in that study.

Our finding that gaze following predicts later vocabulary size extends previous findings in live interactions (Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998; Morales et al., 2000b; Mundy & Gomes, 1998; Mundy et al., 1995) to an eye-tracking paradigm. The presentation of these stimuli in prerecorded video allowed us to present the gaze-following stimuli consistently and without the influence of the child's own social cues on the experimenter.

The lack of correlation between gaze following and receptive language abilities may be related to methodological differences between the current study and previous work demonstrating such a relation (Brooks & Meltzoff, 2005; Carpenter et al., 1998; Morales et al., 2000a; Mundy & Gomes, 1998). These other studies explored receptive vocabulary much earlier (at 12 or 14 months) (Carpenter et al., 1998; Morales et al., 2000a), focused on initiation of joint attention behaviors (i.e. vocalizations combined with looking at the target) (Brooks & Meltzoff, 2005), or used behavioral measures rather than parental report to

estimate receptive language skills (Morales et al., 2000b; Mundy & Gomes, 1998). As anecdotally evidenced by the number of parents asking their children questions such as “Shoe, do you know shoe?” while attempting to complete the receptive portion of the M-CDI, it is possible that our lack of results with respect to receptive language ability reflects the difficulty parents have in assessing receptive vocabulary size. While the first few words a child demonstrates recognition of might be easier to track (those that might be clear by 12 or 14 months), it is reasonable to expect that parents would have a more difficult time tracking receptive language skills by 18 months. In support of the difficulties establishing validity of receptive vocabulary size, the M-CDI has been validated for its productive measures by more than eighteen independent studies (resulting in an average Pearson product moment correlation between the CDI and an independent measure of $.68$), whereas only six have confirmed validity for the receptive measures (average $r = .64$) (Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007).

We had suggested that as manifestations of active scanning for communicative information, attention to the mouth and gaze following would be related to each other and would be stronger predictors of vocabulary size jointly than individually. Because we examined gaze-following and face-scanning behaviors in the same group of infants, we were able to identify a significant relation between these behaviors. That is, at 12 months, attention to the mouth and gaze-following behavior in our paradigm were significantly related to one another. It is particularly interesting that we found this relation, because one might expect attention to the mouth to be inversely related to gaze following given that the patterns of attention are mutually exclusive. We argue that the relation between these behaviors lends support to the claim that these behaviors reflect a common underlying ability to scan the scene productively. In further support of the notion that these behaviors are reflective of a shared capacity that is relevant for word learning, the overall model that included attention to the mouth and gaze-following behaviors was a better predictor of infants’ productive vocabulary at both 18 and 24 months than a model that contained either individual variable alone.

An open question is what sort of mechanism drives the relation between visual attention to the social scene and language development. We argue that gaze-following and face-scanning behaviors are related and predict language development because they both reflect active scanning of a social scene for communicative information. In addition to providing information necessary for language learning (Baldwin, 1995; Brooks & Meltzoff, 2005; Young et al., 2009), these behaviors and the connections between them result from an underlying level of social engagement on the part of the infant. Indeed, previous work has shown that both gaze following (Sheinkopf, Mundy, Claussen & Willoughby, 2004) and attention to the mouth (Young et al., 2009) are predictive of later socialization scores on standardized measures of adaptive behaviors. Young *et al.*, in particular, suggest that the connection between attention to the mouth and socialization scores is mediated by expressive language skills (arguing that attention to the mouth predicts language skills which in turn predict socialization skills).

We would suggest an alternative interpretation. Instead of the child’s language capacities predicting their socialization, we believe that infants’ SOCIAL ENGAGEMENT drives them to attend to and integrate the relevant information in the social scene. The extent to which

infants can integrate information from their attention to the mouth with information available from such social interactions contributes to their productive language skills. This social engagement factor also makes infants more motivated to communicate with others through language, thus providing greater opportunity for practice and success at language development (Falck-Ytter, Fernell, Gillberg & Von Hofsten, 2010; Mundy et al., 1995; Sheinkopf et al., 2004; Trevarthen & Aitken, 2001; Vaughan Van Hecke et al., 2007).

The outcome measures in the current experiment were restricted to vocabulary size, but we argue that active scanning of the social scene for linguistic information could also help children gain an understanding of syntactic structures. Just as gaze following can help a child to assign *ball* to the round object being attended, attention to who is throwing that ball can help them understand the phrase *Billy threw to Sally*. In the case of syntactic structures, we would not necessarily expect attention to the mouth or eyes specifically to predict success, but the underlying social engagement that we posit contributes to active scanning of the scene would likely remain relevant. Further, the patterns of attention we have demonstrated to be predictive of vocabulary size are likely to contribute indirectly to greater skills in other linguistic domains. Indeed, evidence from Marchman and Fernald (2008) suggests that vocabulary size in toddlerhood is predictive of more general expressive language abilities as late as eight years. These abilities included expressive vocabulary, formulating sentences, recalling sentences, and word structure. Vocabulary size has also been linked to speed of sentence interpretation among children and adults, suggesting benefits for receptive language skills as well (Borovsky, Elman & Fernald, 2012).

Some have argued that the relation between attention to the mouth and later language skills might be mediated by infants' exposure to infant directed speech (IDS). Young et al. (2009) suggested that the more IDS a mother uses, the more likely she is to exaggerate her mouth movements, thereby drawing attention to the mouth. As a result, children might have superior language skills because of the IDS rather than the attention to the mouth per se. Our use of a single, unfamiliar speaker rather than the infant's own caregiver allowed us to demonstrate that infants who attend to the mouth of a stranger have greater success at language learning. While this suggests that the effect of attention to the mouth is not contingent upon the mother's tendency to use IDS, our data do not permit us to rule out the possibility that infants of mothers who use exaggerated mouth movements are primed to focus more on the mouth and also have greater success at language learning,

Another alternative interpretation of our results is that attention to the mouth predicts language abilities because lower-level features of the moving mouth, rather than linguistic content, attract the attention of infants who are more advanced and thus able to direct their attention to the more salient areas of the scene (Colombo, 2001; Richards, Reynolds & Courage, 2010). By this logic, infants who attend to the movement of the mouth are faster to acquire language because they are more cognitively advanced in general. The present data, however, are inconsistent with this account for two reasons. First, the effects demonstrated here for attention to the mouth on language development held when overall attention to the stimuli (also a sign of cognitive maturity) was factored out. Second, while it is quite likely that movement of the mouth draws the attention of infants early on, movement alone does not seem to be the full story. Lewkowicz and Hansen-Tift (2012) found that attention to a

speaker's mouth increases when infants hear a foreign language relative to when they hear their native language. This suggests that infants are sensitive to the linguistic relevance of information in the mouth, and not simply the movement. If movement alone were the driving force in infants' attention to the mouth, the linguistic content of the speech would not matter.

To conclude, we suggest that language development is facilitated by an infant's underlying social engagement that pushes them to attend to both linguistically relevant information in the mouth and socially relevant information about the referent (objects or actions) as they follow their interlocutors' social cues. This conclusion suggests possible interventions for language acquisition. If children are gaining relevant information for language learning by seeking out social information and integrating it with information available in the face, then an open question is whether language acquisition could be facilitated by drawing a child's attention to those informative areas. This may be particularly relevant for children with developmental disorders such as autism who might not have the internal social drive to direct their attention to the areas of the visual scene that are most informative for language learning (Chawarska, Macari & Shic, 2013; Chawarska & Shic, 2009; Shic, Bradshaw, Klin, Scassellati & Chawarska, 2011; Shic, Macari & Chawarska, 2013), but who might benefit from the information if directed to those areas (Tenenbaum, Amso, Abar & Sheinkopf, 2014).

Acknowledgments

This work was supported by NIH grant HD32005 to JLM and a graduate research award from the Brain Sciences Program at Brown University to EJT and DMS, and an Autism Speaks postdoctoral fellowship (grant 7897) to EJT and SJS. Thanks go to Lori Rolfe for substantial effort in recruiting and testing, Halie Rando for her face, and the participant families for all of their efforts. We also thank Scott P. Johnson at UCLA for the attention attracting video used to calibrate the infants in this study.

REFERENCES

- Baldwin DA (1995). Understanding the link between joint attention and language. In Moore C & Dunham PJ (eds), *Joint attention: its origins and role in development*, 131–58. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bloom P (2002). Mindreading, communication and the learning of names for things. *Mind & Language* 17, 37–54.
- Borovsky A, Elman JL & Fernald A (2012). Knowing a lot for one's age: vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology* 112, 417–36. [PubMed: 22632758]
- Brooks R & Meltzoff AN (2005). The development of gaze following and its relation to language. *Developmental Science* 8, 535–43. [PubMed: 16246245]
- Brooks R & Meltzoff AN (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *Journal of Child Language* 35, 207–20. [PubMed: 18300435]
- Bruner JS (1983). *Child's talk: learning to use language*. Oxford: Oxford University Press.
- Butterworth G & Jarrett N (1991). What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology* 9, 55–72.
- 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.

- 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, 195–203. [PubMed: 23313640]
- 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, 1663–72. [PubMed: 19590943]
- Colombo J (2001). The development of visual attention in infancy. *Annual Review of Psychology* 52, 337–67.
- Corkum V & Moore C (1995). Development of joint visual attention in infants. In Moore C & Dunham PJ (eds), *Joint attention: its origins and role in development*, 61–84. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Corkum V & Moore C (1998). Origins of joint visual attention in infants. *Developmental Psychology* 34, 28–38. [PubMed: 9471002]
- Csibra G & Gergely G (2006). Social learning and social cognition: the case for pedagogy. *Processes of change in brain and cognitive development. Attention and Performance XXI*(21), 249–274.
- D’Entremont B (2000). A perceptual-attentional explanation of gaze following in 3 and 6 month olds. *Developmental Science* 3, 302–11.
- D’Entremont B, Hains SMJ & Muir DW (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development* 20, 569–72.
- Falck-Ytter T, Fernell E, Gillberg C & Von Hofsten C (2010). Face scanning distinguishes social from communication impairments in autism. *Developmental Science* 13, 864–75. [PubMed: 20977557]
- Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ & Pethick SJ (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* (Vol. 59).
- Fenson L, Marchman VA, Thal DJ, Dale PS, Reznick JS & Bates E (2007). *MacArthur-Bates Communicative Development Inventories – User’s Guide and Technical Manual*, 2nd ed. Baltimore: Brookes Publishing.
- Gredebäck G, Fikke L & Melinder A (2010). The development of joint visual attention: a longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science* 13, 839–48. [PubMed: 20977555]
- Lewkowicz DJ & Hansen-Tift AM (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.
- Locke JL (1996). Why do infants begin to talk? Language as an unintended consequence. *Journal of Child Language* 23, 251–68. [PubMed: 8936686]
- Marchman VA & Fernald A (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science* 11, F9–F16. [PubMed: 18466367]
- Morales M, Mundy P, Delgado CEF, Yale M, Messinger D, Neal R & Schwartz HK (2000a). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology* 21, 283–98.
- Morales M, Mundy P, Delgado CEF, Yale M, Neal R & Schwartz HK (2000b). Gaze following, temperament, and language development in 6-month-olds: a replication and extension. *Infant Behavior and Development* 23, 231–6.
- Morales M, Mundy P & Rojas J (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development* 21, 373–7.
- Mundy P & Gomes A (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development* 21, 469–82.
- Mundy P, Kasari C, Sigman M & Ruskin E (1995). Nonverbal communication and early language acquisition in children with Down Syndrome and in normally developing children. *Journal of Speech and Hearing Research* 38, 157–67. [PubMed: 7537345]
- Perra O & Gattis M (2010). The control of social attention from 1 to 4 months. *British Journal of Developmental Psychology* 28, 891–908.
- Richards JE, Reynolds GD & Courage ML (2010). The neural bases of infant attention. *Current Directions in Psychological Science* 19, 41–6. [PubMed: 20445766]

- Scaife M & Bruner JS (1975). The capacity for joint visual attention in the infant. *Nature* 253, 265–6. [PubMed: 1113842]
- Senju A & Csibra G (2008). Gaze following in human infants depends on communicative signals. *Current Biology* 18, 668–71. [PubMed: 18439827]
- Sheinkopf SJ, Mundy P, Claussen AH & Willoughby J (2004). Infant joint attention skill and preschool behavioral outcomes in at-risk children. *Developmental Psychopathology* 16, 273–91.
- Shic F, Bradshaw J, Klin A, Scassellati B & Chawarska K (2011). Limited activity monitoring in toddlers with autism spectrum disorder. *Brain Research* 1380, 246–54. [PubMed: 21129365]
- Shic F, Macari S & Chawarska K (2013). Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. *Biological Psychiatry*.
- Tenenbaum E, Amso D, Abar B & Sheinkopf SJ (2014). Attention and word learning in autistic, language delayed and typically developing children. *Frontiers in Psychology* 5, 490. [PubMed: 24904503]
- Tenenbaum E, Shah RJ, Sobel DM, Malle BF & Morgan JL (2013). Increased focus on the mouth among infants in the first year of life: a longitudinal eye-tracking study. *Infancy* 18, 534–53. [PubMed: 23869196]
- Tomasello M & Farrar MJ (1986). Joint attention and early language. *Child Development* 57, 1454–1463. [PubMed: 3802971]
- Trevarthen C & Aitken KJ (2001). Infant intersubjectivity: research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 42, 3–48.
- Vaughan Van Hecke A, Mundy PC, Acra CF, Block JJ, Delgado CEF, Parlade MV, & Pomares YB (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Development* 78, 53–69. [PubMed: 17328693]
- Young G,S , Merin N, Rogers SJ & 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, 798–814. [PubMed: 19702771]

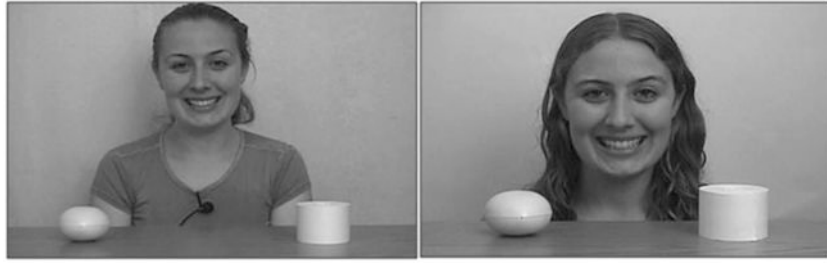


Fig. 1.
Example stimuli from the head-turn (left) and gaze-shift (right) conditions.

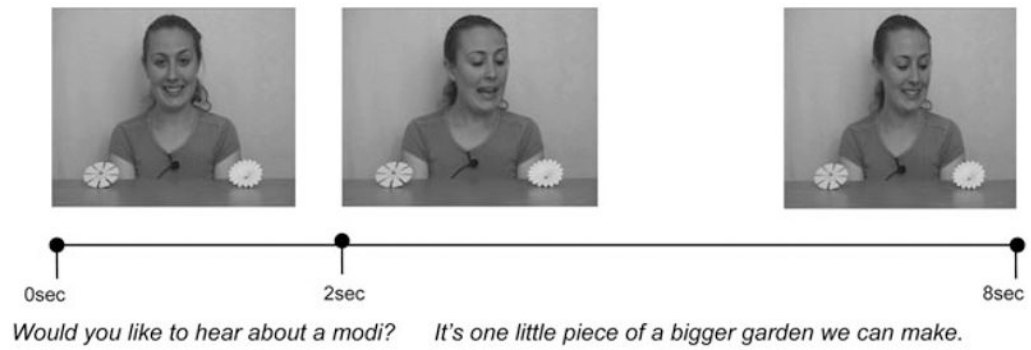


Fig. 2.

An example of a trial. All trials began with the speaker smiling at the camera. Two seconds into the trial, the speaker would label the target object and shift her attention to that object. Her attention would remain fixed on the object for the length of the 8-second trail.



Fig. 3. Frame Capture from Coding Program. Shows the regions of interest (eyes, nose, mouth, left object, and right object) as defined for this particular frame. The solid black crosshairs represent the infant's fixation on this frame and would have been coded as 'mouth'.

TABLE 1.

Stepwise regression predicting receptive vocabulary at 18 months

Variable	β	R ²	Adj R ²	R ²
Model 1 ^a		0.01	-0.02	0.01
Attention overall	-0.05			
Model 2 ^b		0.07	0.03	0.05
Attention overall	-0.07			
Attention to mouth	0.26			
Model 3 ^c		0.11	0.06	0.05
Attention overall	-0.12			
Attention to mouth	0.21			
Gaze following	0.23			

NOTES:

^aOverall model: $F(1,55) = 0.14$, n.s.^bOverall model: $F(2,54) = 1.96$, n.s.; change from Model 1 to 2: $F(1,54) = 3.77$, $p = .06$.^cOverall model: $F(3,52) = 2.26$, n.s.; change from Model 2 to 3: $F(1,52) = 2.74$, n.s.

TABLE 2.

Stepwise regression predicting productive vocabulary at 18 months

Variable	β	R ²	Adj R ²	R ²
Model 1 ^a		0.00	-0.02	0.00
Attention overall	-0.01			
Model 2 ^b		0.12	0.12	0.10 ^{**}
Attention overall	-0.03			
Attention to mouth	0.35 ^{**}			
Model 3 ^c		0.20	0.15	0.08 [*]
Attention overall	-0.09			
Attention to mouth	0.28 [*]			
Gaze following	0.29 [*]			

NOTES:

* = $p < .05$ ** = $p < .01$.^aOverall model: $F(1,55) = 0.01$, n.s.^bOverall model: $F(2,54) = 3.62$, $p < .05$; change from Model 1 to 2: $F(1,54) = 7.24$, $p < .01$.^cOverall model: $F(3,53) = 4.31$, $p < .01$; change from Model 2 to 3: $F(1,53) = 5.14$, $p < .05$.

TABLE 3

Stepwise regression predicting productive vocabulary at 24 months

Variable	β	R ²	Adj R ²	R ²
Model 1 ^a		0.04	0.01	0.04
Attention overall	-0.01			
Model 2 ^b		0.21	0.17	0.17**
Attention overall	-0.03			
Attention to mouth	0.35**			
Model 3 ^c		0.28	0.23	0.08*
Attention overall	-0.09			
Attention to mouth	0.28*			
Gaze following	0.29*			

NOTES:

* = $p < .05$ ** = $p < .01$.^aOverall model: $F(1,41) = 1.61$, n.s.^bOverall model: $F(2,40) = 5.15$, $p < .05$; change from Model 1 to 2: $F(1,40) = 8.40$, $p < .01$.^cOverall model: $F(3,39) = 5.14$, $p < .01$; change from Model 2 to 3: $F(1,39) = 4.28$, $p < .05$.