Published in final edited form as:

J Exp Child Psychol. 2018 September ; 173: 268–283. doi:10.1016/j.jecp.2018.04.011.

How do infants and adults process communicative events in real time?

Amy Yamashiro and Athena Vouloumanos

Department of Psychology, New York University

Abstract

Speech allows humans to communicate and to navigate the social world. By 12 months, infants recognize that speech elicits appropriate responses from others. However, it is unclear how infants process dynamic communicative scenes and how their processing abilities compare to adults. Do infants, like adults, process communicative events while the event is occurring, or only after being presented with the outcome? We examined 12-month-olds' and adults' eye movements as they watched a Communicator grasp one (target) of two objects. During the test event, she could no longer reach the objects, so she spoke or coughed to a Listener, who selected either object. Infants' and adults' patterns of looking to the actors and objects revealed that both groups immediately evaluated the Communicators' speech, but not her cough, as communicative, and recognized that the Listener should select the target object only when the Communicator spoke. Furthermore, infants and adults shifted their attention between the actors and the objects in very similar ways. This suggests that 12-month-olds can quickly process communicative events as they occur with adult-like accuracy. However, differences in looking reveal that 12-month-olds' process slower than adults. This early developing processing ability may allow infants to learn language and acquire knowledge from communicative interactions.

Keywords

Speech perception; communication; infant eye-tracking

Speech allows people to quickly and efficiently transfer information to others. Early in development, understanding this communicative function of speech helps infants navigate the social world and acquire knowledge in social interactions (reviewed in Vouloumanos & Waxman, 2014). What we know about infants' understanding of communication is largely based on violation-of-expectation (VOE; e.g., Cheung, Xiao, & Lai, 2012; Krehm, Onishi, & Vouloumanos, 2014; Martin, Onishi, & Vouloumanos, 2012; Song, Onishi, Baillargeon, & Fisher, 2008; Vouloumanos, Martin, & Onishi, 2014; Vouloumanos, Onishi, & Pogue, 2012) or play-based (e.g., Akhtar, Jipson, & Callanan, 2001; Grosse, Behne, Carpenter, & Tomasello, 2010; Liszkowski, Carpenter, & Tomasello, 2008; Schulze & Tomasello, 2015) methods that only provide evidence for how infants evaluate communicative events and do not assess how infants process the dynamics of communication as it unfolds in real time.

Recent eye-tracking studies have shown that infants recognize that speech is directed at others (Thorgrimsson, Fawcett, & Liszkowski, 2015). But no study of real-time processing has examined whether infants understand that speech (but not non-speech) may transfer information to a listener, eliciting an appropriate response. In this study, we ask whether infants and adults process communicative events involving speech or non-speech in real time while the event is occurring or retrospectively only after seeing the outcome, and moreover, whether infants and adults anticipate the outcome of a communicative event by making predictions about how others will respond.

Within their first year of life, infants use speech to learn about the world and the people around them. Infants treat others' speech and non-speech as functionally distinct when categorizing (Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010; Fulkerson & Waxman, 2007), labeling (Mackenzie, Graham, & Curtin, 2011), and individuating (Xu, 2002; Xu, Cote, & Baker, 2005) objects. Infants also use speech to identify potential communicative partners and learn from communicative interactions. By 5 months, infants match human speech, but not other human vocalizations like laughter, to human rather than monkey faces (Vouloumanos, Druhen, Hauser, & Huizink, 2009). By 6 months, infants use cues like speech and eye gaze as signals of communication and an opportunity for learning (Csibra & Gergely, 2009, 2011). Before their first birthday, infants use others' speech to learn who to communicate with and how to interact with others.

Infants' looking time to the outcome of an event in VOE procedures (reviewed in Baillargeon et al., 2010; see also Krehm et al., 2014; Liberman et al., 2014; Martin et al., 2012; Song et al., 2008; Vouloumanos et al., 2014, 2012) or their behavioral responses in play-based procedures (e.g., Akhtar et al., 2001; Grosse et al., 2010; Liszkowski et al., 2008; Schulze & Tomasello, 2015) suggest that infants understand how speech and gestures communicate information to others. VOE and play-based studies have shown that infants as young as 6 months, who have a limited receptive vocabulary (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012), recognize that speech, but not a non-speech cough, is communicative, allowing the speaker to provide information to a listener (Martin et al., 2012; Vouloumanos et al., 2014). By 12 months, infants also recognize that speech and social actions like clapping can inform others about observable and unobservable intentions (Cheung et al., 2012; Vouloumanos et al., 2012). At 12 months, infants will also tailor their gestures to the knowledge state of their communicative partner: infants gesture more often to an object's location when interacting with an ignorant, compared with a knowledgeable partner (Liszkowski et al., 2008). VOE and play-based studies suggest that, by 12 months, infants have some understanding of how speech and gestures communicate in social interactions and use this knowledge when interacting with others.

How long infants look at the outcome of a scene or how infants respond to an experimenter in a play-based scenario allow for broad inferences about how infants evaluate communication in social interactions. However, communication and social interactions are dynamic processes with actions and vocalizations shifting in time and space. Within this dynamic environment, adults can quickly identify, predict, and evaluate others' behaviors in real time, as the actions are unfolding (e.g., Daum & Gredebäck, 2011; Flanagan & Johansson, 2003; Langdon & Smith, 2005). However, little is known about how infants

process dynamic communicative scenes and whether they perceive and interpret others' actions in real time, as they unfold or evaluate events after they've occurred (Aslin, 2007; Gredebäck & Daum, 2015). Infants may process a social event in at least two different ways (Gredebäck & Melinder, 2010). Infants, like adults, may be evaluating the event in real time and may even make predictions about its outcome while the event is occurring (e.g., Thorgrimsson, Fawcett, & Liszkowski, 2014, 2015). Alternatively, infants may observe the entire sequence of events and only evaluate the social act retrospectively after being presented with its outcome, as in habituation studies of social events (e.g., Beier & Spelke, 2012; Hamlin, 2014). Measuring infants' eye movements with eye-tracking expands on traditional VOE and play-based procedures to reveal underlying cognitive processes (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

The study of infants' eye movements has begun to reveal what infants understand about how people interact (Augusti, Melinder, & Gredebäck, 2010), how they interpret others' behaviors, and how they make predictions about others' actions in real time, while the actions occur (Brandone, Horwitz, Aslin, & Wellman, 2014; Cannon & Woodward, 2012; Falck-Ytter, 2015; Fawcett & Gredebäck, 2013; Gredebäck & Melinder, 2010; Thorgrimsson et al., 2014, 2015). For example, infants make more gaze shifts between two actors when they engage in face-to-face compared to back-to-back conversation, suggesting 6-month-olds are sensitive to conventional patterns of social interaction (Augusti et al., 2010). Furthermore, infants as young as 10 months perform like adults in anticipating the goal of their own and others' reaches by shifting their gaze to the goal of a reach before their own or someone else's hand arrives at the goal (Rosander & von Hofsten, 2011). Infants' eye movements reveal how infants evaluate others' actions as they unfold in real time.

More recently, infants' eye movements have given insight into how infants process and respond to communicative events involving non-verbal gestures and speech in real time, while the event unfolds (Thorgrimsson et al., 2014, 2015). At 14 months, infants can predict the outcome of communicative events that consist of non-verbal gestures: infants look at a bowl in anticipation that a recipient will act on it after a gesturer points or uses a palm-up request, but not when she reaches towards the bowl (Thorgrimsson et al., 2014). Although infants interpret gestures while they are occurring and before the outcome is presented, it is less clear how infants interpret a communicative event involving speech. At 12 and 24 months, infants evaluate speech but not non-speech as being directed at a third-party listener only when two people are face-to-face: after the speaker speaks, infants quickly shift their visual attention from the speaker to the listener (Thorgrimsson et al., 2015). However, these scenarios do not assess communication: because the listener never responds to the speaker, the infants cannot evaluate whether communication has occurred or whether it was successful. Therefore, it is unclear whether 1) infants understand *how* the listener should respond to the speaker's speech (the outcome of the event), and 2) whether infants use that knowledge to anticipate the listener's response to speech in the same way they do for gestures. Given that infants within their first year seem to understand that non-verbal gestures and speech function similarly in communicative scenes (e.g., Csibra & Gergely, 2009, 2011; Gredebäck, Melinder, & Daum, 2010; Martin et al., 2012; Vouloumanos et al., 2014), 12-month-olds may interpret speech like gesture in these events: by making predictions about the outcome during the communicative event.

In the current study, we used a remote eye-tracker to examine how 12-month-old infants and adults perceive and interpret the communicative function of speech in a social interaction. We showed participants a video of pre-recorded actions in a third-party scenario (Augusti et al., 2010; Fawcett & Gredebäck, 2013; Gredebäck et al., 2010; Thorgrímsson et al., 2015) during which an actor (the Communicator), alone, repeatedly grasped one object (target) rather than another object (non-target), suggesting that the Communicator may prefer the target object (e.g., Cannon & Woodward, 2012; Woodward, 1998). In the test scene, the Communicator and a second actor (the Listener) were both present. Due to a change in the scene, however, the Communicator could no longer reach the objects whereas the Listener could. The Communicator could request the target object from the Listener by using an informative vocalization like speech (Speech condition), but should not be able to request the target object using an uninformative non-speech vocalization like coughing (Cough condition).

We tested whether infants and adults process a communicative event in real time, while the event is occurring or retrospectively only after seeing the outcome, by measuring how infants' and adults' looking time to the Listener, Communicator, target, and non-target objects changed during the Communicator's vocalization and the Listener's response. We predicted that participants would distribute their attention to the actors and objects differently depending on the Communicator's vocalization and the Listener's response. Specifically, if participants evaluated the Communicator's speech as communicating about the target object, they may recognize that the Listener would select the target object. However, if the Listener incorrectly selected the non-target object, participants may recognize that this is incongruent with the Communicator's speech, and look at the Listener relatively more than when she selected the target object. Similarly, if participants recognize that the Listener would select the target object in response to the Communicator's speech, they may look anticipatorily at the target object before the Listener reached for it. In contrast, if participants recognize that a non-speech cough may not transfer information, then they may look longer at the Communicator after she coughed than when she spoke because it is unusual to direct an uninformative vocalization towards a partner in a social scenario. Furthermore, participants may be agnostic about how the Listener will respond to an uninformative cough vocalization and may look equally at the Listener regardless of the object she chose. Finally, infants may look longer at the scene overall when the Communicator spoke and the Listener selected the non-target compared to the target object, as in prior studies (Martin et al., 2012; Vouloumanos et al., 2014).

Method

Participants

Thirty-two healthy, full-term infants ($M_{age} = 12$ months, 15 days; range = 12, 0 to 13, 1) participated, 16 infants in the Speech condition (11 females), and 16 infants in the Cough condition (6 females). This sample size is justified by a posthoc power analysis (GPOWER; Erdfelder, Faul, & Buchner, 1996) on two key results: the interaction between vocalization and outcome on infants' looking times to all AOIs in the action segment and in the still image segment of the test trial, which yielded large effect sizes ($f_{action} = .36$, $f_{still image} = .37$)

and 70% power of at alpha level of $p < .05$. Infants were recruited from maternity wards at local hospitals ([blinded for review]). Parents gave informed consent on behalf of their infants and received a certificate and small toys or t-shirts as gifts. Infant race was reported as 17 White, 9 mixed race, 2 African American, and 1 Asian (3 were unknown). Education levels of the primary caregiver were high as 12 had a graduate degree, 15 had a college degree, and 2 completed some college (3 were unknown). Data from 17 additional infants were excluded from analysis due to fussiness or crying (3), inattentiveness (1), insufficient gaze tracking by the eye tracker (11, see Inclusion Criteria below), or experimenter error (2).

Twenty-four adults ($M_{age} = 19.7$ years; range = 18 to 23; 17 females) participated, 12 adults in the Speech condition (9 females) and 12 adults in the Cough condition (8 females). Adults were undergraduate students at [blinded for review] participating for course credit for a psychology class. Adult race was reported as 8 White, 7 Asian, 3 mixed race, and 1 African American (5 were unknown). Data from 4 additional adults were excluded from analysis due to insufficient gaze tracking by the eye tracker (3, see Inclusion Criteria below) or falling asleep during the task (1). All procedures were approved by the University Committee on Activities Involving Human Subjects at [blinded for review].

Procedure

Participants watched a video of two actors interacting with two novel objects, based on a live action version of this study (Martin et al., 2012; Vouloumanos et al., 2014). Each participant participated in either the Speech or Cough condition. For each vocalization condition, each participant saw one Target outcome and one Non-target outcome. Each outcome included three trials: two familiarization trials and one test trial composed of an action segment and a still image segment (see Figure 1). During the familiarization trials, the Communicator was in the back window with the top of her face and arms visible. Two novel objects were within her reach. She looked briefly at each object (500 ms per object) and picked up the target object (1420 ms). The familiarization trial was presented two times, each lasting 4 s with an interstimulus interval of 430 ms. During the action segment of the test trial, both the Communicator and Listener were present. However, the smaller opening did not allow the Communicator's arms to pass through so she could see but no longer reach the objects. The Listener was able to reach both objects. The Communicator turned to the Listener (−750 to −0 ms), and produced a vocalization, either speech (a nonsense word; e.g., “koba”) or cough (0 to 750 ms). The Listener responded by looking at the objects (750 to 1500 ms), leaning forward (1500 to 2250 ms), reaching for an object (2250 to 3000 ms), and grasping an object (Listener's hand begins to close around the object at 3000 ms and grasps the object at 4000 ms). In the Target outcome the Listener selected the target object. In the Non-target outcome the Listener selected the non-target object. The Listener always reached for the object closest to her. Then, the Listener lifted it just below the Communicator's face (4000 to 4750 ms). After the 6 s action segment of the test trial, the final image froze for the 20 s still image segment of the test trial. To replicate previous VOE studies, timing for the still image segment of the test trial ended when the infant had looked away from the scene for two consecutive seconds (Martin et al., 2012; Vouloumanos et al., 2014). Half the participants within each vocalization saw the Target outcome first and half saw the Non-target outcome

first. See Supplemental materials for an example video of the Target and Non-target outcomes.

Stimuli

The objects, actors, and vocalizations differed between the Target and Non-target outcomes. In the Target outcome, the target object was an orange, angular, hourglass shaped object with a purple base, the non-target object was an inverted ring tower covered in pink tape. In the Non-Target outcome, the target object was a gray box with a red base and rod on top, and the non-target object was a yellow cone with a green, coiled wire on top.

Apparatus

We report eye-tracking information following the guidelines of Oakes, 2010. Infants were seated upright in a high chair approximately 60 cm from a 29 cm x 47 cm screen in a sound-attenuated room. At a viewing distance of 60 cm, the stimulus scene measured 27.8° vertical and 43.1° horizontal visual angle. Adults were seated in a chair at approximately 65 cm from the same size screen in a sound-attenuated room. Gaze was measured with the SensoMotoric Instruments (SMI) RED (sampling rate of 60 Hz) or RED-m (sampling rate of 120 Hz) infrared eye tracker system (SensoMotoric Instruments GmbH, Tetlow, Germany; www.smivision.com/). SMI IView X™ (Version 2.8, 2014) and Experiment Center™ (Version 3.4, 2014) controlled the calibration and stimulus presentation. Before beginning data collection, the software obtained a 2-point calibration using a telescoping bulls-eye or animated image of Elmo accompanied by soft sounds. The calibration was repeated if we did not obtain an accurate calibration. Prior to the familiarization trials, participants' attention was drawn to the center of the screen using a telescoping bulls-eye or flashing circle. For adults, the software obtained a 4-point validation using a telescoping bulls-eye at the end of data collection to assess whether tracking accuracy shifted from the initial calibration.

Inclusion Criteria

Infants' data were included if the dwell times to the entire scene recorded by the eye-tracker were congruent with the dwell times for each scene coded offline by a research assistant blind to the vocalization being tested. Infants ($n = 11$) were excluded from analyses when there was less than 50% agreement between the eye-tracker and offline coded data, reflecting more than 50% data loss by the eye-tracker, in more than two trials. Adults ($n = 3$) were excluded from analyses when the eye-tracker did not collect data due to loss or inattention during more than 30% of the task and when validation values from the eye-tracker indicated that the accuracy of tracking shifted more than 1.5 degrees of visual angle from the initial calibration.

Data Reduction and Analyses

Eye-tracking data were processed using SMI BeGaze™ Eye-tracking Analysis Software (Version 3.4, 2014). To assess infants' and adults' visual fixation patterns throughout the video, we defined multiple areas of interest (AOIs) including the Listener's face and body (657090 pixels), the Communicator's face (69276 pixels), the target (70338 pixels), and the

non-target (69552 pixels) objects (see Figure 1). AOIs remained the same size for the duration of the trials. However, the shapes of the Communicator, target, and non-target object AOIs were adjusted to avoid overlap between AOIs when the Communicator lifted the object below her face during familiarization and when the Listener lifted the object to the Communicator during test. The eye-tracker calculated fixation lengths and locations by filtering raw looking data by predetermined criteria (80 ms, 100 pixels of dispersion) for each individual participant.

Results

First, we established that our eye-tracking study replicated overall looking time patterns reported in previous infant VOE studies using a similar communicative scenario (Martin et al., 2012; Vouloumanos et al., 2014). Second, we examined how infants and adults process a communicative event involving speech in real time. Specifically, to determine if infants, like adults, interpreted the communicative event while it was occurring or evaluated it retrospectively after seeing the outcome, we examined infants' and adults' responses during the action segment of the test trial by analyzing infants' overall looking times to all AOIs, and infants' and adults' time course of looking at each AOI as the actions unfolded. As an exploratory analysis, we also examined infants' and adults' changes in pupil diameter as the actions unfolded (see Supplemental Materials).

Overall Looking at Combined AOIs During the Still Image Segment

Infant's overall looking times to the entire scene during the still image segment replicated previous VOE findings (Martin et al., 2012). We ran a 2 (vocalization: Speech, Cough) by 2 (outcome: Target, Non-Target) Analysis of Variance (ANOVA; SPSS vers. 22.0, IBM Corp., 2013) with vocalization as a between-subjects factor and outcome as a within-subjects factors on infants' looking time to all AOIs and found a significant interaction between vocalization and outcome ($F(1,30) = 4.23, p = .05, \eta^2 = .12$). Tests of simple effects suggest that when the Communicator spoke, infants looked longer at the AOIs when the Listener selected the non-target compared to the target object ($t(15) = 2.84, p = .01$). When the Communicator coughed, infants looked equally at the AOIs when the Listener selected either object ($t(15) = .03, p = .98$; for Means and Standard Errors see Table 1). There were no main effects of vocalization or outcome (both $ps > .19$).

Real-time Processing of the Communicative Event During the Action Segment

Overall looking at combined AOIs.—Infants' overall looking times to all AOIs during the action segment showed the same pattern of differential looking at the Non-target outcome for Speech but not Cough as during the still image segment, revealing that infants processed the communicative event while it was occurring. We ran the same 2×2 ANOVA on infants' looking time to all AOIs during the action segment and found a marginal interaction between vocalization and outcome ($F(1,30) = 3.96, p = .06, \eta^2 = .10$). Tests of simple effects suggest that when the Communicator spoke, infants looked longer at the AOIs when the Listener was selecting the non-target compared to the target object ($t(15) = 2.40, p = .02$). When the Communicator coughed, infants looked equally at the AOIs while the Listener was selecting either object ($t(15) = .51, p = .62$; see Table 1). There was a main

effect of vocalization, with infants looking longer at the action segment when the Communicator coughed than when she spoke ($F(1,30) = 12.38, p = .001, \eta^2 = .29$), and a main effect of outcome, with infants looking longer when the Listener was selecting the non-target compared to the target object ($F(1,30) = 4.13, p = .05, \eta^2 = .11$).

The same 2×2 ANOVA on infant looking times to all AOIs during the familiarization trials revealed no significant main effects or interactions (all $ps > .19$), confirming that infants attended equally during the familiarization phase. Finally, a 2 (vocalization: Speech, Cough) $\times 2$ (outcome: Target, Non-target) $\times 2$ (object: target, non-target) ANOVA on infants' looking times to each object in the first 1400 ms of the first familiarization trial, before the Communicator reached for the target object, also revealed no main effects or interactions (all $ps > .10$), confirming that infants did not show an initial preference for either object that may have influenced their looking behavior during the test trial.

Time course of looking at individual AOIs.—Next, we used linear spline models to analyze how infants' and adults' looking behavior to the Communicator, Listener, target, and non-target objects changed as the actions unfolded (see Figure 2 for a timeline of the actions). Linear spline models are a flexible extension of linear or polynomial growth curve models. They can be used to analyze a non-linear trend in which the trajectory of the trend is hypothesized to change at discrete points in time (i.e., when certain actions in the communicative event occur; Cudeck & Klebe, 2002; Grimm & Marcoulides, 2016). The trend is divided into a sequence of connected linear segments and the model simultaneously estimates a linear slope for the segment indicated as the baseline segment. Then, it evaluates how the slope in each remaining segment changes relative to the baseline segment (Fitzmaurice, Laird, & Ware, 2011).

First, we divided the action segment into 25 bins of 250 ms each and we calculated average proportion looking times to each AOI out of looking time to the entire scene. Bins of 250 ms allow us to capture adults' and infants' processing abilities because adults and infants require 200 ms to initiate a gaze shift when observing an action, and anticipate the outcome of an action about 200 ms before the action is complete (Rosander & von Hofsten, 2011). Second, we specified linear segments based on the timing of the actions for which we expected the trajectory of participants' looking to change and based on infants' and adults' response times when processing actions and vocalizations in previous studies (e.g., Rosander & von Hofsten, 2011; Swingle & Aslin, 2000; Thorgrimsson et al., 2015). We selected a baseline segment for each AOI that contained the actions most relevant to that AOI (segment 1 for Communicator; segment 2 for Listener, target, and non-target objects). For example, segment 1 for the Communicator is when the Communicator looked at the Listener and vocalized, which are actions that participants should process by attending to the Communicator. Then, the model can evaluate how participants' slope of looking at the Communicator in each comparison segment (segments 2 and 3) changes relative to participants' slope of looking at the Communicator when she vocalized (segment 1). Similarly, we chose segment 3 as the baseline segment for the Listener which is when the Listener reached for an object and presented it to the Communicator. We hypothesized this would elicit a change in participants' looking behavior to the Listener. Segment 2 is the baseline segment for the target and non-target objects which is when the Listener reached for

and grasped either object, and we predicted would also elicit a meaningful change in participants' looking at the objects which can then be compared to remaining comparison segments. Segment 2 begins 500ms earlier for the target compared to the non-target object because we expected that participants might anticipate where the Listener would reach and increase their looking at the target, not the non-target object before the Listener grasped it.

Finally, we used linear models to analyze how the slope of participants' proportion of looking at each AOI changed from segment to segment (Figure 3 for all AOIs together and Figure 4 for each AOI separately; graphs were made in R using ggplot2 vers. 2.1.0; R Core Team, 2013; Wickham, 2009). Thus, the coefficient for each baseline segment represents the slope of participants' looking during the baseline segment. The coefficients for each comparison segment represent the change in participants' slope of looking from the baseline segment to the comparison segment. We calculated the simple slope for each comparison segment by subtracting the coefficient for the comparison segment from the coefficient for the baseline segment (denoted as b).

To further examine whether participants' changes in looking behavior to each AOI differed based on age (adult vs. infant), the Communicator's vocalization (speech vs. cough), or the experimental outcome of the event (target vs. non-target), we included fixed effects of age, vocalization, experimental outcome, and their interactions with the linear segments. To limit the number of comparisons in each model and decrease the chances of Type I error, we only included the interactions for which we hypothesized an effect. All significant interactions were followed by tests of simple effects, which were corrected for multiple comparisons using the Bonferroni adjustment. Residuals for the linear spline models were approximately normally distributed, though they were slightly heteroscedastic with some clustering at 0 and 1. However, the slight clustering did not warrant transformations to the data, which could make the results more difficult to interpret (Grayson, 2004).

Categorical diagnostic group and experimental outcome predictor variables were effects coded, such that adults, Speech vocalization condition, and Target outcome were coded as -1 , and infants, Cough vocalization condition, and Non-target outcome were coded as 1 . All linear spline models were analyzed using the Proc Mixed package in SAS[®] software (vers. 9.4; SAS Institute Inc., 2014)

Communicator.—Both infants' and adults' looking behavior to the Communicator changed as the actions unfolded. Across outcome, both infants and adults significantly increased their looking at the Communicator when she vocalized (segment 1) and decreased their looking at the Communicator when the Listener leaned forward to reach for either object (segment 2). When the Listener selected an object and presented it to the Communicator (segment 3), both infants and adults increased their looking at the Communicator more after she coughed ($b_{cough} = .14$, $b_{cough} = .11$, $SE = .01$) compared to when she spoke ($b_{speech} = .13$, $b_{speech} = .08$, $SE = .01$, $t(2505) = 2.45$, $p = .03$). There was a marginal interaction between vocalization and segment 1, however pairwise comparisons were not significant ($p = .12$; see Figure 2 for a timeline of the actions, Figure 4 for raw looking behavior, and Table 2 for model estimates).

Listener.—Both infants' and adults' looking behavior to the Listener also changed as the actions unfolded. Across vocalization and outcome, both infants and adults increased their looking at the Listener when the Communicator looked at the Listener (segment 1). When the Communicator vocalized and the Listener leaned forward and began reaching for an object, infants and adults increased their looking at the Listener (segment 2). However, infants ($b_{\text{infants}} = .03$, $b_{\text{infants}} = .11$, $SE = .01$) increased their looking at the Listener more than adults ($b_{\text{adults}} = .01$, $b_{\text{adults}} = .07$, $SE = .01$, $t(2724) = 3.00$, $p = .01$).

Critically for the hypothesis that participants will show differential looking at the Listener after the Communicator speaks but not when she coughed, there was significant three-way interaction between vocalization, outcome, and segment 2 suggesting that both infants and adults showed a larger change (increase) in their looking at the Listener after the Communicator spoke and the Listener leaned forward to reach for the non-target ($b_{\text{non-target}} = .02$, $b_{\text{non-target}} = .09$, $SE = .01$) compared to the target object ($b_{\text{target}} = .01$, $b_{\text{target}} = .095$, $SE = .01$, $t(2724) = 2.36$, $p = .04$). However, when the Communicator coughed, participants showed no difference in their relative change in looking at the Listener when she leaned forward to reach for either the target ($b_{\text{target}} = .03$, $b_{\text{target}} = .11$, $SE = .01$) or the non-target object ($b_{\text{non-target}} = .03$, $b_{\text{non-target}} = .10$, $SE = .01$, $t(2724) = -.44$, $p = 1.0$).

There was a similar three-way interaction between vocalization, outcome, and segment 3 suggesting that when the Listener selected an object and presented it to the Communicator (segment 3), both infants and adults showed a marginally larger decrease in their looking at the Listener after the Communicator spoke and the Listener selected the non-target ($b_{\text{non-target}} = -.06$, $SE = .01$) compared to the target object ($b_{\text{target}} = -.05$, $SE = .01$, $t(2724) = 2.05$, $p = .08$). However, participants showed no difference in looking at the Listener when she selected either the non-target ($b_{\text{non-target}} = -.07$, $SE = .01$) or target object ($b_{\text{target}} = -.08$, $SE = .01$, $t(2724) = 1.21$, $p = .46$) after the Communicator coughed. However, across outcome and vocalization, infants ($b_{\text{infants}} = -.07$, $SE = .01$) decreased their looking at the Listener marginally more than adults ($b_{\text{adults}} = -.06$, $SE = .00$, $t(2724) = 2.00$, $p = .09$).

Thus, when the Communicator spoke, not when she coughed, both groups showed a larger difference between their increase in looking at the Listener in segment 2 and decrease in looking at the Listener in segment 3 when the Listener reached for the non-target compared to the target object. Because spline models assume that the linear functions in each segment are connected, a larger difference in slopes may suggest that the peak of the curve, where segments 2 and 3 meet, was higher and participants immediately showed a greater proportion of looking after the Communicator spoke and the Listener reached for the non-target compared to the target object (see Figure 2 for a timeline of the actions, Figure 4 for raw looking behavior, and Table 2 for model estimates).

Target and non-target objects.—Both infants and adults looked most at the object the Listener selected. When the Communicator spoke or coughed and the Listener leaned forward to reach for an object, both infants and adults showed no change in looking at either object (segment 1). When the Listener reached for and grasped an object (segment 2), participants increased their looking more at whichever object the Listener chose (target object: $b_{\text{target}} = .19$, $SE = .01$, $t(2723) = 12.19$, $p < .001$; non-target object: $b_{\text{non-target}} = .22$,

$SE = .01$, $t(2727) = 13.57$, $p < .001$). Finally, when the Listener presented an object to the Communicator (segment 3), participants decreased their looking at whichever object the Listener presented (target object: $b_{target} = -.04$, $b_{target} = -.22$, $SE = .01$, $t(2723) = 10.43$, $p < .001$; non-target object: $b_{non-target} = -.08$, $b_{non-target} = -.29$, $SE = .02$, $t(2723) = 11.82$, $p < .001$).

Across both outcomes, an age by segment 3 interaction suggested that when the Listener presented the non-target object to the Communicator (segment 3) adults ($b_{adults} = -.07$, $b_{adults} = -.19$, $SE = .02$) showed a greater change (decrease) in their slope of looking at the non-target object compared to infants ($b_{infants} = -.01$, $b_{infants} = -.12$, $SE = .01$, $t(2723) = 3.14$, $p = .002$). A larger decrease in looking at the non-target object when the Listener presented it to the Communicator (segment 3) compared to when the Listener selected the non-target object (segment 2) may suggest that the peak of the curve was higher and adults showed a greater proportion of looking at the non-target object immediately after the Listener selected it compared to infants and looked away from the non-target object faster as the Listener presented it to the Communicator, compared to infants.

Across all segments, main effects of age suggested that adults looked at the target and non-target objects more than infants.

There were significant interactions between age, outcome and segments 1 and 2 for looking at the target object and age, vocalization, and segment 2 for looking to the non-target object, however, pairwise comparisons were not significant ($ps > .61$; see Figure 2 for a timeline of the actions, Figure 4 for raw looking behavior and Table 2 for model estimates).

Discussion

While previous VOE studies (e.g., Krehm et al., 2014; Martin et al., 2012; Song et al., 2008; Vouloumanos et al., 2014, 2012), play-based procedures (e.g., Grosse et al., 2010; Liszkowski et al., 2008; Schulze & Tomasello, 2015), and eye-tracking studies (Thorgrímsson et al., 2015) have revealed much about what infants understand about communicative social interactions, we examined how infants and adults process communicative events in real time, while they occur. Tracking infant and adult eye movements and during a communicative event and its outcome suggests that 12-month-olds, like adults, recognize the communicative function of speech in a social exchange, and they process the communicative event while it is occurring. At the same time, differences in 12-month-olds' and adults' looking behaviors suggest that infants' processing speeds can be slower than adults.

Infants' looking behavior during the event suggested that they begin processing communicative events while they are occurring. Infants' overall looking time to the AOIs during the action segment of the test trial reflected similar patterns of looking as infants' overall looking times to the still image segment, suggesting that, as the event was occurring, infants recognized that the Listener would select the target object after the Communicator spoke but not when she coughed. Thus, infants understand how others should respond to speech, but not non-speech, as the actions are unfolding and do not need to wait until the

actions are complete to begin processing or interpreting the scene. In other words, infants do not need to see the outcome of a communicative event before they begin to process it. Our findings first replicate previous findings from VOE studies (Martin et al., 2012; Vouloumanos et al., 2014) in showing that infants understand how speech communicates in social interactions because they increase their looking to the incongruent outcome of communicative scenes. Our findings expand on these VOE studies to suggest that infants use their understanding of the communicative function of speech to interpret or process communicative events while they are occurring, but do not necessarily anticipate the outcome of a communicative interaction.

The way infants and adults processed the communicative event by shifting their attention between the AOIs revealed that infants process communicative events with adult-like accuracy. When the Communicator spoke or coughed, both infants and adults looked at the Communicator. Then, they looked at the Listener when she leaned forward and reached for an object. Just as the Listener grasped an object, participants looked at whichever object she selected. Finally, when the Listener presented the object to the Communicator, participants looked back at the Communicator. Just as infants process nonverbal actions similarly to adults (reviewed in Gredebäck & Daum, 2015; see also Rosander & von Hofsten, 2011), our findings suggest that infants and adults can also process communicative vocal events in very similar ways.

Both infants and adults differentiated between potentially communicative events involving speech and events involving non-speech, showing that both groups understand that speech but not non-speech transfers information in a social scenario. Consistent with our predictions, infants' and adults' time course of looking at the Listener suggest that participants recognized the Listener would choose the target object only if the Communicator spoke and looked faster and longer if she selected the non-target object instead. Both infants and adults seemed agnostic about what the Listener would do when the Communicator coughed, showing similar looking behaviors whether the target or non-target objects were chosen. For infants, speech is not only directed towards others (Thorgrímsson et al., 2015) but also influences how others will respond. Both infants and adults also identified a cough as an incongruent vocalization in a communicative context. When the Listener presented an object to the Communicator, participants looked at the Communicator more after she coughed than after she spoke. This may reflect infants' and adults' uncertainty at how an uninformative cough vocalization functions in a communicative event and suggests they may not expect coughing to elicit a response from the Listener at all (Thorgrímsson et al., 2015).

Overall, infants and adult participants show striking similarities in the way they understand and process communicative events involving speech: they processed the communicative event as it was occurring but did not anticipate the outcome of the communicative event. However, infants also differed from adults in their attention to the Listener and the non-target object as the actions unfolded, suggesting that infants' processing speeds can be slower than adults. Across vocalization and outcome, infants looked at the Listener more than adults when the Communicator vocalized and the Listener leaned forward and began reaching for an object. Furthermore, for both speech and coughs, adults compared to infants

looked away from the non-target object faster when the Listener presented the non-target object to the Communicator. These differences in looking behavior between infants and adults suggest that infants may have required more time looking at the Listener when she responded to the Communicator, and at the non-target object when the Listener presented it to the Communicator to process these actions, compared to adults. These findings are consistent with previous research suggesting that infants take slightly longer to shift their gaze when processing others' object-directed actions compared to adults (e.g., Flanagan & Johansson, 2003; Rosander & von Hofsten, 2011). Despite differences in processing speeds, overwhelming similarities in infants' and adults' looking behaviors suggest that 12-month-olds understand and process communicative events with adult-like accuracy.

We hypothesized that infants and adults may make predictions about the outcome of the communicative event before it occurred by looking anticipatorily at the target object after the Communicator spoke and before the Listener selected the object (Brandone et al., 2014; Cannon & Woodward, 2012; Fawcett & Gredebäck, 2013; Thorgrimsson et al., 2014). We found no evidence that infants and adults looked in anticipation at the object that was congruent with the Communicator's vocalization. Both infants and adults began increasing their looking at the target object 750 ms before the Listener grasped the target object, which is 500 ms earlier than when they began increasing their looking at the non-target object. However, this pattern did not differ based on the Communicator's vocalization. This lack of clear anticipation may be because the scene was complex and participants did not have enough time to look before the Listener responded to the Communicator's vocalization. Previous studies assessing infants' anticipatory looks used simple scenes that only examined the anticipation of a single reaching action (Brandone et al., 2014), as compared with the two actors vocalizing and interacting with multiple objects in the current study. Furthermore, infants were often allowed a longer period of time (6 to 8 s compared with 700 ms in the current study) to observe an action before anticipating its outcome (Brandone et al., 2014; Fawcett & Gredebäck, 2013) or the outcome of the action was never presented (Brandone et al., 2014; Cannon & Woodward, 2012). Our complex scene and short time window may have precluded infants and adults making clear anticipatory looks.

Limitations to the current study included the fact that each infant and adult only saw the action segment of the test trial once with the Target outcome and once with the Non-target outcome. Eye-tracking data from multiple trials (e.g., Thorgrimsson et al., 2014, 2015) of the action segment may yield a more stable measure of participants' looking behavior. Furthermore, while we did measure the amount of data we lost from the eye-tracker and set appropriate inclusion criteria based on the data loss, we did not have an accurate measure of the precision of the eye-tracking data for infants which could influence the looking behaviors to the specific AOIs (Oakes, 2010, 2012). Verifying the calibration by presenting objects at certain locations on the screen would allow us to better evaluate whether infants gaze location is accurately approximated by the eye-tracker (Oakes, 2012). Finally, as an exploratory analysis, we examined infants' and adults' changes in pupil diameter during the action segment. We found that participants' pupil diameter changed as the actions of the scene unfolded. However, the results were inconsistent with participants' changes in looking behavior and were not clearly interpretable. The results of pupil diameter analyses are included as supplemental information.

An important question for future research is whether infants' processing of communicative events shapes language and social abilities later in life. The ability to recognize the communicative function of speech without knowing its lexical content and quickly process the event while it is occurring may be an early mechanism for detecting the meaning of novel words (Akhtar et al., 2001) and for acquiring knowledge from third-party communicative interactions (reviewed in Vouloumanos & Waxman, 2014). The degree to which early processing of communicative events predicts later language and social abilities may also be valuable for identifying early difficulties in social and communicative skills, which are often characteristic of individuals with autism spectrum disorder (American Psychiatric Association, 2013). Consistent with this possibility, 6-month-olds who are later diagnosed with autism spectrum disorder attend less to social scenes overall than typical infants, and especially attend less to the actor's face and actions (Chawarska, Macari, & Shic, 2013).

Conclusion

Twelve-month-olds, like adults, recognize that speech, but not non-speech, can communicate information to others, and process the communicative event online as the event is unfolding. Infants and adults show many similarities in the way they process communicative events involving speech, however, differences in the magnitude and timing of their looking behaviors revealed that, at times, 12-month-olds process more slowly than adults. We found no evidence that infants or adults are making predictions about the outcome of a communicative event before the outcome is presented, though it is possible that participants could anticipate the outcome if they were given more time between the communicative event and its outcome. Even before knowing or saying many words, infants can already use their knowledge about the communicative function of speech to quickly and efficiently process a communicative event. These processing abilities may be a mechanism not only for learning language, but also for acquiring knowledge from third-party communicative interactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R01HD072018. We'd like to thank members of the NYU Infant Cognition and Communication Lab, the parents and infants who participated, and Pat Shrout for his assistance with analyses.

References

- Akhtar N, Jipson J, & Callanan MA (2001). Learning words through overhearing. *Child Development*, 72(2), 416–430. 10.1016/S0163-6383(98)91471-0 [PubMed: 11333075]
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*.
- Aslin RN (2007). What's in a look? *Developmental Science*. 10.1111/j.1467-7687.2007.00563.x

- Augusti EM, Melinder A, & Gredebäck G (2010). Look who's talking: Pre-verbal infants' perception of face-to-face and back-to-back social interactions. *Frontiers in Psychology*, 1, 161 10.3389/fpsyg.2010.00161 [PubMed: 21833226]
- Baillargeon R, Scott RM, & He Z (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*. 10.1016/j.tics.2009.12.006
- Balaban MT, & Waxman SR (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, 64(1), 3–26. 10.1006/jecp.1996.2332 [PubMed: 9126625]
- Beier JS, & Spelke ES (2012). Infants' developing understanding of social gaze. *Child Development*, 83(2), 486–496. 10.1111/j.1467-8624.2011.01702.x [PubMed: 22224547]
- Bergelson E, & Swingley D (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253–3258. 10.1073/pnas.1113380109
- Brandone AC, Horwitz SR, Aslin RN, & Wellman HM (2014). Infants' goal anticipation during failed and successful reaching actions. *Developmental Science*, 17(1), 23–34. 10.1111/desc.12095 [PubMed: 24112439]
- Cannon EN, & Woodward AL (2012). Infants generate goal-based action predictions. *Developmental Science*, 15(2), 292–298. 10.1111/j.1467-7687.2011.01127.x [PubMed: 22356184]
- Chawarska K, Macari S, & Shic F (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, 74(3), 195–203. 10.1016/j.biopsych.2012.11.022 [PubMed: 23313640]
- Cheung H, Xiao W, & Lai CM (2012). Twelve-month-olds' understanding of intention transfer through communication. *PLoS ONE*, 7(9), e46168 10.1371/journal.pone.0046168 [PubMed: 23029427]
- Csibra G, & Gergely G (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153. 10.1016/j.tics.2009.01.005 [PubMed: 19285912]
- Csibra G, & Gergely G (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1567), 2010.1098/rstb.2010.0319
- Cudeck R, & Klebe KJ (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods*, 7(1), 41–63. 10.1037//1082-989X.7.1.41 [PubMed: 11928890]
- Daum MM, & Gredebäck G (2011). The development of grasping comprehension in infancy: Covert shifts of attention caused by referential actions. *Experimental Brain Research*, 208(2), 297–307. 10.1007/s00221-010-2479-9 [PubMed: 21082313]
- Falck-Ytter T (2015). Gaze performance during face-to-face communication: A live eye tracking study of typical children and children with autism. *Research in Autism Spectrum Disorders*, 17, 78–85. 10.1016/j.rasd.2015.06.007
- Fawcett C, & Gredebäck G (2013). Infants use social context to bind actions into a collaborative sequence. *Developmental Science*, 16(6), 841–849. 10.1111/desc.12074 [PubMed: 24118711]
- Ferry AL, Hespos SJ, & Waxman SR (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development*, 81(2), 472–479. 10.1111/j.1467-8624.2009.01408.x [PubMed: 20438453]
- Fitzmaurice GM, Laird NM, & Ware JH (2011). *Applied longitudinal analysis*. John Wiley & Sons, Inc (2nd ed.).
- Flanagan JR, & Johansson RS (2003). Action plans used in action observation. *Nature*, 424(6950), 769–771. 10.1038/nature01861 [PubMed: 12917683]
- Fulkerson AL, & Waxman SR (2007). Words (but not Tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, 105(1), 218–228. 10.1016/j.cognition.2006.09.005 [PubMed: 17064677]
- Grayson D (2004). Some myths and legends in quantitative psychology. *Understanding Statistics*, 3(1), 101–134. 10.1207/s15328031us0302_3
- Gredebäck G, & Daum MM (2015). The microstructure of action perception in infancy: Decomposing the temporal structure of social information processing. *Child Development Perspectives*, 9(2), 79–83. 10.1111/cdep.12109 [PubMed: 27642367]

- Gredebäck G, & Melinder A (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition*, 114(2), 197–206. 10.1016/j.cognition.2009.09.004 [PubMed: 19800056]
- Gredebäck G, Melinder A, & Daum M (2010). The development and neural basis of pointing comprehension. *Social Neuroscience*, 5(5–6), 441–450. 10.1080/17470910903523327 [PubMed: 20162491]
- Grimm KJ, & Marcoulides K (2016). Individual change and the timing and onset of important life events: Methods, models, and assumptions. *International Journal of Behavioral Development*, 40(1), 87–96. 10.1177/0165025415580806
- Grosse G, Behne T, Carpenter M, & Tomasello M (2010). Infants communicate in order to be understood. *Developmental Psychology*, 46(6), 1710–1722. 10.1037/a0020727 [PubMed: 20873924]
- Hamlin JK (2015). The case for social evaluation in preverbal infants: Gazing toward one's goal drives infants' preferences for Helpers over Hinderers in the hill paradigm. *Frontiers in Psychology*, 6(JAN), 1563 10.3389/fpsyg.2015.00563 [PubMed: 26578995]
- Krehm M, Onishi KH, & Vouloumanos A (2014, October 2). I See Your Point: Infants Under 12 Months Understand That Pointing Is Communicative. 10.1080/15248372.2012.736112
- Langdon R, & Smith P (2005). Spatial cueing by social versus nonsocial directional signals. *Visual Cognition*, 12(8), 1497–1527. 10.1080/13506280444000805
- Lieberman Z, Kinzler KD, & Woodward AL (2014). Friends or foes: Infants use shared evaluations to infer others' social relationships. *Journal of Experimental Psychology: General*, 143(3), 966–971. 10.1037/a0034481 [PubMed: 24059843]
- Liszkowski U, Carpenter M, & Tomasello M (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, 108(3), 732–739. 10.1016/j.cognition.2008.06.013 [PubMed: 18721918]
- Mackenzie H, Graham SA, & Curtin S (2011). Twelve-month-olds privilege words over other linguistic sounds in an associative learning task. *Developmental Science*, 14(2), 249–255. 10.1111/j.1467-7687.2010.00975.x [PubMed: 22213898]
- Martin A, Onishi KH, & Vouloumanos A (2012). Understanding the abstract role of speech in communication at 12months. *Cognition*, 123(1), 50–60. 10.1016/j.cognition.2011.12.003 [PubMed: 22209584]
- Oakes LM (2010). Infancy guidelines for publishing eye-tracking data. *Infancy*, 15(1), 1–5. 10.1111/j.1532-7078.2010.00030.x
- Oakes LM (2012). Advances in eye tracking in infancy research *Infancy*, 17(1), 1–8. 10.1111/j.1532-7078.2011.00101.x
- R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria Retrieved from <http://www.r-project.org/>
- Rosander K, & von Hofsten C (2011). Predictive gaze shifts elicited during observed and performed actions in 10-month-old infants and adults. *Neuropsychologia*, 49(10), 2911–2917. 10.1016/j.neuropsychologia.2011.06.018 [PubMed: 21722655]
- SAS Institute Inc (2014). SAS 9.4 Users Guide. Cary, NC, USA: SAS Institute, Inc.
- Schulze C, & Tomasello M (2015). 18-month-olds comprehend indirect communicative acts. *Cognition*, 136, 91–98. 10.1016/j.cognition.2014.11.036 [PubMed: 25497519]
- SensoMotoric Instruments GmbH. (n.d.). Tetlow, Germany Retrieved from www.smivision.com
- Song H joo, Onishi KH, Baillargeon R, & Fisher C (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, 109(3), 295–315. 10.1016/j.cognition.2008.08.008 [PubMed: 18976745]
- Swingle D, & Aslin RN (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147–166. 10.1016/S0010-0277(00)00081-0 [PubMed: 10856741]
- Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, & Sedivy JC (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. 10.1126/science.7777863 [PubMed: 7777863]

- Thorgrimsson GB, Fawcett C, & Liszkowski U (2014). Infants' expectations about gestures and actions in third-party interactions. *Frontiers in Psychology*, 5(April), 1–7. 10.3389/fpsyg.2014.00321 [PubMed: 24474945]
- Thorgrimsson GB, Fawcett C, & Liszkowski U (2015). 1- and 2-year-olds' expectations about third-party communicative actions. *Infant Behavior and Development*, 39, 53–66. 10.1016/j.infbeh.2015.02.002 [PubMed: 25766104]
- Tincoff R, & Jusczyk PW (1999). Some beginnings of word comprehension in 6-Month-Olds. *Psychological Science*, 10(2), 172–175. 10.1111/1467-9280.00127
- Tincoff R, & Jusczyk PW (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4), 432–444. 10.1111/j.1532-7078.2011.00084.x
- Vouloumanos A, Druhen MJ, Hauser MD, & Huizink AT (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, 106(44), 18867–72. 10.1073/pnas.0906049106 [PubMed: 19846770]
- Vouloumanos A, Martin A, & Onishi KH (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science*, 17(6), 872–879. 10.1111/desc.12170 [PubMed: 24835877]
- Vouloumanos A, Onishi KH, & Pogue A (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences of the United States of America*, 109(32), 12933–7. 10.1073/pnas.1121057109 [PubMed: 22826217]
- Vouloumanos A, & Waxman SR (2014). Listen up! Speech is for thinking during infancy. 10.1016/j.tics.2014.10.001
- Wickham H (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. 10.1007/978-0-387-98141-3
- Woodward A (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1–34. 10.1016/S0010-0277(98)00058-4 [PubMed: 9871370]
- Xu F (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, 85(3), 223–250. 10.1016/S0010-0277(02)00109-9 [PubMed: 12169410]
- Xu F, Cote M, & Baker A (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science*, 16(5), 372–377. 10.1111/j.0956-7976.2005.01543.x [PubMed: 15869696]

Figure 1.

Method. Infants saw a sequence of three trials in a pre-recorded video of a third-party communicative interaction. During the Familiarization trials, the Communicator looked at two novel objects, and grasped the target object. During the Action segment of the test trial, the Communicator could no longer reach the objects so she vocalized to the Listener using speech or cough. The Listener selected either the target or non-target object. In the Still Image segment, the final image of the test trial froze for the remainder of the test trial. AOIs used in all trials are shown in the image of the Still Image segment of the test trial.

Figure 2. Results. Timeline of the start and end of each linear segment included in the linear spline models for each AOI aligned with the start and end of each action in the action segment of the test trial. “C” is Communicator and “L” is Listener.

Figure 3. Results. Time course of fixations to the Communicator (red), Listener (green), target (blue), and non-target (purple) objects during the actions segment of the test trial before the Target outcome and Non-target outcome in the Speech condition (left panel) and Cough condition (right panel) for adults (top panel) and infants (bottom panel). Gray lines indicate the times during which each action occurred, where “C” is Communicator and “L” is Listener.

Figure 4.

Results. Time course of fixations $\pm$ *SEM* (ribbon) to the A) Communicator (red), B) Listener (green), C) Target Object (blue) and D) Non-target Object (purple) before the Target outcome (solid lines) and Non-target outcome (dashed lines) in the Speech condition (left panel) and Cough condition (right panel) for adults (top panel) and infants (bottom panel). Colored boxes indicate the linear segments included in analyses for looking times to each AOI.

Table 1.

Marginal Means and Standard Errors of Infants' Looking Times (ms) to all AOIs During the Action and Still Image Segments of the Test Trial

	Still Image		Action	
	Target	Non-Target	Target	Non-Target
Speech	7677 (1430)	10439 (1207) *	3836 (254)	4549 (242) **
Cough	7484 (1430)	6900 (1207)	5252 (253)	5260 (242)

Note.

*
p < .05.

**
p < .01.

Table 2.

Fixed effects of linear spline models for age, outcome, time segments, and their interactions predicting proportion of looking at the Communicator, Listener, target, and non-target objects.

	Communicator					Listener					Target object					Non-target object				
	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>
Intercept	0.21	0.05	0.00	0.11	0.32	0.57	0.02	0.00	0.52	0.62	0.04	0.01	0.02	0.01	0.07	0.37	0.02	0.00	0.33	0.40
Age	-0.07	0.05	0.20	-0.18	0.04	-0.08	0.02	0.00	-0.13	-0.04	0.04	0.01	0.00	0.01	0.07	0.07	0.02	0.00	0.03	0.11
Vocalization	-0.06	0.04	0.10	-0.14	0.01	-0.08	0.02	0.00	-0.12	-0.03	0.01	0.01	0.54	-0.02	0.03	0.01	0.01	0.61	-0.01	0.02
Outcome	-0.01	0.01	0.32	-0.04	0.01	-0.03	0.02	0.11	-0.07	0.01	0.01	0.01	0.69	-0.02	0.03	-0.33	0.02	0.00	-0.37	-0.29
Segment 1	0.04	0.01	0.00	0.02	0.06	0.11	0.02	0.00	0.08	0.14	-0.09	0.01	0.00	-0.11	-0.08	-0.12	0.01	0.00	-0.13	-0.10
Segment 2	-0.11	0.01	0.00	-0.13	-0.08	0.09	0.01	0.00	0.07	0.10	0.09	0.01	0.00	0.08	0.11	0.11	0.01	0.00	0.10	0.13
Segment 3	0.10	0.01	0.00	0.08	0.11	-0.07	0.00	0.00	-0.07	-0.06	-0.12	0.01	0.00	-0.14	-0.10	-0.16	0.01	0.00	-0.18	-0.14
Age*Segment 1	0.01	0.01	0.37	-0.01	0.03	0.02	0.02	0.17	-0.01	0.05	0.00	0.01	0.60	-0.01	0.02	0.00	0.01	0.70	-0.02	0.01
Age*Segment 2	-0.01	0.01	0.46	-0.04	0.02	-0.02	0.01	0.00	-0.03	-0.01	0.00	0.01	0.89	-0.02	0.01	0.01	0.01	0.27	-0.01	0.02
Age*Segment 3	0.01	0.01	0.15	0.00	0.02	0.01	0.00	0.05	0.00	0.01	-0.02	0.01	0.14	-0.04	0.00	-0.04	0.01	0.00	-0.06	-0.01
Vocalization*Segment 1	0.01	0.00	0.06	0.00	0.01	-0.02	0.01	0.00	-0.03	-0.01	0.00	0.00	0.46	0.00	0.01	0.00	0.00	0.61	-0.01	0.01
Vocalization*Segment 2	-0.01	0.01	0.01	-0.02	0.00	0.01	0.00	0.01	0.00	0.02	-0.02	0.00	0.58	-0.01	0.00	0.00	0.00	0.60	-0.01	0.00
Vocalization*Segment 3	-0.01	0.01	0.01	-0.02	0.00	0.01	0.00	0.01	0.00	0.02	-0.09	0.01	0.00	-0.11	-0.07	0.10	0.01	0.00	0.08	0.12
Outcome*Segment 1						-0.01	0.01	0.18	-0.02	0.00	0.09	0.01	0.00	0.08	0.11	-0.10	0.01	0.00	-0.12	-0.09
Outcome*Segment 2						0.00	0.00	0.56	0.00	0.01	-0.11	0.01	0.00	-0.13	-0.09	0.14	0.01	0.00	0.11	0.16
Outcome*Segment 3						-0.04	0.02	0.04	-0.08	-0.00	-0.01	0.01	0.68	-0.03	0.02	0.00	0.01	0.79	-0.02	0.02
Vocalization*Outcome											0.00	0.00	0.92	0.00	0.00					
Age*Vocalization*Segment 1	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.78	-0.01	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.98	0.00	0.00
Age*Vocalization*Segment 2																				
Age*Vocalization*Segment 3	0.00	0.00	0.78	-0.01	0.01	0.00	0.00	0.82	-0.01	0.00										
Age*Outcome*Segment 1											0.00	0.00	0.01	0.00	0.01					
Age*Outcome*Segment 2						0.00	0.00	0.40	0.00	0.01	-0.01	0.00	0.00	-0.01	0.00	0.00	0.04	0.00	0.00	0.00
Age*Outcome*Segment 3	0.00	0.00	0.92	-0.01	0.01	0.00	0.00	0.78	-0.01	0.00										
Vocalization*Outcome*Segment 1											0.00	0.00	0.80	-0.01	0.01					
Vocalization*Outcome*Segment 2						-0.01	0.00	0.04	-0.02	-0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.36	0.00	0.00	0.00
Vocalization*Outcome*Segment 3	0.00	0.00	0.59	-0.01	0.00	0.01	0.00	0.02	0.00	0.01										

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Communicator				Listener				Target object				Non-target object			
	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>	<i>q</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>
Age*Vocalization*Outcome*Segment 1	0.00	0.00	0.16	−0.01	0.00	0.00	0.50	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.74	0.00
Age*Vocalization*Outcome*Segment 2	0.00	0.00			0.00	0.00	0.56	−0.01	0.00	0.00	0.27	0.00	0.00	0.00	0.74	0.00
Age*Vocalization*Outcome*Segment 3	0.00	0.00			0.00	0.00	0.50	0.00	0.01							0.00

Note. Age was coded as adult = −1, infant = 1. Vocalization was coded as Speech = −1, Cough = 1. Outcome was coded as Target = −1, Non-target = 1. Segment 1 was indicated as the baseline segment for the Communicator and Segment 2 was indicated as the baseline segment for the Listener, Target, and Non-target objects. Coefficients for comparison segments represent the change in slope from the baseline segment to the comparison segment.