



Published in final edited form as:

*Child Dev.* 2011 May ; 82(3): 902–915. doi:10.1111/j.1467-8624.2011.01579.x.

## Third-Party Social Interaction and Word Learning from Video

**Katherine O’Doherty,**

Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Pl., Nashville, TN 37203-5721

**Georgene L. Troseth,**

Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Pl., Nashville, TN 37203-5721

**Priya M. Shimpi,**

Department of Psychology, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064

**Elizabeth Goldenberg,**

Department of Psychology, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064

**Nameera Akhtar, and**

Department of Psychology, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064

**Megan M. Saylor**

Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Pl., Nashville, TN 37203-5721

Katherine O’Doherty: katherine.odoherty@vanderbilt.edu

### Abstract

In previous studies, very young children have learned words while “overhearing” a conversation, yet they have had trouble learning words from a person on video. In Study 1, 64 toddlers (mean age = 29.8 months) viewed an object-labeling demonstration in one of four conditions. In two, the speaker (present or on video) directly addressed the child and in two, the speaker addressed another adult who was present or was with her on video. Study 2 involved two follow-up conditions with 32 toddlers (mean age = 30.4 months). Across the two studies, the results indicated that toddlers learned words best when participating in or observing a *reciprocal social interaction* with a speaker who was present or on video.

---

From early infancy, children are exposed to information from varied sources, including parents and siblings, overheard conversations (“third party” social interactions --Akhtar, 2005), and television. A general challenge young children face is determining when and from whom to take information. By age two, children are sensitive to the presence or absence of referential social cues signaling a learning opportunity. Baldwin and Moses (2001) relay an anecdote in which their 2-year-old son, playing with new toys while they watched the evening news, suddenly declared, “No legal precedent!” Although he obviously had liked the sound of the phrase when it had emanated from the TV, this discerning youngster realized that the words did not refer to the objects he had been examining, and did not begin to use them as object labels. Baldwin and Moses go on to point out the frequency

---

Priya M. Shimpi is now at Mills College, Oakland, CA.

of errors that would occur if children relied on temporal associations to learn words, rather than on social “clues” to a speaker’s referential intentions.

Across the first year of life, children become aware of a range of such clues that indicate the intention to communicate information. Csibra and Gergely (2006) argue that from birth, infants are aware of social signals including eye contact and the prosody of infant-directed speech. Young infants’ attention is directed by such cues: newborns attend more to an individual using infant-directed than adult-directed speech (Cooper & Aslin, 1990) and 6-month-olds follow the gaze of a person who first makes eye contact with them (but not a person with averted gaze) to a specific target object (Csibra & Gergely, 2009). Social signals indicate to infants that a “pedagogical” or teaching situation is taking place, facilitating knowledge transfer (Csibra & Gergely, 2006; Gergely, Egyed, & Király, 2007). Throughout the first years of life, the ability to read meaning into social signals becomes increasingly sophisticated. By 9 to 12 months, infants use cues such as eye gaze to draw conclusions about an actor’s goals (Woodward, 2003) and, shortly after their first birthday, to share an understanding of an actor’s intentions (Tomasello, Carpenter, Call, Behne, & Moll, 2005).

By their second year, children are skilled at recognizing opportunities to learn from other people. Individual social cues (e.g., infant-directed prosody) do not appear to automatically “trigger” toddlers’ perception of a learning situation; rather, children rely on evidence from the context in which these cues are presented to determine when intentional communication is occurring. For example, 18-to 20-month-olds learned a novel object label uttered by a person looking at the object with them; in contrast, an utterance in infant-directed speech from a disembodied voice, contingent on children’s own attention to an object but lacking cues to reference, was not sufficient for children’s learning (Baldwin, Markman, Bill, Desjardins, Irwin, & Tidball, 1996). In the context of social interaction, when a particular cue (e.g., information about gaze direction) is uninformative, toddlers adaptively use any of a variety of communicative cues (such as emotional expressions) that may be available (Akhtar & Tomasello, 1996; Tomasello & Barton, 1994; Tomasello, Strosberg, & Akhtar, 1996).

Along with their awareness of social cues, toddlers appear to understand, and can flexibly adopt, the complementary roles involved in social interaction. In one study, an adult taught 18-month-olds a “placing” game (held out a plate for the child to deposit a toy); when given the plate, the children took the adult’s role, offering the plate to allow the adult to place the toy (Carpenter, Tomasello, & Striano, 2005). Carpenter and her colleagues point out that collaborative role reversal such as this involves recognizing and then adopting the intentions of the other, and seeing the self and other as in some ways “the same” and interchangeable.

By the middle of the second year, children also appear to have different expectations for information conveyed along with referential cues than they do when these cues are absent (e.g., when the information comes from a person with their back turned, or a radio speaker). For example, sixteen-month-old children looked longer when an adult seated next to them *looked at and mislabeled* an object (called a shoe “*a ball*”) compared to when the person correctly labeled the object (Koenig & Echols, 2003). However, they looked longer at an adult whose back was turned to an object who *correctly* labeled it (compared to when labeling was incorrect). Children of this age apparently expected accuracy from an adult who was facing the display and offered referential cues (i.e., eye gaze toward the display during labeling), but not when such cues were missing (from an adult facing away from the display). Children’s attention to an audio speaker did not differ whether a correct or an incorrect object label came out of it, suggesting that children had no expectations of correct reference from a human voice in the absence of a person providing referential cues.

Reliance on the presence of social partners and the cues they provide may at least partially explain why very young children seem not to learn efficiently from video, a source of information that is becoming increasingly pervasive in infants' and toddlers' lives (Rideout & Hamel, 2006). This result, dubbed the "video deficit" (Anderson & Pempek, 2005) has been found at several ages and across various learning tasks. For example, at an age when infants' speech perception is narrowing toward their native language, 9-month-olds maintained the ability to discriminate non-native (Mandarin Chinese) speech sounds after a Mandarin speaker interacted with them face-to-face for 5 hours across the course of a month, but other 9-month-olds lost this ability after watching the same speaker on video for the same amount of time (Kuhl, Tsao, & Liu, 2003). In other research, 12- to 30-month-olds imitated the actions of a person who was present more often than they imitated the same person appearing on a TV screen, even though she made apparent eye contact and offered attention-directing comments (e.g., "Look at this!") in both cases (Barr & Hayne, 1999; Hayne, Herbert, & Simcock, 2003). Similarly, 24-month-olds followed the instructions of a person who was present, using this information to find a hidden toy, but were only a third as likely to use the same information offered by the same person on video (Evans Schmidt, Crawley-Davis, & Anderson, 2007; Troseth, Saylor, & Archer, 2006). Although 30-month-olds (compared to younger children) showed some improvement in using information from people on video in search tasks (Schmitt & Anderson, 2002; Troseth & DeLoache, 1998), children of this age still did better when instructed by a person who was present (Schmitt & Anderson, 2002). In a difficult labeling task that demanded reliance on referential social cues, 24- and 30-month-olds learned a word when an adult who was present gazed into an opaque container while offering the novel label, ignoring a visible distracter; in contrast, toddlers of the same age did not learn the word from a person on video who offered the same cues (Troseth, Saylor, & Strouse, 2009; also see Krcmar, Grela, & Lin, 2007).

In all of these studies, a person on video offered particular social cues, such as apparent eye contact with the viewing child and infant-directed language. However, other aspects of a normal social situation were missing. For instance, the person on television did not engage in contingent, reciprocal interaction with the child—a characteristic of social engagement to which infants are sensitive from the middle of the first year (Bigelow, MacLean, & MacDonald, 1996; Hains & Muir, 1996). Additionally, the person on screen did not share attention with the viewing child to objects present in the child's environment, as usually occurs in a triadic interaction. TV watching may not have met toddlers' expectations for a social, "pedagogical" situation in which information relevant to the child is being transmitted. An exception is a study in which a person on video provided evidence of engagement and relevance by conversing with the child's parent via 2-way closed-circuit video while the child watched. The person on TV then played "Simon Says" with the child, talked about and discussed an item *in the child's environment* (a sticker on the child's shirt), and responded contingently to whatever the child and parent said and did. After interacting with the person on video for 5 minutes, 24-month-olds learned efficiently from her (Troseth et al., 2006; also see Nielsen, Simcock, & Jenkins, 2008).

Do children need to be involved in a contingent, reciprocal interaction with another person in order to learn from that person? Young children also learn as *onlookers* to third-party interactions (Akhtar, 2005). Children as young as 18 months used an adult's referential behavior to learn words while "overhearing" an exchange between the speaker and another person who also was present in the room, learning as well as children who were directly addressed by the speaker (Akhtar, Jipson, & Callanan, 2001; Floor & Akhtar, 2006). It is important to note that the speaker's cues were not directed to the "overhearing" child; the speaker made eye contact and interacted only with the adult confederate while stating her intention to show that person a named object. She removed the item from a container, held it up with a gasp of pleasure, demonstrated its function, and handed it over to the other adult

before returning it to the container. No social behavior was directed at the children, who nevertheless learned the word as easily as children who were directly addressed.

In the current study, we examined whether 30-month-olds learn from a person on video in a commercially-realistic format (i.e., one that is possible on commercial television) in which expected social cues are offered, *but are directed toward another adult on the screen*. We used Akhtar's word learning task to see if toddlers' ability to learn as onlookers to a *live interaction* would extend to learning as onlookers to a *televised interaction*. We chose to test 30-month-olds because toddlers of this age did not reliably learn words using the referential cues of a person on video in an earlier study (Troseth et al., 2009). However, we reasoned that toddlers might learn words watching an intact interaction between two people on video yet not learn from a person on video who apparently addressed them directly but who was unresponsive to them. Based on previous research, we expected that children would learn from someone present in their environment whether they were onlookers to a conversation or were directly addressed.

## Study 1

### Method

**Participants**—Sixty-four children (31 boys) participated, ranging in age from 27.0 to 31.7 months ( $M = 29.8$   $SD = 0.9$ ). Half of the participants (15 boys), from a city in the southeast United States, were recruited from state birth records. The rest of the participants (16 boys) from a community on the U. S. west coast, were recruited from a database of children whose parents had expressed interest in being included in studies of child development. In both of the studies reported here, participants were mostly of European American descent and were native English speakers. Across both studies, primary caregiver education ranged from high school diploma to a doctoral/professional degree (85% had a college degree or above) and family income from \$30,000 to over \$100,000 a year (79% earned \$50,000 or above).

Children were randomly assigned to one of four conditions: *Live Addressed* ( $M = 29.7$  months, 7 boys) *Live Onlooker* ( $M = 29.9$  months, 8 boys) *Video Addressed* ( $M = 29.6$  months, 8 boys), and *Video Onlooker* ( $M = 29.7$  months, 8 boys). Data from 12 children were excluded from analysis for uncooperativeness (6), English not being the child's primary language (2), parental interference (1), suspected developmental delay (1) and experimenter error (2).

**Materials**—Four familiar items (e.g., plastic horse, banana, turtle, and truck) were used for a warm up comprehension task. Four distinctive wooden toys with movable parts were used as novel test objects. All of these objects afforded interesting actions that 2-year-olds could perform. Four different-colored opaque jars with screw-on lids, placed in a row, and attached to a 36" (91 cm) wood plank were used as the hiding apparatus for the objects. During the warm-up and testing phases, participants sat at a small table across from an experimenter while parents sat on a nearby couch. During the labeling of the novel objects, children sat on the couch or on a small chair, facing the researcher(s) in the live conditions and a 27" or 32" (69 or 81 cm, depending on location) television set that displayed a pre-recorded video of the researcher(s) in the video conditions. One video camera filmed the entire experimental setup, including child and researchers/TV. A second camera was directed at the participant during the labeling phase; the resulting video was used to code children's attentiveness.

**Procedure**—Testing in both locations took place in small laboratory playrooms and lasted approximately 30 minutes. The researcher explained the study to parents, obtained consent, and then asked parents to complete the short form of the MacArthur Communicative

Development Inventories (Level II) to assess their child's expressive vocabulary (Fenson, Pethick, Renda, Cox, Dale, & Reznick, 2000). The researcher also confirmed that none of the children were familiar with any of the four novel objects.

**Warm up:** After a brief warm up activity, the researcher introduced a showing game and comprehension test using the four familiar items. After placing the hiding apparatus on the table, the researcher said, "I'm going to show you what's in here", removed the lid from the first container, and pulled out the familiar item inside. She then handed the item to the child, allowed the child to handle it briefly, replaced this object in the container and continued until all four containers had been opened and their contents examined. After introducing the four familiar items, the researcher placed all the items in a tray in front of the participant and asked the child to choose one (e.g., "Can you pick the horse?"). After the child correctly chose the requested item, the experimenter replaced it and asked the child to choose a different object, until the child had correctly picked two items in a row, indicating they understood the instructions.

Next, the researcher placed the hiding apparatus, now filled with the four novel objects, on the table and said, "I'm going to show you what's in here. Let's see what's in here. I'm going to show you this one". She then removed the lid from the first jar, pulled out the novel item inside, handed the item to the child and then asked the child to return it to the jar. These steps were repeated for the remaining containers until the child had been familiarized with all four novel objects. No labeling occurred during this warm up. The warm-up familiarized the children with the novel objects and the apparatus, and confirmed that they understood what they were being asked to do in the comprehension task.

**Labeling:** In the *live* conditions, the child sat on the couch approximately 3–4 feet (1 m) from the researcher who was directly facing the child (*Live Addressed*) or an adult confederate (*Live Onlooker*). In the *video* conditions, the child sat approximately the same distance from the television set which showed a pre-recorded video of the researcher facing the camera (*Video Addressed*) or an adult confederate (*Video Onlooker*) (see Figure 1).

In all four conditions, the researcher sat at a small table that held the containers. Each of the novel objects was always placed in a particular container, but which object was the target was counterbalanced across children. Prior to introducing the target object, the researcher stated its label three times: "I'm going to show you the toma. Let's see the toma. I'm going to find the toma". Prior to introducing each of the distracter objects the researcher made three non-labeling statements: "I'm going to show you this one. Let's see this one. I'm going to find this one".

In the *Live Addressed* and *Video Addressed* conditions, all words and actions were directed at the child: the researcher made eye contact with the child (or the camera) and uttered the three statements, then opened the container, looked inside while gasping and smiling, pulled out the object, performed a distinctive action on it (e.g., shaking it; rocking it from side to side) for approximately 5 seconds, replaced the object in the container and, moved on to the next one.

In the *Live Onlooker* and *Video Onlooker* conditions, the same words and actions were directed at the confederate instead of the child. The speaker performed the distinctive action on the object for approximately 2 seconds, and then handed the object to the confederate who imitated the action for approximately 3 seconds before handing the object back to the speaker. The *Live Onlooker* condition was the same as that used previously by Akhtar et al. (2001). The *Live Addressed* condition was the same as Akhtar et al.'s *Addressed* condition except that the researcher did not hand the toys to the child (this was done to equate the *Live*

and *Video* conditions). Across all conditions, each object was out of its container and visible to the child for approximately 5 seconds (range = 2 to 10 seconds). In no condition did the child touch the toys during the labeling phase.

In all conditions, the researcher proceeded from right to left through each of the four containers in the apparatus, repeating the process for a total of three times. Thus, children heard the target object labeled the “toma” a total of nine times. This demonstration took approximately three minutes. Once labeling began, the researcher/video was not stopped for any reason (e.g., if the child got up or turned away from the demonstration).

**Testing:** Children sat across from the researcher at the small table and were allowed to handle the novel objects (without labeling) for approximately 30 seconds. The objects were then put into a tray, which the researcher shook up and placed in front of the child. The researcher looked up from the tray, made eye contact with the child, and asked a comprehension question, “Which one is the toma?” as well as a preference question, “Which one is your favorite?” Children’s answers to the preference question helped clarify whether or not their object selection on the comprehension question was meaningful; if the children picked the target as their favorite at a rate greater than chance, mere preference might be motivating their toy choice on the other question. To ensure that children did not think that the two questions were the same, there was a minimum of 30 seconds between questions in which the researcher asked about the child’s outfit, siblings, etc. If the child did not pick an item (or picked several), the researcher asked again until the child chose one object. If children were distracted and would not answer after several repeated attempts, the researcher took a short break and then asked again; the average number of times she repeated the comprehension and preference question was 3 (max = 8 and 10, respectively). The order of questions (comprehension and preference) was counterbalanced across children; there was no effect of question order on children’s word learning.

**Attention:** During the demonstration, the speaker named the target object nine times and manipulated it three times. To learn a novel word, children needed to attend to at least part of the demonstration. As a measure of attentiveness, the proportion of time children spent looking at the researcher(s) or video during the labeling session was assessed for all of the videotaped sessions by one coder. An additional coder independently recorded children’s attentiveness for 25% of the videotaped sessions (4 participants from each condition). Inter-rater reliability was high ( $\rho_I = .96, p < .001$ ). Another coder, blind to hypotheses, also coded 25% of the videotaped sessions; reliability with the main coder again was high ( $\rho_I = .96, p < .001$ ). Two participants were excluded from attention analyses because of technical problems with their videotapes.

**Learning:** An assistant recorded the participant’s answer to the comprehension and preference questions during the session. An independent coder who was blind to the hypotheses viewed all of the videotaped sessions and recorded the child’s choices for the comprehension and preference questions; agreement between the two raters was 90% ( $\kappa = .87$ ) for comprehension and 95% ( $\kappa = .96$ ) for preference responses. We noted which children chose the target in response to the comprehension question only, to the preference question only, to both questions, or to neither question.

## Results and Discussion

**Vocabulary**—In preliminary analyses, we examined children’s raw scores on the MacArthur CDI. Girls’ scores ( $M = 90.2, SD = 16.9$ ) were significantly higher than boys’ scores ( $M = 79.1, SD = 20.8$ );  $t(62) = 2.35, p = .02, d = .59$ , but there were no significant differences by age or testing location. In a one-way analysis of variance, there were no

significant differences in vocabulary scores across the four conditions,  $F(3, 60) = .046$ ,  $p = .99$ ,  $\eta^2 = .002$ . Furthermore, parental reports of vocabulary did not correlate with learning,  $r = .008$ ,  $n = 64$ ,  $p = .94$ .

**Attention**—In preliminary tests of children's attentiveness, there were no effects of gender, age, or testing location. A two-way between groups analysis of variance exploring the effects of medium (*Live* vs. *Video*) and conversation type (*Addressed* vs. *Onlooker*) on children's attentiveness revealed a significant main effect of medium,  $F(1, 58) = 6.12$ ,  $p = .02$ ,  $\eta^2 = .095$ . Children in the *Live* conditions spent a higher proportion of time ( $M = 92\%$ ,  $SD = .07$ ) looking at the demonstration than children in the *Video* conditions ( $M = 84\%$ ,  $SD = .15$ ). There was no main effect of conversation type,  $F(1, 58) = .379$ ,  $p = .54$ ,  $\eta^2 = .006$ : children's attentiveness in the *Addressed* and the *Onlooker* conditions was equivalent. There also was no interaction between medium and conversation type,  $F(1, 58) = 1.59$ ,  $p = .21$ ,  $\eta^2 = .027$ . During the demonstration, a few children did get up from their seats/ turn away at least once (4 in *Live Addressed*, 3 in *Video Addressed*, and 1 each in *Live* and *Video Onlooker*). However, all participants watched at least a portion of the demonstration ( $M = 88\%$ , range: 40% to 100% of the time) and no participant was dropped from the study for not attending to the demonstration at all.

**Learning**—In preliminary tests, there were no effects of gender, age, or testing location, on children's learning. For each condition, we ran a binomial test to compare the number of children who chose the target object for the comprehension question against the number expected by chance (25%). Children in both *Onlooker* conditions chose the target at rates that exceeded chance (*Live Onlooker*, 10 of 16 children and *Video Onlooker*, 10 of 16 children, both  $ps = .002$ ). In contrast, in the two *Addressed* conditions the number of children who chose the target did not significantly differ from chance (*Live Addressed*, 7 of 16 children and *Video Addressed*, 6 of 16 children,  $ps = .08$  and  $.19$ ).

Children's responses to the *preference* question clarified the meaning of their choices on the comprehension question. If children simply happened to like the target object and did not actually learn the novel label, they would demonstrate preference of the target at a rate above chance in binomial tests. However, the number of children who chose the target in response to the preference question was not significantly different from chance in any of the four conditions (*Live Onlooker* = 3 of 16,  $p = .41$ ; *Video Onlooker* = 3 of 16,  $p = .41$ ; *Live Addressed* = 5 of 16,  $p = .37$ ; and *Video Addressed* = 4 of 16,  $p = .63$ ).

To further compare children's selection of the target object in comprehension and preference trials, we conducted sign tests (see Floor & Akhtar, 2006), which can be used to compare the likelihood of patterns of answers on two trials occurring by chance. There were four possible patterns of responses in our data: choosing the target for 1) comprehension but not preference; 2) preference but not comprehension; 3) both; or 4) neither. Table 1 shows the number of children who exhibited each pattern across the four conditions. The sign test eliminates pairs when there is no difference between the two trials (our "both" and "neither" patterns). As shown in the middle column of Table 1, in both *Onlooker* conditions, the number of children who chose the target in response to comprehension but not preference questions was above chance, whereas the opposite response pattern was not. In both *Addressed* conditions neither pattern was above chance.

Next, we ran binomial tests in which the children who chose the target for *both* comprehension and preference questions were not credited with learning. We wanted to see if the number of children who chose the target for *comprehension only* still differed from chance. Children's choices of the target for both *Onlooker* conditions were still significantly above chance (9 of 16 in *Live Onlooker* and 8 of 16 in *Video Onlooker*,  $ps = .001$  and  $.005$ ,

respectively), and children's choices in both *Addressed* conditions still did not differ from chance (4 of 16 in *Live Addressed* and 4 of 16 in *Video Addressed*,  $ps = .35$ ). Thus, the pattern of results remained the same when we eliminated those children who also happened to prefer the target.

Therefore, children who observed the speaker addressing another adult and labeling the novel object (either "in person" or on video) did learn the word. Children who were directly addressed by a person who was present or on video did not. Although attention was higher for the live events than for the video events, this did not predict the pattern of learning and there was no correlation between the amount of time children looked at the demonstration and learning,  $r = -.154$ ,  $n = 62$ ,  $p = .23$ .

Based on previous research (e.g., Akhtar, 2005; Akhtar et al., 2001) we expected children in the *Live Addressed* condition to learn the novel label. In Akhtar's previous studies, the researcher handed the children in the *Addressed* condition each object after talking about it (in the *Onlooker* condition, she handed the object to the adult confederate). In the present study, in order to match our *Video Addressed* condition, the speaker in the *Live Addressed* condition did not hand the child the toy after discussing it. We assume that a triadic interaction between a toddler and adult regarding new toys—especially one that involved multiple comments on and displays of each object—typically would include offering the child the objects of interest. In addition, the researcher had offered the child the toys when she was introducing the apparatus and finding game during the warm-up. In our *Live Addressed* condition, we may have violated toddlers' expectations of what a social interaction should include. Behaviors such as reaching for the objects or approaching the demonstration may indicate children's expectation to be a part of the interaction. Eight of 16 children in the *Live Addressed* condition reached for the objects and/or approached the demonstration, whereas only 2 of 16 in the *Live Onlooker* condition, and none of the children in either video condition, did so. A chi-square test showed that the difference in reaching/approaching behaviors between *Live Addressed* and *Live Onlooker* was significant,  $\chi^2(1, N = 32) = 5.24$ ,  $p = .022$ ,  $\phi = -.41$ . *Handing* the toys immediately after labeling clearly was not necessary for learning (the children in the *Onlooker* conditions were not handed the toys, yet they learned the word). However, *reciprocal engagement* between social partners (as evidenced by the *handing over* of toys) may have been vital. We hypothesize that unfulfilled expectations regarding normal social interaction hindered children's learning in the *Live Addressed* condition. To test this hypothesis, in Study 2A, a group of children participated in a *Modified Live Addressed* condition in which they were handed each object after labeling and thus clearly were included as part of the interaction.

## Study 2A

### Method

**Participants**—Sixteen children (8 boys) participated, drawn equally from the same two communities and populations, and recruited in the same way as in Study 1. Participants ranged in age from 28.5 to 31.6 months ( $M = 30.2$ ;  $SD = .82$ ).

**Materials and Procedure**—The same materials were used as in Study 1. The procedure was the same as in Study 1, *Live Addressed* condition, with some slight variations. During the *Labeling* phase, the researcher and child sat across the small table from each other (slightly closer than in Study 1, to allow the toys to be passed from researcher to child). Different from Study 1, after the speaker pulled each object from the container and performed an action on it, she gave the object to the child to handle briefly before replacing the object.



## Results and Discussion

A binomial test indicated that the number of children who chose the target object in response to the comprehension question (chance = 25%) was significantly above chance (12 of 16,  $p = .000$ ) whereas the number who chose the target object as the one they preferred (4 of 16,  $p = .63$ ) was not different from chance. As predicted, when an adult addressed children and then handed the toys to them, they learned the novel word. We also ran a sign test to assess individual patterns of performance. Nine of 16 participants chose the target for comprehension but not preference, whereas 1 showed the opposite pattern ( $p = .011$ ). Three children chose the target for both comprehension and preference questions; in a binomial test, the number of participants who chose the target for comprehension but not preference (9 of 16) is still significantly above chance ( $p = .001$ ).

Retrieving the toys from the children took longer than retrieving them from the adult confederate in the Overhearing conditions. Children in the *Modified Live Addressed* condition therefore were *exposed* to the target object for an average of 9 seconds--longer than the average amount of exposure in the *Live Addressed* condition of Study 1 (4 seconds). Within each condition, the average exposure to *target and non-target* objects was the same (*Live Onlooker* = 5 seconds, *Video Onlooker* = 7 seconds, *Video Addressed* = 6 seconds). Importantly, exposure time to target (range = 4 to 23 seconds) was not correlated with children's word learning in the *Modified Live Addressed* condition (Spearman rank order correlation,  $r = -.07$ ,  $n = 16$ ,  $p = 0.8$ ) or across all Study 1 conditions,  $r = .16$ ,  $n = 64$ ,  $p = .12$ .

There also was no correlation between the length of time children *handled* the target ( $M = 5$  seconds, range = 0 to 17 seconds, both for the target and non-target objects) and learning,  $r = -.38$ ,  $n = 16$ ,  $p = .15$ . Therefore, it is unlikely that differences in exposure time (or handling time, for the *Modified Live Addressed* condition) can explain the differences in learning among the conditions.

When we incorporated handing the object to children into our original *Live Addressed* procedure, children learned. During the warm-up to the showing game, children in all conditions were handed the novel objects, which may have raised their expectations that this was how the game was played. Sharing objects clearly indicates a reciprocal engagement with the social partner (whether one directly benefits from such sharing, as occurred in Study 2A, or merely observes it, as in the *Onlooker* conditions of Study 1). When such evidence of reciprocal engagement was missing (in Study 1's *Addressed* conditions), toddlers may not have recognized that the speaker had shared information. Simply *handling* the objects during the demonstration was unlikely to be the cause of learning; children in Study 1's *Onlooker* conditions never touched the toys during the demonstration (they watched an engaged recipient do so) yet learned, and the amount of time children were exposed to the target (in any condition) or handled it (in the *Modified Live Addressed* condition) was unrelated to learning.

In Study 2B, we further explored the effect of obvious evidence of reciprocal engagement by testing whether or not toddlers would learn a novel word as onlookers to a "live" situation in which the speaker did not hand the toy to the confederate. To do so, we modified the *Live Onlooker* condition used in Study 1 and in Akhtar's earlier research.

## Study 2B

### Method

**Participants**—Sixteen children (8 boys) participated, drawn equally from the same two communities and populations, and recruited in the same way as in Study 1. Participants ranged in age from 29.1 to 31.5 months ( $M = 30.6$ ,  $SD = .64$ ).

**Materials and Procedure**—For this *Modified Live Onlooker* condition, the same materials were used as in Study 1. The procedure was the same as in Study 1's *Live Onlooker* condition, with some slight variations. Similar to Study 1, the speaker sat across from the adult confederate while the child looked on. Different from Study 1, after the speaker pulled each object from the container and performed an action on it, she did *not* hand the object to the confederate to handle before replacing it. However, similar to Study 1 the speaker directed eye contact to the confederate and held up the objects, and the confederate indicated she was attending to the demonstration by making eye contact with the speaker and looking at the objects.

### Results and Discussion

The number of children who chose the target in response to the comprehension question was not significantly above the chance level of 25% in a binomial test (7 of 16,  $p = .08$ ) nor was the number of children who preferred the target (4 of 16,  $p = .63$ ). Therefore, children failed to learn a novel word as an onlooker to an interaction in which the speaker addressed another adult but *did not hand over* the object of interest. We also ran a sign test to assess the individual patterns of performance. Four of 16 participants chose the target for comprehension but not preference whereas 1 showed the opposite pattern ( $p = .188$ ). Three children chose the target for both the comprehension and preference trials; in a binomial test, the 4 of 16 who chose the target for comprehension but not preference was still not significantly above chance ( $p = .63$ ).

In Study 2B, children were exposed to the target object an average of 4 seconds (range = 3 to 6 seconds). As in all of the other conditions, the amount of exposure time to the target was not correlated with children's learning,  $r = -.11$ ,  $n = 16$ ,  $p = .69$ .

The results of Study 2B suggest that toddlers need significant evidence that a *reciprocal interaction* is taking place to learn a word by observing a speaker talking to someone else. In the *Onlooker* conditions of Study 1, the speaker handed the object to the confederate, who demonstrated that she was part of the interaction by accepting the object and imitating the action the speaker had demonstrated. Results of Study 2B indicate that evidence of sharing on the part of the "teacher" and engagement by the "learner" may be needed for toddlers to learn a novel word from a third-party interaction.

### General Discussion

The results reported here demonstrate that 30-month-olds can learn a novel word as an onlooker to a conversation between two people on video as well as they learn from a live third-party conversation in their environment. Study 1 expands on previous research by providing evidence of a learning situation involving video in which toddlers do not contingently interact with the person on screen, yet learn as well as they do from overhearing those who are actually present. Consistent with previous research, we found that toddlers did not reliably learn a new word after being directly addressed by a person on a pre-taped video. These results shed light on the kind of situation on video that may be most informative for toddlers' learning. Given that third-party interactions can easily be shown on

commercial television (unlike actual contingent interactions with the viewing child), this result may have implications for the design of educational programs for toddlers.

An unexpected result in Study 1 was that toddlers did not reliably learn a new word after being directly addressed by a person who was present. Results from Study 2A indicate that this unexpected finding may be due to the situation not truly being a shared *interaction*: children did learn the word when directly addressed by someone present in their environment who showed engagement by handing them the objects. Results from Study 2B indicate that evidence of engagement by both parties also is important for children's learning when they are onlookers; that is, children did not learn the word when they watched one adult label and demonstrate an object for another adult, who served as a passive observer. This research therefore deepens our understanding of the conditions under which children learn from third-party interactions.

In the current studies, toddlers learned a novel name for an object as onlookers to a conversation on video, onlookers to a conversation present in their environment, and when directly addressed by someone in their environment who offered them the objects of interest. What signaled to children that each of these scenarios was a learning situation? We believe the key element of these three contexts is *reciprocal interaction*. That is, children in these situations were either onlookers to knowledge exchanged via social interaction or actively took part in a reciprocal interaction. Specifically, in the two overhearing conditions, the confederate watched intently as the speaker talked about the object to be revealed, then followed the speaker's gaze as she looked into each container and extracted an object. The speaker held each object up and manipulated it, then handed it to the confederate who briefly manipulated the object in the same way. This scene may have visibly demonstrated to onlooking children the occurrence of teaching (and learning). In the *Modified Live Addressed* condition (Study 2), the speaker handed each object to the viewing child after manipulating it, thereby allowing the child to participate in a shared interaction. In the three conditions in which children did not learn (*Video* and *Live Addressed, Modified Live Onlooker*) children did not take part in or observe a reciprocal interaction; they were mere observers of an extended, one-sided offering of information.

By 30 months of age, children are becoming discerning consumers of socially transmitted knowledge. Toddlers are sensitive to typical components of social "scripts" (both as observers and as participants), and in the absence of expected components may fail to recognize a learning opportunity (see Nelson, 1986). We hypothesize that toddlers are especially attuned to reciprocal social interaction, including behaviors exhibited by both actor and recipient in social games (Carpenter et al., 2005). By 12 to 15 months of age, infants have begun to take an active role in collaborative learning (Tomasello et al., 2005). For instance, in a study examining 15- and 18-month-old infants' behavior during adult-initiated interruptions in social game playing, 60% of infants' actions were interpreted as communicative attempts to reengage the adult in the game (Ross & Lollis, 1987). In addition to taking an active role in social learning, older infants apparently count on others to play their role in a typical manner (displaying expected social cues). For example, 14-month-olds did not imitate a demonstrated action in an "incidental learning context" where the actor did not provide any communicative cues (Király, Csibra, & Gergely, 2004). Nielsen (2006) found that 18-month-olds were significantly more likely to imitate an adult's exact actions when she demonstrated engagement (i.e., played with the child in the warm up and made eye contact and smiled during the demonstration, alternating gaze between toy and child) than when she acted aloof (no prior playing, no eye contact, no smiling), whereas 24-month-olds were equally likely to imitate regardless of the adult's demeanor. Nielsen hypothesized that in imitating the person who was engaged with them, children of both ages showed they wanted to sustain interaction; additionally, the older children may have been attempting to

*initiate* interaction by imitating the aloof actor. Toddlers' strong motivation to attend to and engage in reciprocal interactions therefore may facilitate their learning.

In the *Addressed (Live and Video)* conditions of Study 1, following a warm-up in which children were allowed to touch each item after it was removed from its container, children were placed in the role of observer of a one-sided labeling demonstration that went on for 3 minutes. Although the adult carrying out the demonstration made eye contact with and directed remarks to children (following a script), she may not have seemed fully engaged with them, given their prior experience in the game during warm up. Procedural differences may explain why in other research (e.g., studies of early imitation) infants and toddlers have learned after merely observing an individual's behavior with a novel toy (e.g., Barr, Dowden, & Hayne, 1996; Meltzoff, 1985). As part of the imitation procedure in these studies, children typically are not given access to the toy before the novel behaviors are modeled (thus, expectations are not raised that toys will be handed over), and demonstrations are very short (20 to 60 seconds). In another type of study, infants and toddlers viewed an adult's communicative cues and learned the location of a hidden toy (Behne, Carpenter, & Tomasello, 2005). However, children did not need to wait an extended period of time for "their turn" to find the toy, which immediately followed a very brief presentation. In contrast, after handing over the toys in the warm-up with the apparatus, the researcher in the current studies discussed, revealed, and manipulated each of the four objects three times while directly addressing children; thus, in the *Addressed* conditions in Study 1, children sat through 12 object "showings" before it was their turn to play. Toddlers in the *Modified Live Addressed* condition, who remained involved by being handed each toy right away (as had occurred in the warm-up), learned the novel word. In the *Onlooker* conditions of Study 1, toddlers who watched the speaker maintain engagement with the listener throughout the 12 object "showings" (by handing her the toys) also learned the word.

It is important to emphasize that children did not need to handle an object in order to learn; children learned the novel word in both *Onlooker* conditions in which they did not handle the toys at all during the demonstration. Additionally, in the condition in which children were handed the toy (Study 2A), learning was not related to how long children handled the target object. Rather, it apparently was the *handing over* of the toys, serving as evidence of the speaker's engagement with the partner (the child or the confederate), that was the crucial difference.

Akhtar and colleagues' research (and the research reported here) indicates that toddlers do not need to be an active part of a social interaction to view it as a learning situation; they learn as an onlooker to a conversation as well as when they are directly addressed. In Akhtar's paradigm (and our own) the confederate in the *Onlooker* learning situation is an important part of the interaction; she handles each novel object and demonstrates that she has learned to perform an action on it. In a recent study, Herold and Akhtar (2008) hypothesized that the ability to learn as an onlooker relies on understanding self-other equivalence (Moore, 2007), or what Meltzoff (2005) describes as seeing others as "like me", because it allows children to imagine themselves as part of the interaction. Herold and Akhtar assessed 18- to 20-month-olds' self-recognition and their ability to take another person's perspective. Both factors predicted children's imitation of novel behaviors that the children learned as onlookers to a "third-party interaction". We hypothesize that children learn as onlookers (both in this study and from the real interactions they observe daily) because they recognize that knowledge is being transferred via social interaction and possibly imagine themselves as part of the exchange. In contrast, Study 2B indicates that toddlers do not learn as onlookers to a situation in which the knowledge recipient is a passive observer of an extended, one-sided demonstration. Toddlers in the current studies

needed some indication that a reciprocal interaction was taking place (such as the handing over of the object of interest) in order to learn the information presented.

In the research reported here, toddlers exhibited no “video deficit” in learning after observing a reciprocal social interaction on video, in contrast to many other studies in which learning from video was depressed compared to learning from an equivalent “live” event (e.g., Barr & Hayne, 1999; Deocampo & Hudson, 2005; Evans Schmidt et al., 2007; Hayne et al., 2003; Kuhl et al., 2003; Schmitt & Anderson, 2002; Strouse & Troseth, 2008; Troseth et al., 2006). In a recent word learning study, 24- and 30-month-olds failed to use subtle referential social cues (e.g., the labeler’s gaze into an opaque bucket containing a target object, in the presence of a visible distracter) to learn a word from a person on a video compared to a person who was present (Troseth et al., 2009). Even when a person’s social cues were straightforward (gaze toward one of two visible toys), toddlers more often learned a word on trials when a person was present compared to on video trials (Krcmar et al., 2007). Additionally, 30- to 35-month-old children did not learn verbs after repeatedly watching an event on video narrated in a voiceover using infant-directed speech; they did learn when the first two video demonstrations were replaced with live social interaction (an adult using a doll or puppet to demonstrate and label the action—Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009). Only children over age 3 learned the verbs from video alone.

Given that 18- to 30-month-old infants have learned from the conversations and reciprocal behaviors of people who are present in the environment (Floor & Akhtar, 2006, Akhtar et al., 2001), we would predict that they also could learn from videos of 3<sup>rd</sup> party interactions. In the word learning study described above that was carried out with 24- to 30-month-olds, learning was basically the same in both age groups (Troseth et al., 2009). Therefore, we assume that children younger than 30 months of age would also learn when observing the conversations of others on video.

The presence of an engaged recipient of the social interaction, as in the current research, may have promoted toddlers’ awareness of the relevance of referential social cues presented on video. Kuhl (2007; Kuhl et al., 2003) points out that a basis for infants’ language acquisition may be seeing a person’s social cues as referential; social cues that are informative when the speaker is present in the environment and directing attention to real objects may not seem referential coming from a non-contingent person on video directing his or her gaze/points/actions at 2-dimensional images of objects on the screen. In the present study, when both parties to the interaction were together on screen and an obvious exchange of information took place, toddlers appeared to treat this scene as an opportunity to learn.

Nielsen et al. (2008) suggest ways in which social interaction affects learning from video. In their study, a modeler demonstrated an arbitrary and somewhat awkward action (using a stick to press a switch to open a box, rather than using hands). Toddlers were more likely to imitate the exact behaviors of a person who was present and responsive than a person on a pre-taped video. In a follow-up study, children were more likely to imitate the exact actions of a modeler who had been contingently responsive to them via closed-circuit video; in contrast, when the same modeler on a pre-taped video was non-responsive, toddlers tended to open the box with their hands. Nielsen et al. reason that social interaction affected imitation because children viewed the responsive person on video as a social partner with whom they could affiliate (i.e., forge an interpersonal bond and sustain interaction—Meltzoff & Moore, 2002; Uzgiris, 1981). In the current study, toddlers observed an interaction on video between a teacher and a learner who copied the teacher’s exact actions, sustaining the interaction. The teacher treated the learner as she had the child during the

warm-up, handing over each toy to her. The viewing children may have learned by establishing “like me” correspondence, putting themselves in the place of the responsive learner on the screen (Herold & Akhtar, 2008; Meltzoff, 2005).

Referential cues help children determine the intended target in situations where the object is not visible during labeling (as in the current studies) or when there is more than one possible referent for an utterance (Baldwin, 1993; Baldwin & Moses, 1996). For toddlers, the offering of such cues by a person on video, in the absence of reciprocal interaction, usually was not sufficient for them to learn words. However, there is a second way that a video can direct attention to a labeled object: by presenting a *close-up* of the target (thus eliminating other potential referents) accompanied by a voiceover. Voices draw children’s attention to a TV screen; preschoolers playing with toys in the presence of a television look at the screen at “noisy” moments such as when a woman speaks (Alwitt, Anderson, Lorch, & Levin, 1980). Research on word learning takes advantage of this fact: in the Intermodal Preferential Looking Paradigm (e.g., Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008), participants hear a label while watching an object/action displayed on a screen. After repeated pairings of the word and video, children are presented with the label accompanied by the old video and a new one showing a different object or action. Very young children tend to look to the matching video (e.g., Scofield, Williams, & Behrend, 2007; Maguire et al., 2008; Werker, Cohen, Lloyd, Casasola, & Stager, 1998), showing that they can learn associations between words and what they see on screen. Werker and colleagues (also Roseberry et al., 2009) express the need for caution in describing such matching as “word learning”, reserving the term for situations in which children respond appropriately to a request (e.g., hand a questioner the object labeled with the novel word) or extend the label to other instances of the same category. At minimum, word learning from video would seem to require extending the label to the real object depicted on the screen (see Allen & Scofield, 2008).

In another word learning study, 27-month-olds learned novel action verbs that were uttered in a voiceover while a person on television demonstrated the actions while exhibiting behavioral cues to her intentions (Poulin-Dubois & Forbes, 2002). The two similar actions (e.g., “tiptoe” and “circle”) differed only in manner. Children needed to distinguish the demonstrator’s manner to attach the novel label to an action. Children of this age were able to do so, mapping the word from the voiceover to the target action during test trials. This study differs from the current one in the kind of intentional cues that were available. Although the demonstrator’s intent could be read in her manner of acting, she did not offer *referential social* cues (such as looking at the viewer, and then gazing into a bucket while providing a label). The actor on video did not offer *social* cues at all; she did not attempt to communicate, although the intentions underlying her actions could be read by an aware observer. Because the voiceover co-occurred with the actor’s behavior, the children may have inferred the unseen labeler’s communicative intent, or merely associated the word with the action.

Some children’s television programs (e.g., *Blue’s Clues*, *Dora the Explorer*) attempt to provide social cues, even a kind of contingency, by having the character on the screen ask questions and leave pauses for the child to answer. Research to date (involving preschoolers older than age 3) indicates that even though a person on television is not actually responsive, repeated viewing leads to more interaction on the part of the child, as well as increased comprehension of program content (Anderson, Bryant, Wilder, Santomero, Williams, & Crawley, 2000; Crawley, Anderson, Wilder, Williams, & Santomero, 1999). By the age of 3, of course, preschoolers have been shown to learn words and other information from video (e.g., Rice, Huston, Truglio, & Wright, 1990; Rice & Woodsmall, 1988), possibly because they have begun to see videos (as well as pictures and scale models) as representations that can convey information (DeLoache, 2002; Troseth & DeLoache, 1998). We are currently

conducting research to determine whether infants and toddlers also treat watching so-called “interactive” video as a learning situation.

By 30 months of age, toddlers have established scripts for the ways adults typically interact with them and teach them new information. They therefore have expectations about what learning situations are like, and may be more likely to learn in contexts that match their expectations. Although scripts for pedagogical interactions may vary across cultural contexts (Rogoff, 2003), the children we studied likely are used to teaching contexts that involve reciprocal (contingent) interactions with eye contact--expectations that cannot be met by a “talking head” on television who is unresponsive to the viewer. However, watching two other people share eye contact and respond to each other seems to meet children’s script for a learning situation. Children appear to put themselves into the place of the learner, viewing the recipient of this type of teaching interaction as similar to themselves, and therefore learning the information being offered (Herold & Akhtar, 2008; Meltzoff, 2005; Moore, 2007). Observing reciprocal social interactions--both those observed directly and those appearing on video--seems to be an effective way for toddlers to learn words.

## Acknowledgments

This research was supported by Vanderbilt University, by a Faculty Research Grant from the University of California, Santa Cruz Senate Committee on Research to Nameera Akhtar, by an NIH postdoctoral training grant #T32 HD046423-10 to Priya M. Shimpi and by NICHD grant # P30 HD15052 to the Kennedy Center at Vanderbilt University. We thank Gabrielle Strouse, Brian Verdine, Lauren Deisenroth, Shazia Ansari, and Paige Holden for help with data collection and coding, and the parents and children for participating.

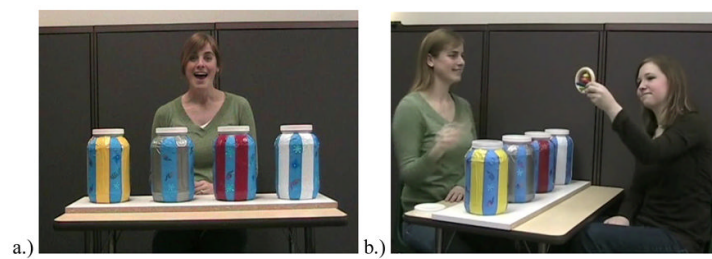
## References

- Akhtar, N. Is joint attention necessary for early language learning?. In: Homer, BD.; Tamis-LeMonda, CS., editors. *The development of social cognition and communication*. Mahwah, NJ: Erlbaum; 2005. p. 165-179.
- Akhtar N, Jipson J, Callanan M. Learning words through overhearing. *Child Development*. 2001; 72:416–430. [PubMed: 11333075]
- Akhtar N, Tomasello M. Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology*. 1996; 14:79–93.
- Allen, R.; Scofield, J. Word learning from videos: Evidence from 2-year-olds. 2008. Manuscript submitted for publication
- Alwitt LF, Anderson DR, Lorch EP, Levin SR. Preschool children’s visual attention to attributes of television. *Human Communication Research*. 1980; 7:52–67.
- Anderson DR, Bryant J, Wilder A, Santomero A, Williams M, Crawley AM. Researching *Blue’s Clues*: Viewing behavior and impact. *Media Psychology*. 2000; 2:179–194.
- Anderson DR, Pempek TA. Television and very young children. *American Behavioral Scientist*. 2005; 48:505–522.
- Baldwin DA. Infants’ ability to consult the speaker for clues to word reference. *Journal of Child Language*. 1993; 20:395–418. [PubMed: 8376476]
- Baldwin DA, Markman EM, Bill B, Desjardins RN, Irwin JM, Tidball G. Infants’ reliance on social criterion for establishing word-object relations. *Child Development*. 1996; 67:3135–3153. [PubMed: 9071774]
- Baldwin DA, Moses LJ. The ontogeny of social information gathering. *Child Development*. 1996; 67:1915–1939.
- Baldwin DA, Moses LJ. Links between social understanding and early word learning: Challenges to current accounts. *Social Development*. 2001; 10:309–329.
- Barr R, Dowden A, Hayne H. Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development*. 1996; 19:159–170.

- Barr R, Hayne H. Developmental changes in imitation from television in infancy. *Child Development*. 1999; 70:1067–1081. [PubMed: 10546335]
- Behne T, Carpenter M, Tomasello M. One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*. 2005; 8:492–499. [PubMed: 16246240]
- Bigelow AE, MacLean BK, MacDonald D. Infants' response to live and replay interactions with self and mother. *Merrill Palmer Quarterly*. 1996; 42:596–611.
- Carpenter M, Tomasello M, Striano T. Role reversal imitation and language in typically developing infants and children with autism. *Infancy*. 2005; 8:253–278.
- Cooper RP, Aslin RN. Preference for infant-directed speech in the first month after birth. *Child Development*. 1990; 61:1584–1595. [PubMed: 2245748]
- Crawley AM, Anderson DR, Wilder A, Williams M, Santomero A. Effects of repeated exposures to a single episode of the television program *Blue's Clues* on the viewing behaviors and comprehension of preschool children. *Journal of Educational Psychology*. 1999; 91:630–637.
- Csibra, G.; Gergely, G. Social learning and social cognition: The case of pedagogy. In: Johnson, MH.; Munakata, YM., editors. *Processes of change in brain and cognitive development. Attention and Performance*. Vol. XXI. Oxford: Oxford University Press; 2006. p. 249-274.
- Csibra G, Gergely G. Natural pedagogy. *Trends in Cognitive Sciences*. 2009; 13:148–153. [PubMed: 19285912]
- DeLoache, JS. The symbol-mindedness of young children. In: Hartup, W.; Weinberg, RA., editors. *Child psychology in retrospect and prospect: In celebration of the 75th anniversary of the Institute of Child Development*. Vol. 32. Mahwah, NJ: Lawrence Erlbaum Associates; 2002. p. 73-101.
- Deocampo JA, Hudson JA. When seeing is not believing: Two-year-olds' use of video representations to find a hidden toy. *Journal of Cognition and Development*. 2005; 6:229–260.
- Evans Schmidt MK, Crawley-Davis AM, Anderson DR. Two-year-olds' object retrieval based on television: Testing a perceptual account. *Media Psychology*. 2007; 9:389–409.
- Fenson L, Pethick S, Renda C, Cox JL, Dale PS, Reznick JS. Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*. 2000; 21:95–116.
- Floor P, Akhtar N. Can 18-month-olds learn words by listening in on conversations? *Infancy*. 2006; 9:327–339.
- Gergely G, Egyed K, Király I. On pedagogy. *Developmental Science*. 2007; 10:139–146. [PubMed: 17181712]
- Hains SM, Muir DW. Effects of stimulus contingency in infant-adult interactions. *Infant Behavior and Development*. 1996; 19:49–61.
- Hayne H, Herbert J, Simcock G. Imitation from television by 24- and 30-month-olds. *Developmental Science*. 2003; 6:254–261.
- Herold KH, Akhtar N. Imitative learning from a third-party interaction: Relations with self-recognition and perspective taking. *Journal of Experimental Child Psychology*. 2008; 101:114–123. [PubMed: 18635193]
- Király, I.; Csibra, G.; Gergely, G. The role of communicative-referential cues in observational learning during the second year. Poster presented at the 14th biennial International Conference on Infant Studies; Chicago. 2004 May.
- Koenig MA, Echols CH. Infants' understanding of false labeling events: the referential roles of words and the speakers who use them. *Cognition*. 2003; 87:179–208. [PubMed: 12684199]
- Krcmar M, Grela BG, Lin Y-J. Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology*. 2007; 10:41–63.
- Kuhl PK. Is speech learning 'gated' by the social brain? *Developmental Science*. 2007; 10:110–120. [PubMed: 17181708]
- Kuhl PK, Tsao F-M, Liu H-M. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 2003; 100:9096–9101.
- Maguire M, Hirsh-Pasek K, Golinkoff RM, Brandone A. Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*. 2008; 11:628–634. [PubMed: 18576970]



- Meltzoff AN. Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development*. 1985; 56:62–72.
- Meltzoff, AN. Imitation and other minds: The “Like Me” hypothesis. In: Hurley, S.; Chater, N., editors. *Perspectives on Imitation: From Neuroscience to Social Science*. Vol. 2. Cambridge, MA: MIT Press; 2005. p. 55-77.
- Meltzoff AN, Moore MK. Imitation, memory and the representation of persons. *Infant Behavior and Development*. 2002; 25:39–61.
- Moore, C. Understanding self and others in the second year. In: Brownell, CA.; Kopp, CB., editors. *Socioemotional development in the toddler years: Transitions and transformations*. New York, NY: The Guilford Press; 2007. p. 43-65.
- Nelson, K. Event knowledge: Structure and function in development. Hillsdale, NJ: Erlbaum; 1986.
- Nielsen M. Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*. 2006; 42:555–565. [PubMed: 16756445]
- Nielsen M, Simcock G, Jenkins L. The effect of social engagement on 24-month-olds’ imitation from live and televised models. *Developmental Science*. 2008; 11:722–731. [PubMed: 18801128]
- Poulin-Dubois D, Forbes JN. Toddlers’ attention to intentions-in-action in learning novel action words. *Developmental Psychology*. 2002; 38:104–114. [PubMed: 11806694]
- Rice ML, Huston AC, Truglio R, Wright J. Words from “Sesame Street”: Learning vocabulary while viewing. *Developmental Psychology*. 1990; 26:421–428.
- Rice ML, Woodsmall L. Lessons from television: Children’s word learning when viewing. *Child Development*. 1988; 59:420–429. [PubMed: 3359862]
- Rideout, VJ.; Hamel, E. *The media family: Electronic media in the lives of infants, toddlers, preschoolers and their parents*. Menlo Park, CA: Kaiser Family Foundation; 2006.
- Rogoff, B. *The cultural nature of human development*. Oxford: Oxford University Press; 2003.
- Roseberry S, Hirsh-Pasek K, Parish-Morris J, Golinkoff RM. Live action: Can young children learn verbs from video? *Child Development*. 2009; 80:1360–1375. [PubMed: 19765005]
- Ross HS, Lollis SP. Communication within infant social games. *Developmental Psychology*. 1987; 23:241–248.
- Schmitt KL, Anderson DR. Television and reality: Toddlers’ use of visual information from video to guide behavior. *Media Psychology*. 2002; 4:51–76.
- Scofield J, Williams A, Behrend DA. Word learning in the absence of a speaker. *First Language*. 2007; 27:297–311.
- Strouse GA, Troseth GL. “Don’t try this at home”: Toddlers’ imitation of new skills from people on video. *Journal of Experimental Child Psychology*. 2008; 101:262–280. [PubMed: 18675431]
- Tomasello M, Barton M. Learning words in nonostensive contexts. *Developmental Psychology*. 1994; 30:639–650.
- Tomasello M, Carpenter M, Call J, Behne T, Moll H. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*. 2005; 28:675–735. [PubMed: 16262930]
- Tomasello M, Strosberg R, Akhtar N. Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*. 1996; 23:157–176. [PubMed: 8733565]
- Troseth GL, DeLoache JS. The medium can obscure the message: Young children’s understanding of video. *Child Development*. 1998; 69:950–965. [PubMed: 9768480]
- Troseth GL, Saylor MM, Archer AH. Young children’s use of video as a source of socially relevant information. *Child Development*. 2006; 77:786–799. [PubMed: 16686801]
- Troseth, GL.; Saylor, MM.; Strouse, GL. Do toddlers learn words from video?. 2009. Unpublished manuscript
- Uzgiris IC. Two functions of imitation during infancy. *International Journal of Behavioral Development*. 1981; 4:1–12.
- Werker JF, Cohen LB, Lloyd VL, Casasola M, Stager CL. Acquisition of word-object associations by 14-month-olds. *Developmental Psychology*. 1998; 34:1289–1309. [PubMed: 9823513]
- Woodward AL. Infants’ developing understanding of the link between looker and object. *Developmental Science*. 2003; 6:297–311.



**Figure 1.** Views of the (a) researcher in the *Live* and *Video Addressed* conditions and the (b) researcher and confederate in the *Live* and *Video Onlooker* conditions.

**Table 1**

The number of children in the four conditions who displayed each of four possible patterns of response to test questions

	Comprehension not Preference	Preference not Comprehension	Sign Test p-value	Both	Neither
<i>Live Onlooker</i>	9	2	$p = .033^*$	1	4
<i>Video Onlooker</i>	8	1	$p = .020^*$	2	5
<i>Live Addressed</i>	4	2	$p = .344$	3	7
<i>Video Addressed</i>	4	2	$p = .344$	2	8

Note. The p-values are from sign tests comparing the first two patterns ( $*p < .05$ )