



Published in final edited form as:

Child Dev. 2009 ; 80(5): 1360–1375. doi:10.1111/j.1467-8624.2009.01338.x.

Live Action: Can Young Children Learn Verbs From Video?

Sarah Roseberry,
Temple University

Kathy Hirsh-Pasek,
Temple University

Julia Parish-Morris, and
Temple University

Roberta Michnick Golinkoff
University of Delaware

Abstract

The availability of educational programming aimed at infants and toddlers is increasing, yet the effect of video on language acquisition remains unclear. Three studies of 96 children aged 30–42 months investigated their ability to learn verbs from video. Study 1 asked whether children could learn verbs from video when supported by live social interaction. Study 2 tested whether children could learn verbs from video alone. Study 3 clarified whether the benefits of social interaction remained when the experimenter was shown on a video screen rather than in person. Results suggest that younger children only learn verbs from video with live social interaction while older children can learn verbs from video alone. Implications for verb learning and educational media are discussed.

Can infants and toddlers learn language from video? A recent upsurge in children’s “edutainment” media (Brown, 1995) has promised to enhance cognitive and language skills in infants and toddlers (Garrison & Christakis, 2005). Despite the prevalence of toddler-directed educational programming and high rates of television viewing among very young children (Zimmerman, Christakis & Meltzoff, 2007a), only a handful of studies have investigated whether children under 3 years can learn words from video. These studies yield mixed results, with a plurality of research suggesting that children learn language through *responsive* and *interactive* exchanges with adults and other children rather than by passive viewing (Krcmar, Grela & Lin, 2007; Naigles, Bavin & Smith, 2005). The current research explores mechanisms that drive word learning in young children. In particular, we focus on how children acquire verbs, the gateway to grammar. Can children learn verbs when passively exposed to video versus when they interact with a person while watching videos?

Two distinct lines of research are relevant to the question of whether children can learn language from video. The first, correlational, investigates the relation between television usage patterns and language outcomes. The second, experimental, involves laboratory studies that use video displays to test for vocabulary acquisition.

Research on television viewing and word learning outcomes in children under age three is rare, but at least two relevant studies exist. One compares individual youngsters’ viewing habits while the other is a laboratory-based word learning experiment. The first is a multi-

state survey by Zimmerman, Christakis, and Meltzoff (2007b), who asked parents of 8- to 24-month-olds about the content and frequency of their children's exposure to screen media. A standardized parent-report language measure was also given over the phone. Results revealed that the vocabulary sizes of 8- to 16-month-olds varied inversely as a function of the number of hours of television that they watched per day. That is, infants who spent more time with screen media had smaller vocabulary sizes than same-aged peers who watched less television. Notably, no association between television use and vocabulary was found in children aged 17 to 24 months. In conjunction, these results might indicate a developmental trajectory in the ability to manage input from screen media, or that the process of language learning becomes more robust over time.

The second study, conducted by Krcmar and colleagues (2007), investigated word learning from video in a laboratory setting. Using a repeated measures design, researchers presented novel objects and their names to children aged 15 to 24 months in four ways: A real life interaction with an adult that used joint attention; a real life interaction that incorporated discrepant adult and child attention; an adult on video; and a clip from a children's television show, *Teletubbies*. Following each training, children were asked to fast map the novel label onto the object that was named. Results revealed that when children were trained using children's television shows, 62% of children aged 22 to 24 months selected the target object among four distracters while only 23% of children aged 15 to 21 months were successful. In contrast, word learning was significantly better in both older and younger children when taught by a live person using joint attention (93% and 52% accurate, respectively). Thus, even though *all* children learn language best in the context of social interaction, older toddlers may learn words from video alone. Taken together, these studies suggest that as children age and have more experience with television, they are better able to conquer the "video deficit" (Anderson & Pempek, 2005) to learn words.

In contrast to the dearth of research on how young children learn words from video, there is a significant body of data examining television and language outcomes in older children. For example, Singer and Singer (1998) demonstrated in a naturalistic study that 3- and 4-year-old children learn nouns from *Barney & Friends*. Another study suggests social interaction improves recall from educational videos among 4-year-olds (Reiser, Tessmer & Phelps, 1984). In this experimental study, half of the children watched an educational program while participating in a content-related interaction with a live experimenter, while the other half of the children simply watched the video. Three days later, children who interacted with an adult while viewing were better able to identify numbers and letters that had been presented in the video. Finally, Rice and Woodsmall (1988) showed children an animated program in a naturalistic setting that contained 20 novel words from a variety of word classes (e.g., nouns, verbs, adjectives). Results of pre- and posttests suggested that both 3- and 5-year-olds learned to pair a novel word heard on television with a picture that depicted that word. Even here, 5-year-old children learned only 5 out of a possible 20 words, while 3-year-old children learned only 2 out of 20 words.

Although several studies of television and language have been conducted, the literature as a whole has several limitations. First, only two studies to date have examined video and word learning outcomes in children younger than 3 years of age, despite the recent push for educational programming targeting that demographic. Second, video language studies with very young children have primarily investigated noun learning (e.g., Krcmar et al., 2007). To become sophisticated speakers, children must also acquire verbs, adjectives, and prepositions. In particular, verbs enable children to speak about a dynamic world of relations between objects, and may form the bridge into grammar (Marchman & Bates, 1994). Verbs are also, however, more difficult to learn than nouns (Bornstein et al., 2004; Gentner, 1982; Hirsh-Pasek & Golinkoff, 2006). Finally, it is critical to go beyond correlational research to

better understand the relationship between language used on video and vocabulary acquisition (Linebarger & Walker, 2005).

Are there any other sources of information that can shed light on children's ability to learn from video? In addition to studies examining the direct effects of television on language outcomes, there are a number of language experiments that use video as a stimulus at training or test but were not specifically designed to test the effects of television viewing, *per se*.

Many developmental studies use video paradigms to teach children new linguistic concepts through controlled exposures or to test children's inherent knowledge of language. Although the videos used in these studies do not approximate the quality of those used in commercial programming, results nonetheless indicate that children under age 3 are able to learn information from video displays. The practice of using video clips to teach new words is particularly useful in studies of verb learning, because dynamic scenes best illustrate the meaning of an action word. For example, studies have used the Intermodal Preferential Looking Paradigm to investigate children's emerging grammar (Gertner, Fisher & Eisengart, 2006; Hirsh-Pasek, Golinkoff & Naigles, 1996). In this video based paradigm, children were able to link a transitive (Cookie Monster is *licking* Big Bird) or intransitive (e.g., Cookie Monster is *licking*) verb with a video depicting the appropriate action at 28 months of age, even when the verb was unfamiliar.

Recent research has explored whether children can use televised social cues to distinguish words for intentional actions like *pursuing* from words for unintentional actions like *wandering* (Poulin-Dubois & Forbes, 2002; 2006). Toddlers watched videos that depicted pairs of intentional and unintentional actions performed by a real person. The actions were perceptually similar, and therefore distinguishable only by the videotaped actor's social cues indicating intent or lack of intent. Results suggested that by 27 months of age, children succeeded at a matching task using relatively subtle social cues like eye gaze and facial expression – even when those social cues are presented on video.

Other studies using video paradigms further illuminate the optimal conditions for learning from video, and suggest a strong role for social cues, particularly during infancy. For example, young children are better able to reproduce an action after a delay when the demonstration is presented live, rather than on video (Barr, Muentener, Garcia, Fujimoto & Chavez, 2007). However, when 12- to 21-month-olds saw videotaped actions twice as many times as the live action, they imitated just as well as children who had seen the live demonstration with fewer repetitions. In addition to repetition, several studies indicate that contingent social interactions also enable children to learn from video (Nielsen, Simcock & Jenkins, in press; Troseth, Saylor & Archer, 2006). For example, Nielsen and colleagues (in press) demonstrated that providing children with a contingent social partner who interacted via a closed circuit video increased children's imitation. In fact, children were just as likely to imitate a contingent partner on video as they were to imitate a live model. Thus, although children typically learn better from live social interactions than from video (see also Conboy, Brooks, Taylor, Meltzoff & Kuhl, 2008; Kuhl, Tsao & Liu, 2003; Kuhl, 2004), repetition and contingent social interactions improve video-based learning. Finally, a study by Naigles and colleagues (2005) also highlighted the importance of social interaction when using video to test children's verb learning. First, novel verbs were presented to 22- and 27-months-old children in interactive play settings. Here, children were able to interact with the live experimenter and even perform the actions themselves while hearing the novel verb sixteen times. Then, videotaped scenes were used to test children's ability to match each novel verb with the corresponding target action. Children were able to match a novel verb to the correct action when two scenes were presented side-by-side on a video screen, but only

when the novel verb was presented in a rich syntactic context (e.g., “Where is she *lorping* the ball?”) and not in a bare context (e.g., “Where’s *lorping*?”). Importantly, although Naigles and colleagues relied on an interactive teaching session and used video only for testing, their results suggest that children are able to transfer knowledge about novel verbs to a video format. Although these studies do not speak to the specific ability of children to learn words from video, these results suggest that social interaction and repetition are central features in children’s successful training and learning from video in other domains (e.g., imitation; Barr et al., 2007; Nielsen et al., in press; Troseth et al., 2006), and that social interaction during training allows young children to generalize their knowledge of novel verbs to video (Naigles et al., 2005). These conclusions are useful in considering the demands of learning words from video.

The current research adds to our knowledge in four ways. First, we study verbs rather than nouns. This move is important because verbs are the architectural centerpiece of language and ultimately control the shape of sentences. Second, word learning demands more than matching words to objects or actions (e.g., Brandone, Pence, Golinkoff & Hirsh-Pasek, 2007; Krcmar et al., 2007). Rather, a child who truly understands a word forms a category of the referent and can *extend* the word to a new, never-before-seen instance of the referent. For example, after a child sees two or three different colored cups and hears the label *cup*, they should be able to label a green cup even if they have never seen one before. Third, most existing research on television and language outcomes is focused on children aged 3 years and older. As the television industry increases programming for children under age three, however, it is imperative for both researchers and the public to understand how television might affect language at this sensitive age. Finally, prior research underscores the role of social support in verb learning (Brandone et al., 2007; Naigles et al., 2005; Tomasello & Akhtar, 1995), particularly in the case of video-mediated learning (Poulin-Dubois & Forbes, 2002; 2006).

This research is among the first to explore the utility of television for childhood language learning by combining the television and language learning literatures that have, up to now, presented mixed evidence about the effect of television on language. We use highly engaging videos similar to the commercial programming used in research on the direct effects of television on language, and we couple these with the tightly controlled conditions that are characteristic of laboratory-based word learning studies. In this way, we hope to provide children with the optimal video verb-learning experience. We hypothesize that 2-year-olds will be able to learn verbs from video, but only when the video is accompanied by live social interaction. Three studies examine these claims.

Study 1: Do children learn verbs from video in an optimal learning environment?

In light of the finding that social cues, live or demonstrated on video, help children learn verbs (e.g., Naigles et al., 2005; Poulin-Dubois & Forbes, 2002, 2006), the first study is designed to optimize the word-learning situation. We provide children with stimuli from child-oriented commercial programming, and supplement the video with interactive teaching by a live experimenter. When provided with video *and* social support, will children be able to learn verbs?

Method

Participants—Forty monolingual, English-reared children were tested, twenty in each of two age groups (half male and half female): 30- to 35-months (mean age = 33.74 months; range = 30.66–35.66) and 36- to 42-months (mean age = 39.36 months; range = 36.21–

42.87). Data from an additional 5 children were discarded for prematurity (2), bilingualism (1), low attention (1), and experimenter error (1). Children included in the final sample were predominantly white and from middle-class homes in suburban Philadelphia.

Paradigm and dependent variable—Children were tested using the Intermodal Preferential Looking Paradigm (IPLP; see Figure 1; Golinkoff, Hirsh-Pasek & Cauley, 1987; Hirsh-Pasek & Golinkoff, 1996). The IPLP is an established method of studying language acquisition in infants and toddlers. Children see a series of events on a video screen, and are shown two different events (one on each side of split-screen) at test. For example, the video might show a mother performing an action with her child (e.g., bouncing her child up and down on her lap). The scene is accompanied by audio presenting a novel word such as *blicking* (e.g., *Look, she's blicking!*). At test, the same child is shown a split screen wherein the action from training appears on one side of the screen and a scene depicting another novel action appears on the other side of the screen (e.g., a mother turning her child around). The child hears, “*Where is blicking?*” A number of studies show that if children learn that the word *blick* refers to a particular novel action, then they prefer looking at the action labeled by the word *blick* at test. Comprehension is measured by looking time in seconds to the matching versus the non-matching screen.

Design and procedure—Videos were created using high quality clips from *Sesame Beginnings*, a video series produced by Sesame Workshop for children as young as six months (Hassenfeld, Nosek & Clash, 2006a; 2006b). Children were seated on a parent’s lap, four feet in front of the video monitor. Parents were instructed to close their eyes during the video, and all parents complied. An experimenter sat in a chair next to the parent/child dyad, also with eyes closed, except during the prepared live interaction sequence. The task was presented in four phases: *Introduction, Salience, Training, and Testing* (see Table 1). With the exception of the *Training Phase*, each phase was presented entirely in video format. Each video segment was separated by a 3 second “centering trial” in which attention was redirected to the middle of the video monitor. Centering trials showed a baby laughing and were accompanied by audio instructions to look at the screen (“Wow! Look up here!”). All children recentered for at least .75 seconds.

In the *Introduction Phase*, children were shown one of four possible characters (e.g., Cookie Monster). Two of the characters were Sesame puppets and two were real babies. The *Introduction Phase* acquainted children with the single character who would perform the target action, because children learn verbs better when they are familiar with the agent in a scene (Kersten, Smith & Yoshida, 2006). Video clips in the *Introduction Phase* appeared twice: first on one half of the screen, and then on the other. The left/right presentation order was counterbalanced across condition. This demonstrated to the child that actions could appear on either side of the screen. Importantly, the video clips used to introduce the characters were not included in any other phase of the video.

During the *Salience Phase*, children saw a preview of the exact test clips used during the *Test Phase*. Measuring looking time to this split-screen presentation before training allows detection of *a priori* preferences for either of the paired clips. Lack of an *a priori* preference in the *Salience Phase* indicates that differences in looking time to one clip or another at test are due to the effect of the *Training Phase*.

Next, the *Training Phase* was presented in four 6-second trials. In each of the trials, children saw the same character from the *Introduction Phase* perform one of four possible actions (see Table 2 for novel words and descriptions). During the first two trials, the experimenter sitting next to the parent and child presented children with a novel word, and the second two trials used video to present children with the same novel word. For the live interaction

sequence, the videotape was programmed to show a solid black screen for 15 seconds (the equivalent of two 6-second trials plus one 3-second centering trial). When the black screen appeared, the experimenter produced the same puppet (or doll) as was seen in the *Introduction Phase* and performed an action with the puppet (or doll) while presenting a novel verb six times using full syntax (three times for each of two training trials; e.g., “Look at Cookie Monster *wezzling!* He’s *wezzling!* Cookie Monster is *wezzling!*”). The experimenter was trained to use infant-directed speech and to look at the child during the live action sequence. Following the live interaction sequence, children saw two 6-second clips of the same animated character (e.g., Cookie Monster) demonstrating the same action (e.g., *wezzling*) that they had just seen with the live experimenter. Each video clip was accompanied by pre-recorded audio that matched the script used during the live interaction. Thus, children heard the novel verb 12 times during the *Training Phase*. Nonsense words are commonly used instead of real words in language training studies (e.g., Brandone et al., 2007; Hirsh-Pasek et al., 1996; Naigles et al., 2005) and were used in the current study to control for children’s previous word knowledge.

Finally, the *Test Phase* consisted of four trials. In each of the trials, a split-screen simultaneously presented two novel clips. One of the clips showed a new actor (e.g., a baby) performing the same action that children saw during the *Training Phase* (e.g., *wezzling*). The other clip showed the same new actor (e.g., the baby) doing a different action, never seen before (e.g., *glorping*). Children who saw a Sesame puppet during the *Training Phase* saw a real baby at test and children who saw a real baby during the *Training Phase* saw a Sesame puppet at test. Importantly, and a strength of this design, *all* conditions required children to extend their verb knowledge to a new actor.

Of the four Test Trials, trials 1 and 2 were designed to test children’s ability to generalize the trained verb to an action performed by a novel actor (e.g., from puppet to baby or from baby to puppet). In this *Extension Test*, pre-recorded audio asked children to find the action presented during training, using the novel word (e.g., “Where is *wezzling*? Can you find *wezzling*? Look at *wezzling!*”). If children learned the target verb, they should look at the matching action screen during each of the first two test trials.

Test Trials 3 and 4, the *Stringent Test*, provided an additionally strong test of word learning (Hollich, Hirsh-Pasek & Golinkoff, 2000) by asking whether children have truly mapped the novel verb to the particular novel action. Here, we test whether children will accept any verb for the action presented during training or whether children expect the trained verb to accompany the trained action. Based on the theory of mutual exclusivity (Markman, 1989), children should prefer attaching only *one* verb to any given action. Thus, Test Trial 3, the *new verb trial*, asked children to find a novel action that was *not* labeled during training (“Where is *glorping*? Can you find *glorping*? Look at *glorping!*”; *glorping* and *spulking* were terms used respectively for non-trained verbs). If children truly learned the target verb (e.g., *wezzling*), they should *not* look toward the action previously labeled *wezzling* during the *new verb trial*, and should look instead toward the unfamiliar action, hereafter the non-matching action. That is, if children thought that the *wezzling* action had already been named, they should now look toward the non-matching action. Test Trial 4, the *recovery trial*, asked children to renew their attention to the trained action by asking for it again by name (“Where is *wezzling*? Can you find *wezzling*? Look at *wezzling!*”). If children have mapped the original verb to the originally named action, they should redirect their gaze to the original action.

In sum, if children learned the target word during the *Training Phase*, they should look more at the side of the screen that showed *wezzling* during the *Extension Test* (trials 1 and 2). If children were able to do more than simply extend, they were also able to use mutual

exclusivity and show that they remembered the action's name, then they should show a quadratic pattern of looking in the *Stringent Test* (trials 3 and 4; see Figure 2). In other words, children should look toward the matching screen in Test Trials 1 and 2, look away in Test Trial 3 (the *new verb trial*), and look back to the matching screen during Test Trial 4 (the *recovery trial*). This v-shape in visual fixation time during the Test Trials 1 and 2 (the *Extension Test*), Test Trial 3 and Test Trial 4 forms a quadratic pattern of looking. Because the *Stringent Test* was designed to probe successful extension of the novel word, analyses for the *Stringent Test* were conducted only when children were able to extend the novel word in the *Extension Test*.

Following the four Test Trials, the entire sequence was repeated for a second novel verb (e.g., *blicking*). The child's assignment to one of four possible conditions determined the particular verbs a child was exposed to during the *Training and Test Phases*. All test trials for a given verb used an identical set of video clips and all conditions were counterbalanced in a between-subjects design (see Table 1). For example, *wezzling* appeared both as the novel verb in the first position and as the novel verb in the second position, depending on condition. Finally, in half of the conditions, puppets were used to train the novel verb while babies were used in test sequences. In the other half of conditions, babies were featured in the videos used to train the novel verb while puppets were used in the test clips.

Coding—Each child's head and shoulders were videotaped for offline coding of gaze duration. Gaze direction was also coded during phases where the child saw a split-screen. Each participant's gaze direction and duration was coded two times by a coder blind to condition and experimental hypotheses. A second blind coder recoded 20% of the videotapes. Reliability within and between coders was $r = .97$ and $.96$, respectively.

Several studies in the literature suggest that older children respond very quickly to auditory instructions at test (e.g., "Where is *wezzling*?"), and tend to make a discriminatory response during the first 2 seconds of a trial (Fernald et al., in press; Gertner et al., 2006; Hollich & George, 2008; Meints, Plunkett, Harris & Dimmock, 2002; Swingley, Pinto & Fernald, 1999). Longer test trials give older children enough time to make a discriminatory response and then to visually explore both sides of the screen. Measures of looking time for older children that are averaged across such a long trial are likely to mask the response and reflect instead the visual exploration after the response (Fernald et al., in press). Therefore, although the current test trials were arranged to show children the full 6-second video clip, we elected to code gaze information during the first 2 seconds of each test trial in addition to the standard coding. Separate coders were used for the standard coding and the two-second coding.

For each child, we calculated the percentage of looking time to either side of the screen during salience trials and test trials. For the salience trials, we divided the number of seconds spent looking at the target action by the total number of seconds spent looking at the video screen. For test trials, percentages were calculated for the first 2-seconds of the trial only. Any score greater than 50% indicated that the child spent more time looking toward the target action than the non-target action, while any score less than 50% indicated that the child spent more time looking at the non-target.

Results

A preliminary multivariate analysis of variance (MANOVA) indicated no effect of gender on mean looking time and no *a priori* preference for one test clip over the other in any condition during the *Salience Phase*. Thus, the data were collapsed across gender and condition.

Extension Test of verb learning—Because Test Trials 1 and 2 were identical, both asking the child to find the trained action using the trained verb, data from Test Trials 1 and 2 were averaged for both the first and then the second verb. Averaging two test trials increases the reliability of children’s responses. Because each child saw two verbs, a 2 (age group: 30–35 months, 36–42 months) x 2 (verb position: 1st verb, 2nd verb) repeated measures analysis of variance (ANOVA) was used to determine the effect of age group and verb position on learning. Results indicate a main effect of verb position, $F(1,38) = 4.75, p < .05, \eta_p^2 = .11$, but no main effect of age group, $p > .05, \eta_p^2 = .04$, and no interaction effect, $p > .05, \eta_p^2 = .01$. This suggests that children reacted differently to the first verb than they did to the second verb across both age groups, and that the *pattern* of reaction was the same for both groups (e.g., if the younger children reacted a certain way to the first verb, so did the older children). Due to the absence of an age effect, data were pooled across age groups for further analyses.

To decipher the differences between child performance on the first verb and the second verb, we conducted planned paired-sample *t*-tests comparing children’s looking times to the matching action versus the non-matching action for the verbs in both positions. Results indicated that children looked equally toward the matching and non-matching action for the first verb, $t(39) = .77, p > .05$, but looked significantly *longer* to the matching action than the non-matching action for the second verb, $t(39) = 4.67, p < .001$. This mean looking time was also significantly greater than chance (50%), $ps < .001$. Thus, the training trials changed children’s looking behavior for the second verb in the sequence. Children who did not initially prefer to look at either action over the other (as demonstrated by chance rates of looking during the salience trials) now preferred to look at the matching action during the second verb.

The *Extension Test* provides evidence of learning in both age groups, but only for the second verb. To determine whether children succeeded in the stronger test of verb learning, we further analyzed the looking patterns for the second verb only.

Stringent Test of verb learning—Recall that Test Trial 3 (*new verb trial*) asked children to find the action labeled by a new novel verb and Test Trial 4 (*recovery trial*) asked children to find the action that was labeled during training. This strong test of verb learning used the average looking time during the first two test trials, looking time in Test Trial 3 (*new verb trial*) and looking time in Test Trial 4 (*recovery trial*) to examine whether a quadratic pattern emerged from the three measures. Although children’s responses in Test Trials 1 and 2 (*Extension Test*) were averaged, Test Trials 3 and 4 ask different questions and success in these trials would be evidenced by looking in opposite directions. Therefore, the test trials in the *Stringent Test* were analyzed individually.

Data were analyzed using a repeated measures one-way ANOVA with three conditions (extension test, new verb trial, recovery trial). A significant quadratic pattern emerged, $F(1,39) = 6.16, p < .05, \eta_p^2 = .14$, indicating that children learned the second verb, even by the standards of the strong test of verb learning (see Figure 3). Furthermore, paired samples *t*-tests revealed that children looked equally to both sides of the screen in Test Trial 3, $t(39) = 0.02, p > .05$, but children looked significantly longer toward the matching screen than the non-matching screen in Test Trial 4, $t(39) = 3.23, p < .05$. Thus, children succeeded in the *Extension Test*, showed no preference in the *new verb trial*, and again preferred the matching screen in the *recovery trial*.

Discussion

Our results reveal that children as young as 30 to 35 months of age can learn verbs from a combination of video and social interaction. This finding is consistent with prior research,

and furthers it in several ways. First, our test of verb learning requires children to *extend* word meaning from one agent to another, rather than simply glue a word onto a scene (which may be mere association rather than word learning). For example, children who fail to *extend* might learn that Cookie Monster can *wezzle*, but not that a baby can also *wezzle*. The current study used a particularly difficult extension task, asking children to extend verb knowledge from puppets to humans and vice versa. Although several researchers have noted that children's ability to generalize increases as linguistic ability increases (Forbes & Farrar, 1995; Maguire, Hish-Pasek, Golinkoff, & Brandone, 2008), few studies have used extension.

Second, in addition to extension, the current study offered a *Stringent Test* of verb learning. In Trials 3 and 4, children were asked to look away from the trained action upon hearing a *new* novel word and then redirect their attention to the trained action. While the extension task ensures that word learners are able to generalize their knowledge of the novel verb, the stringent test of verb learning asks children to constrain their generalization to the trained action *only* and exclude other referents. Success in this task indicates that children in the current study acquired a sophisticated understanding of the trained verb through video and live action. Interestingly, although results from Test Trial 3 (the *new verb trial*) suggests that children did not look significantly toward either side of the screen, children did show a significant quadratic pattern of looking. In this case, the v-shaped quadratic pattern of looking indicates that children succeeded in the *Extension Test*, then looked away from the trained action in Test Trial 3 (but did not prefer either side of the screen), and finally recovered their looking toward the trained action in Test Trial 4 (the *recovery trial*). In general, data from individual children followed this pattern, as opposed to the expected quadratic pattern that would have required children to significantly look away from the trained action in Test Trial 3 (see Figure 2). Thus, even though children did not attach the novel verb in the *new verb trial* to the non-matching action, they were nonetheless able to shift their attention away from the trained action. Notably, children's familiarity with either the trained or the untrained action did not appear to inhibit (or drive) learning. If familiarity with an action had prevented children from attaching a second label to it, we would be more likely to see such an effect with the trained actions than the untrained actions. Whereas the nonsense words for the trained actions all had one-word English equivalents (see Table 2), the non-matching actions shown opposite the trained actions in the Salience Phase and Test Phase were not easily described by a single word in English (e.g., child laying on the parent's leg while the parent swings their leg up and down). Importantly, that children show evidence of verb learning in the current study suggests that familiarity with the trained actions did not impede children's ability to learn a verb.

Finally, we combined videos and live action to provide children with a rich, but complex learning environment. The videos used in the current study were taken from the *Sesame Beginnings* series and contained an action in an intricate setting. For instance, the video clip showing Cookie Monster *wezzling* places him in a living room, in front of a couch, with framed pictures in the background. Thus, children were challenged to zoom in on the action while faced with potentially distracting stimuli. The live action sequence in the current study was also limited in length and scope. Specifically, each live action sequence lasted 15 seconds, which was counterbalanced to encompass two training trials, but only comprised a small amount of the total experiment. In addition, the live action sequence was included between video clips, meaning that children had to redirect attention several times during the experiment. Switching between the modalities of live interaction and video may have been difficult for children, as evidenced by improved performance on the second verb. Success on the second verb in the sequence, regardless of the particular verb itself, may indicate that children required the first verb sequence to acclimate themselves to the demands of the procedure.

The current study suggests that children gleaned specific information from a complex situation to learn a verb. Moreover, both younger and older children were able to generalize a novel word to a new actor and show evidence of constraining novel word use. This first study, however, does not allow us to separate the effect of live action from the effect of video on word learning. Study 2 was designed to disentangle these variables.

Study 2: Do children learn verbs from video alone?

Both younger and older children showed evidence of verb learning in Study 1 when video displays were accompanied by social interaction. To determine if the live action sequence contributed to learning, it was important to evaluate verb learning through video alone. Given that previous research highlights the role of social interaction in verb learning, especially in younger children (e.g., Naigles et al., 2005), we predict that only the older children will be able to learn verbs from video alone.

Method

Participants—Participants included 20 monolingual children, balanced for gender, aged 30- to 35-months (mean age=33.00, range=30.09–35.60) and 20 children aged 36- to 42-months (mean age=39.39, range=36.21–41.90). An additional 6 children were discarded due to fussiness (3), experimenter error (1), parental interference (1), and side bias (1). As in Study 1, children were predominantly white and from middle-class homes in suburban Philadelphia.

Design—The design of Study 2 was identical to Study 1 with two important exceptions. First, all four *Training Phase* clips were televised, and second, no experimenter was present in the viewing room during the study. All four training trials used an identical video clip from *Sesame Beginnings* because recent research suggests that initial verb learning is improved when children are not distracted by multiple novel characters (Maguire et al., 2008). This second study was matched to the first study for the number of exposures to the novel verb, the specific audio script, and the total length of video to allow for direct comparison. All conditions were fully counterbalanced and videotapes were coded as in Study 1.

Results

As in Study 1, no gender differences emerged based on preliminary tests. Children also did not show *a priori* preferences during the *Saliency Phase* in any condition, so data were collapsed across gender and condition. For each child, a percentage of looking time to each side of the screen was calculated for each saliency trial and for the first two seconds of each test trial. Again, we conducted both the *Extension Test* and the *Stringent Test* of verb learning.

Extension Test of verb learning—Data from Test Trials 1 and 2 were averaged for both the first and then the second verb. Using the average looking time to the matching action for both verbs, a repeated measures ANOVA revealed a significant main effect of age group on verb learning, $F(1,38) = 10.69, p < .01, \eta_p^2 = .22$, but no main effect of verb order, $p > .05, \eta_p^2 = .01$, and no interaction between age group and verb order, $p > .05, \eta_p^2 = .01$. This result indicates that younger children performed differently than older children on the test trials, but that each age group performed similarly on both verbs. The data were pooled across verb position for further analyses.

Planned paired-samples *t*-tests suggest that only older children looked significantly longer toward the matching than non-matching during Test Trials 1 and 2, $t(19) = 4.36, p < .001$.

Older children's looking time toward the matching action was also significantly different than chance (50%), $ps < .05$. In contrast, younger children looked equally to the matching and non-matching actions, $ps > .05$. Together, these standard tests of verb learning suggest that only older children were able to learn a verb with video alone. To determine if older children succeeded in the stronger test of word learning, we further analyzed data from the older children only.

Stringent Test of verb learning—We conducted a strong test of verb learning using the average of Test Trials 1 and 2, Test Trial 3 (the *new verb trial*), and Test Trial 4 (the *recovery trial*). Data for the first and second verb were combined. A repeated measures one-way ANOVA with three factors (extension test, new verb trial, recovery trial) was used to examine the quadratic pattern of looking time. No significant quadratic pattern emerged, $ps > .05$, suggesting that the older age group was not successful in the stringent test of verb learning (see Figure 3). Interestingly, paired samples *t*-tests indicate that older children do not show a significant preference for either screen in Test Trial 3, $t(39) = 0.85$, $p > .05$, but do look significantly longer to the matching screen in Test Trial 4, $t(39) = 3.59$, $p < .05$. Although these results might indicate the presence of a quadratic pattern, the relative distances between the test trials are not sufficient to produce a significant quadratic pattern.

Thus, while older children succeeded in the *Extension Test* and were able to “glue” the word onto the target action even when the actor was different, they did *not* glean sufficient information from video alone to succeed in a *Stringent Test* of verb learning.

Discussion

As found in prior studies, Study 2 demonstrates that children older than 3 years show some evidence of verb learning from video alone (Reiser et al., 1984; Rice & Woodsmall, 1988; Singer & Singer, 1998), but that children younger than 36 months do not. It is noteworthy, however, that learning in this case was especially difficult; children had to extend the label for an action from a Sesame Street puppet to a human or from a human to a Sesame Street puppet. Children's ability to move between people and puppets in a word learning task suggests greater flexibility in the process of language acquisition than has been previously demonstrated. Furthermore, although older children in this study were able to demonstrate the ability to extend a novel word to a new actor, even these older children failed at the *Stringent Test* of verb learning. A relatively older age of acquisition combined with difficulty in the *Stringent Test* of verb learning indicates that video alone does not provide enough information for children to gain more than a cursory understanding of novel verbs.

Although the results of Study 2 indicate that video alone is insufficient for young children to learn a verb, perhaps video was equally useless to children in Study 1. It is possible that younger children in Study 1 used only the live action sequence to learn a verb and did not gain any supplemental information from video. If this were true, younger children in Study 1 would have learned a verb, both by *Extension* and the *Stringent Test* of verb learning, from only 15 seconds of live interaction and six presentations of the novel verb. Such a finding would be noteworthy in and of itself and is unprecedented in the verb learning literature. Future studies should further investigate this possibility. At minimum, however, the current results suggest that adding a short, constrained live action sequence to video was important for verb learning.

Study 3: Does social interaction merely provide more information?

Results from Study 2 suggest that children are better able to learn verbs when they experience real life social demonstrations as well as video displays, but what if the social interaction was only useful because it provided an example of a different agent during

training? This is a possible explanation because children in Study 1 saw two different kinds of exemplars of the target action (live and televised) during training whereas participants in Study 2 only saw one type of exemplar (televised). Although the live interaction was matched in time and content to Study 2, the experimenter was an integral part of the target action in Study 1, but not in Study 2. The purpose of Study 3 was to tease apart the effect of the live interaction from the effect of having two types of exemplars of the target action.

Method

Participants—Since only the younger age group showed differential performance in Study 1 and Study 2, only that age group was tested in Study 3. Eighteen children aged 30 to 35 months were recruited. Of these children, two were discarded due to fussiness, leaving a final sample of 16 children ($M=33.08$, range=29.66–35.21). Children in the final sample were predominantly white and from middle-class homes in suburban Philadelphia.

Design—Study 3 used the same video sequence as Study 1. In place of the first two training trials, a video clip of the experimenter appeared on the video screen, performing the actions with puppets or dolls as in the live interaction. In essence, the participant saw the same presentation as they would during the live interaction component of Study 2, except the experimenter was on the video screen, instead of live. The experimenter faced the camera and appeared life-size on the video screen. Length of training and audio were matched to both Study 1 and Study 2. If seeing an action demonstrated by two agents was responsible for increased learning in Study 1, results should mirror those from Study 1, and the younger age group should now show enhanced learning. If, however, some aspect of the live interaction, *per se*, was the driving force behind better performance in Study 1, the results of the control study should be the same as Study 2: the younger group should demonstrate no evidence of verb learning. We predict the latter.

Results

Preliminary analyses revealed no gender differences and no *a priori* preference for either test clip in any condition, $p>.05$, so data were pooled across gender and condition for further analyses.

Extension Test of verb learning—As in our previous studies, data from Test Trials 1 and 2 were averaged for both the first and then the second verb. A one-way repeated measures ANOVA revealed no main effect of verb position (e.g., first verb, second verb), so data from verb 1 and verb 2 were averaged for further analysis.

Paired-samples *t*-tests comparing looking time to the matching versus the non-matching actions revealed no significant looking patterns, $ps>.05$, and looking time toward the target was not significantly different than chance, $ps>.05$.

Thus, children in this younger age group failed to learn a verb from video even when two distinct sources (experimenter and animated character) presented the information. We did not analyze the results of the *Stringent Test* of verb learning in this study since children did not succeed in the *Extension Test* of verb learning.

Discussion

Study 3 was designed to clarify the results of Studies 1 and 2 by pinpointing the mechanism behind improved verb learning in Study 1. Was better performance due to some aspect of social interaction, or did children learn better because they saw more than one example of the action being labeled? If children succeeded in Study 1 because they learned the verb from two sources, then it should not matter whether the information is delivered live or on

video. If, however, the results of Study 1 emerged due to some benefit of live social interaction *per se*, then the children should fail in Study 3 when they see an experimenter on the video screen. This was, in fact, the result. Although the experimenter's positive affect and child-directed speech were preserved in both the live (Study 1) and televised (Study 3) experiments, children succeeded in learning a verb in Study 1 and failed to learn a verb in Study 3. These results suggest that some aspect of the live demonstration in Study 1, although it was highly controlled, was responsible for better verb learning.

General Discussion

Three studies revealed that it is possible for children to learn action words from video, but that specific conditions are necessary to promote such learning in children under age 3. We posed two questions: First, are children able to learn verbs at an early age when provided with video *and* social support? Second, can children learn verbs from video alone? We correctly predicted that younger children would be able to learn a verb from video when given support in the form of a live demonstration, but they would not learn a verb from watching video alone.

Our finding that young children only learned a verb when video was supplemented by live interaction is consistent with prior research emphasizing the importance of social cues to word learning (Baldwin, 1991; Brandone et al., 2007; Krcmar et al., 2007; Naigles et al., 2005; Sabbagh & Baldwin, 2001; Tomasello & Akhtar, 1995). Impressively, children in Study 1 not only extended a new verb to a new actor in a novel scene, but also passed a more stringent test of word learning that required them to remember the originally trained word-action pairing over time and to resist mapping a new word (e.g., *glorping*) onto the trained action. This test of verb learning goes beyond much of what is offered in the literature. Additionally, Study 2 affirmed that only older children can learn a verb without social support, but even then cannot pass the stringent test of word learning. Study 3 underscores that it is likely *social support* and not simply varied types of exposure to the actions and words that is the key to verb learning from video. These findings raise a number of questions about learning from video and have implications for research on both television and language.

On the Nature of Social Interaction

The live action sequence in Study 1 was designed to provide children with social interaction during video viewing, and thus create an environment where the child was an active, rather than passive, viewer. Although the social interaction was minimal and tightly controlled, children seemed to benefit nonetheless from its inclusion. Why did the live action matter? Several aspects of the live social interaction may have been beneficial to children.

First, although attention during familiarization was equal across conditions, it is possible that children's arousal was heightened during the live action sequence. Research suggests that arousal may be an important factor in infants' ability to encode and remember information (Kuhl, 2007). The current research did not measure children's arousal, but future studies might explore arousal as a mechanism by which children learn through live interaction.

Second, it is possible that the experimenter provided additional verbal and non-verbal information during the live action sequences. For instance, social labeling provided *in vivo* labeling of the target action during the live action sequence, which may have unintentionally provided the child with subtle contingencies, even in the controlled situation. The script used in the live action sequence was identical to the audio used in the videotaped training trial and the experimenter was instructed to follow the script regardless of the child's behavior. In

any interaction, however, some unintentional contingencies are possible. That is, the volume of the experimenter's voice might have been louder when the child was inattentive, or the experimenter might have shown signs of recognition if the child spoke. Although these subtle contingencies were not revealed in a review of the videotaped interactions, it is reasonable that some contingencies may occur in even the most controlled interactions. Furthermore, the experimenter's non-verbal behavior during the live action sequence may have been a source of additional information. Specifically, the experimenter's eye gaze unavoidably differed in the televised and live action sequences. Baldwin and Tomasello (1998) suggest that eye gaze is one of the elements that create a common ground for the participant and the experimenter. This common ground, they argue, forms the basis for social learning. Although the experimenter displayed positive affect and used child-directed speech in both the live and televised conditions, the live experimenter looked directly at the child whereas the televised experimenter looked toward the camera, as if it were the child. Thus, the experimenter on video may not have made eye contact with the child whereas the live experimenter was able to naturally establish eye contact with the child. If children used the experimenter's eye gaze to direct their attention to the action named by the novel word, we would have correctly anticipated differential performance on the live and videotaped studies. In sum, children may have used either the verbal or non-verbal cues present in the live action sequence. Because the current studies do not allow us to determine the relative effects of social labeling and eye gaze during the live interaction, it is unclear whether children used both verbal and non-verbal cues or whether only one element would have enabled children's verb learning.

Third, although Study 3 was designed to control for the number of sources presenting the novel verb, the mode of presentation was not controlled. That is, children in Study 1 saw the verb presented both live and on a video screen whereas children in Study 3 saw the verb presented by an experimenter on a video screen and by a character on a video screen. Thus, Study 1 used both live action and televised displays to present the verb, but Study 3 used two forms of video (an experimenter and a character) to demonstrate the verb. Further research is needed to disentangle these factors, but at minimum, we can conclude that a video medium alone is insufficient for young children's verb learning.

Finally, while the current research directly tests verb learning from social interaction and video as well as from video alone, we did not examine children's potential to learn verbs from social interaction alone. This choice was made because of the copious research that has shown that children *can* and *do* learn entire languages (including verbs) based on social interaction alone. Thus, we presume that children in a "live only" condition would be able to learn verbs better than with any combination of live action and video (e.g., Childers & Tomasello, 2002; Tomasello & Akhtar, 1995).

On the Nature of Word Learning

The results of this study are relevant to at least two disparate literatures: the television viewing literature and the word learning literature. Our results suggest that children older than 3 years may be able to learn words from television programs, and that children as young as 30 months may also learn words when provided with (live) social support. However, our results do *not* speak to the utility of word learning through watching television for children younger than 30 months of age. In fact, pilot data for the current study indicated that children under the age of 30 months were *unable* to learn a verb from video, *even with live social support*. Several studies in the television literature may explain why children do not learn from videos at 30 months but do by 36 months. For example, a naturalistic observational study of children viewing a commercial television program revealed that children younger than 30 months only glanced occasionally toward the video display. By 30 months, however, children began to intentionally watch the video screen (Anderson &

Levine, 1976). Although the participants in this study were younger than the target audience for the television program, the current research suggests that even with clips from age-appropriate programming, children younger than 30 months have difficulty using video as a source of information. Thus, it is possible that at around 30 months, children begin to develop the cognitive capacities needed to effectively view television. Troseth and DeLoache (1998) suggest this is the case, noting that infants younger than 30 months are not readily able to understand video displays as a symbolic source of information. Although many studies have shown that even very young infants attend to televised displays (Anderson & Levin, 1976; Schmitt, 2001), the current research suggests that video may not be an effective *educational tool* until 30 months of age.

We found that children under the age of 3 required social support to learn a verb from video. While many educational programs encourage parents to view television with their children, our study is among the first to demonstrate the utility of learning verbs from video with even the mildest (and relatively non-contingent) social interaction. Moreover, the results of our studies also provide an important link between television research using older children and research using younger children. Until now, research exploring television and language in infants and toddlers had a void between 24 months and 3 years. If the ability to learn from video develops over time, this is a critical juncture. As the educational television market continues to target younger children, the facilitative effects of social support require continued investigation.

The current research also has several important implications for the word learning literature. First, the test of word learning used in these studies was much more demanding than the typical tests used in most studies of word learning that require only linking a word to a specific object or action. Rather than asking children to identify the exact agent/action pairing seen before, all of our test trials required children to *extend* their knowledge of a novel verb to a new agent. Additionally, this extension task required children to learn a verb from puppets during training and to transfer that to babies during test (and vice versa). Both younger and older children succeeded in this demanding task when provided with social interaction in addition to televised displays.

In addition to the extension test of word learning, we included a more stringent test. This stronger test not only asked children to generalize the learned verb to a novel character, but also tested the limits that children put on the meaning of the trained verb (e.g., Mutual Exclusivity; Markman, 1989; see also Golinkoff, Mervis & Hirsh-Pasek, 1994). Based on the principle of Mutual Exclusivity (Markman, 1989), we expected that children would look away from the trained verb when presented with an alternative verb, demonstrating their unwillingness to assign two verbs to one action. This pattern of looking would demonstrate robust word learning. We found evidence of Mutual Exclusivity in both age groups, but only when the video display was accompanied by live social support. Although some studies have used similarly extensive tests of word learning (Brandone et al., 2007; Golinkoff, Jacquet, Hirsh-Pasek & Nandakumar, 1996; Hollich et al., 2000; Pruden, Hirsh-Pasek, Golinkoff & Hennon, 2006), this method is not widely used in the field. We believe success in this stringent test of word learning is indicative of secure word knowledge.

A second contribution to the word learning literature comes from our evidence that children are able to learn *verbs* from video. Although verbs are the building blocks of grammar, researchers have only recently begun to study verb acquisition in children (Brandone et al., 2007; Gertner et al., 2006; Hirsh-Pasek et al., 1996; Poulin-Dubois & Forbes, 2002; 2006). Given that children much younger than 30 months of age are able to learn object words from video (e.g., Krcmar et al., 2007), it is quite surprising that children are unable to learn verbs from video (even with optimal support) until they are older. Although children have verbs in

their natural vocabularies from a very young age, the current research corroborates previous findings that verbs are *still* more difficult for children to learn than nouns (e.g., Bornstein et al., 2004; Gentner, 1982; Golinkoff et al., 1996). This research is a crucial step toward understanding language development in multiple contexts. Furthermore, by moving beyond the study of nouns, we begin to understand the complex process by which children become sophisticated speakers.

Conclusion

Taken together, these studies explore language learning in relatively uncharted territory by asking how children learn verbs from video with and without social support. Our results indicate that verb learning from video is possible for children older than 3 years, but remains challenging for children younger than three even under socially supportive conditions. Our results confirm prior findings that social interaction plays a critical role in word learning, and contributes to our understanding of young children's ability to learn from video. However, social interaction provides a wealth of information for children, and our research is only a first step toward pinpointing the *specific mechanism* within social support that makes a difference for children's word learning from video. Future research might further explore what subtle aspect(s) of social interaction make the key difference when an experimenter teaches a child live versus on video.

Considering that the prevalence of television programs marketed toward children aged 3 years and younger has been on the rise in recent years, our findings serve both as a boon and a warning. On the up side, 3-year-olds can learn verbs from video with or without social support. On the down side, children under age 3 *do not* appear to glean any verb learning benefit from watching videos, unless given social support. Although findings on learning from video are mixed, the preponderance of research in the language literature suggests that live social interaction is the most fertile ground for language development.

Acknowledgments

This research was supported by an NICHD grant 5R01HD050199 and by NSF grant BCS-0642529.

We thank Nora Newcombe for her valuable comments on earlier versions of this paper and we also thank Wendy Shallcross for her assistance in data collection.

References

- Anderson DR, Levin SR. Young children's attention to *Sesame Street*. *Child Development*. 1976; 47:806–811.
- Anderson DR, Pempek TA. Television and very young children. *American Behavioral Scientist*. 2005; 48:505–522.
- Baldwin D. Infants' contribution to the achievement of joint reference. *Child Development*. 1991; 62:875–890. [PubMed: 1756664]
- Baldwin, DA.; Tomasello, M. Word learning: A window on early pragmatic understanding. In: Clark, EV., editor. *The proceedings of the twenty-ninth annual child language research forum*. Chicago, IL: 1998. p. 3-23.
- Barr R, Muentener P, Garcia A, Fujimoto M, Chavez V. The effect of repetition on imitation from television during infancy. *Developmental Psychobiology*. 2007; 49:196–207. [PubMed: 17299795]
- Bornstein MH, et al. Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*. 2004; 75:1115–1139. [PubMed: 15260868]
- Brandone AC, Pence KL, Golinkoff RM, Hirsh-Pasek K. Action speaks louder than words: Young children differentially weight perceptual, social, and linguistic cues to learn verbs. *Child Development*. 2007; 78:1322–1342. [PubMed: 17650141]

- Brown, E. That's edutainment. New York, NY: McGraw Hill; 1995.
- Childers JB, Tomasello M. Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*. 2002; 38:967–978. [PubMed: 12428708]
- Conboy, BT.; Brooks, R.; Taylor Meltzoff, A.; Kuhl, PK. Joint engagement with language tutors predicts brain and behavioral responses to second-language phonetic stimuli. Canada. Poster presented at the 16th International Conference on Infant Studies, Vancouver; 2008 March.
- Fernald, A.; Zangl, R.; Portillo, AL.; Marchman, VA. Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In: Sekerina, IA.; Fernández, EM.; Clahsen, H., editors. *Developmental Psycholinguistics: On-line methods in children's language processing*. Amsterdam: John Benjamins; p. 97-135.in press
- Forbes JN, Farrar MJ. Learning to represent word meaning: What initial training events reveal about children's developing action verb concepts. *Cognitive Development*. 1995; 10:1–20.
- Garrison MM, Christakis DA. A teacher in the living room? Educational media for babies, toddlers, and preschoolers. 2005 The Henry J. Kaiser Family Foundation.
- Gentner, D. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In: Kuczaj, SA., editor. *Language development: Vol. 2. Language, thought, and culture*. Hillsdale, NJ: Erlbaum; 1982. p. 301-334.
- Gertner Y, Fisher C, Eisengart J. Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*. 2006; 17:684–691. [PubMed: 16913951]
- Golinkoff RM, Hirsh-Pasek K, Cauley KM. The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*. 1987; 14:23–45. [PubMed: 3558524]
- Golinkoff RM, Jacquet RC, Hirsh-Pasek K, Nandakumar R. Lexical principles may underlie the learning of verbs. *Child Development*. 1996; 67:3101–3119. [PubMed: 9071772]
- Golinkoff RM, Mervis CB, Hirsh-Pasek K. Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*. 1994; 21:125–155. [PubMed: 8006089]
- Hassenfeld, J.; Nosek, D., Producers; Clash, K, Co-Producer/Director. *Beginning together* [Motion picture]. New York, New York: Available from Sesame Workshop, One Lincoln Plaza; 2006a. 10023
- Hassenfeld, J.; Nosek, D., Producers; Clash, K, Co-Producer/Director. *Make music together* [Motion picture]. New York, New York: Available from Sesame Workshop, One Lincoln Plaza; 2006b. 10023
- Hirsh-Pasek, K.; Golinkoff, RM. *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press; 1996.
- Hirsh-Pasek, K.; Golinkoff, RM. *Action meets word: How children learn verbs*. New York: Oxford University Press; 2006.
- Hirsh-Pasek, K.; Golinkoff, RM.; Naigles, L. Young children's use of syntactic frames to derive meaning. In: Hirsh-Pasek, K.; Golinkoff, RM., editors. *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press; 1996. p. 123-158.
- Hollich, G.; George, K. Modification of preferential looking to derive individual differences. Vancouver, Canada. Poster presented at the 16th International Conference on Infant Studies; 2008 March.
- Hollich GJ, Hirsh-Pasek K, Golinkoff RM, With Hennon E, Chung HL, Rocroi C, Brand RJ, Brown E. Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*. 2000; 65 3, Serial No. 262.
- Kersten, AW.; Smith, LB.; Yoshida, H. Influences of object knowledge on the acquisition of verbs in English and Japanese. In: Hirsh-Pasek, K.; Golinkoff, RM., editors. *Action meets word: How children learn verbs*. New York: Oxford Press; 2006. p. 499-524.
- Krcmar M, Grela BG, Lin Y. 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. Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience*. 2004; 5:831–843.

- Kuhl PK, Tsao F, Liu H. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *PNAS*. 2003; 100:9096–9101. [PubMed: 12861072]
- Linebarger DL, Walker D. Infants' and toddlers' television viewing and language outcomes. *The American Behavioral Scientist*. 2005; 48:624–645.
- Maguire MJ, Hirsh-Pasek K, Golinkoff RM, Brandone AC. Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*. 2008; 11:628–634. [PubMed: 18576970]
- Marchman VA, Bates E. Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*. 1994; 21:339–366. [PubMed: 7929685]
- Markman, EM. Categorization and naming in children: Problems of induction. Cambridge, MA: MIT Press; 1989.
- Meints K, Plunkett K, Harris PL, Dimmock D. What is 'on' and 'under' for 15-, 18-, and 24-month-olds? Typicality effects in early comprehension of spatial prepositions. *British Journal of Developmental Psychology*. 2002; 20:113–130.
- Naigles LR, Bavin EL, Smith MA. Toddlers recognize verbs in novel situations and sentences. *Developmental Science*. 2005; 8:424–431. [PubMed: 16048515]
- Nielsen M, Simcock G, Jenkins L. The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science*. in press.
- Poulin-Dubois D, Forbes JN. Toddler's attention to intentions-in-action in learning novel action words. *Developmental Psychology*. 2002; 38:104–114. [PubMed: 11806694]
- Poulin-Dubois, D.; Forbes, JN. Word, intention, and action: A two-tiered model of action word learning. In: Hirsh-Pasek, K.; Golinkoff, RM., editors. *Action meets word: How children learn verbs*. New York: Oxford Press; 2006. p. 499-524.
- Pruden SM, Hirsh-Pasek K, Golinkoff RM, Hennon EA. The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*. 2006; 77:266–280. [PubMed: 16611171]
- Reiser RA, Tessmer MA, Phelps PC. Adult-child interaction in children's learning from "Sesame Street." *Educational Communication & Technology Journal*. 1984; 32:217–223.
- Rice M, Woodsmall L. Lessons from television: Children's word-learning while viewing. *Child Development*. 1988; 59:420–429. [PubMed: 3359862]
- Sabbagh MA, Baldwin D. Learning words from knowledgeable versus ignorant speakers: Link between preschoolers' theory of mind and semantic development. *Child Development*. 2001; 72:1054–1070. [PubMed: 11480934]
- Schmitt KL. Infants, toddlers, and television: The ecology of the home. *Zero to Three*. 2001; 22:17–23.
- Singer, JL.; Singer, DG. Barney & Friends as entertainment and education: Evaluating the quality and effectiveness of a television series for preschool children. In: Asamen, JK.; Berry, GL., editors. *Research paradigms, television, and social behavior*. Thousand Oaks: Sage Publications; 1998. p. 305-367.
- Swingley D, Pinto JP, Fernald A. Continuous processing in word recognition at 24 months. *Cognition*. 1999; 71:73–108. [PubMed: 10444905]
- Tomasello M, Akhtar N. Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*. 1995; 10:201–224.
- 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 socially relevant information. *Child Development*. 2006; 77:786–799. [PubMed: 16686801]
- Zimmerman FJ, Christakis DA, Meltzoff AN. Television and DVD/videoviewing in children younger than 2 years. *Archives of Pediatric and Adolescent Medicine*. 2007a; 161:473–479.
- Zimmerman FJ, Christakis DA, Meltzoff AN. Associations between media viewing and language development in children under age 2 years. *Journal of Pediatrics*. 2007b; 151:364–368. [PubMed: 17889070]

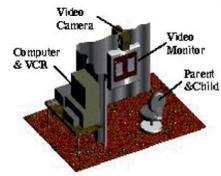


Figure 1.
The Intermodal Preferential Looking Paradigm (see Hollich, et al., 2000).

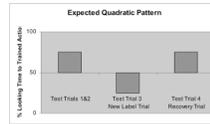


Figure 2.
Expected quadratic pattern for the Stringent Test of Verb Learning.

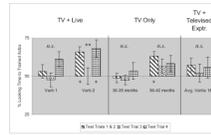


Figure 3. Percentage Looking Time to the Trained Action in Studies 1, 2, and 3. Error Bars Represent the Standard Error of the Mean.
 *denotes a significant percentage of looking time to target during individual test trials
 **denotes significant quadratic patterns
 n.s. denotes quadratic patterns which are not significant

Table 1

Sequence of Phases for Studies 1, 2 and 3.

Trial Type	Audio	Left side of screen	Center of screen	Right side of screen
Center ^a	Hey, look up here!	smiling baby		
Introduction phase	This is Cookie Monster. Do you see Cookie Monster?	Cookie Monster smiling		
Introduction phase	This is Cookie Monster. Do you see Cookie Monster?			Cookie Monster smiling
Center	Wow, what's up here?	smiling baby		
Salience phase	Hey, what's going on up here?	Baby <i>wezzling</i>		Baby playing on parent's leg
Center	Cool, what's up here?	smiling baby		
Training phase^b (4 presentations)	Look at Cookie Monster wezzling! He's wezzling! Cookie Monster is wezzling!	Cookie Monster wezzling		
Center	Find wezzling!	smiling baby		
Test trial 1	Where is wezzling? Can you find wezzling? Look at wezzling!	Baby <i>wezzling</i>		Baby playing on parent's leg
Center	Find wezzling!	smiling baby		
Test trial 2	Where is wezzling? Can you find wezzling? Look at wezzling!	Baby <i>wezzling</i>		Baby playing on parent's leg
Center	Now find glorping!	smiling baby		
Test trial 3 - new verb trial	Where is glorping? Can you find glorping? Look at glorping!	Baby <i>wezzling</i>		Baby playing on parent's leg
Center	Find wezzling again!	smiling baby		
Test trial 4 - recovery trial	Where is wezzling? Can you find wezzling? Look at wezzling!	Baby <i>wezzling</i>		Baby playing on parent's leg

^aCentering trials were 3 seconds in duration; every other trial was 6 seconds in duration.

^bIn Study 1, two of the Training Phase trials comprised the live action sequence and two trials used video clips. In Study 2, all four Training Phase trials were conducted via video clips. Finally, the Training Phase in Study 3 consisted of two video clips of an experimenter and two video clips from *Sesame Beginnings*.

Table 2

Description of the Actions, the Nonsense Verbs They Were Paired With, and the Closest English Equivalent.

Nonsense Word	English Equivalent	Description
Frep	Shake	Character moves object in hand from side to side rapidly
Blick	Bounce	Parent moves character up and down on knee
Wezzle	Wiggle	Character rotates torso and arms rapidly
Twill	Swing	Parent holds character in arms and rotates from side to side