Co-viewing supports toddlers' word learning from contingent and noncontingent video

CrossMark

Gabrielle A. Strouse [a,*], Georgene L. Troseth [b], Katherine D. O'Doherty [c], Megan M. Saylor [b]

[a] Division of Counseling and Psychology in Education, University of South Dakota, Vermillion, SD 57069, USA
[b] Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37203, USA
[c] Department of Psychology, University of Chicago, Chicago, IL 60637, USA

ARTICLE INFO

ABSTRACT

Social cues are one way young children determine that a situation is pedagogical in nature—containing information to be learned and generalized. However, some social cues (e.g., contingent gaze and responsiveness) are missing from prerecorded video, a potential reason why toddlers' language learning from video can be inefficient compared with their learning directly from a person. This study explored two methods for supporting children's word learning from video by adding social-communicative cues. A sample of 88 30-month-olds began their participation with a video training phase. In one manipulation, an on-screen actress responded contingently to children through a live video feed (similar to Skype or FaceTime "video chat") or appeared in a prerecorded demonstration. In the other manipulation, parents either modeled responsiveness to the actress's on-screen bids for participation or sat out of their children's view. Children then viewed a labeling demonstration on video, and their knowledge of the label was tested with three-dimensional objects. Results indicated that both on-screen contingency and parent modeling increased children's engagement with the actress during training. However, only parent modeling increased children's subsequent word learning, perhaps by revealing the symbolic (representational) intentions underlying this video. This study highlights the importance of adult co-viewing in helping toddlers to interpret communicative cues from video.

© 2017 Elsevier Inc. All rights reserved.

---

* Corresponding author.
E-mail address: gabrielle.strouse@usd.edu (G.A. Strouse).

Introduction

From infancy to adulthood, humans learn better when they believe that content is intentionally directed to them (Shafto & Goodman, 2008; Topál, Gergely, Miklósi, Erdohegyi, & Csibra, 2008). Across the preschool period, children use a growing range of behavioral and cognitive cues to determine a social partner's pedagogical intent, including pragmatic and semantic information during communication, as well as characteristics of the informant, such as the person's confidence in stating information (see Diesendruck & Markson, 2011, for a review). For very young learners, however, simple social cues presented during teaching are key. According to the "pedagogical stance" hypothesis (Csibra & Gergely, 2006; Gergely, Egyed, & Király, 2007), young children use social cues from adults as signals that content is being presented intentionally for them, shaping the way they attend to it and learn.

For example, during a toy play session, an adult offered infants a number of social cues—eye contact, gaze shifting, an engaging emotional expression, name referral, and talking in infant-directed speech about actions she wanted to demonstrate for them—or she offered comparable nonsocial actions. Then with only her arm visible, the adult showed how to use a hook to capture and pull a toy that was out of arm's reach. Infants previously exposed to the infant-directed social cues used the hook to grasp more out-of-reach toys than infants did who had experienced nonsocial actions (Sage & Baldwin, 2011). The social cues were given *prior to* the demonstration, setting up a pedagogical context for infants that carried forward; cues did not simply highlight action being simultaneously demonstrated but also signaled to children that information that followed was pedagogical in nature. Similarly, when an experimenter made eye contact and merely told 4-year-olds, "Look, watch this," children subsequently generalized the adult's actions more than if no prior pedagogical cues were offered (Butler & Markman, 2013).

In some situations, however, cues to pedagogy may necessarily be incomplete. As others have noted (O'Doherty et al., 2011; Richert, Robb, & Smith, 2011; Strouse, O'Doherty, & Troseth, 2013; Strouse & Troseth, 2014; Troseth, Strouse, Verdine, & Saylor, 2017), social cues from television are comparatively limited; the on-screen adult's gaze is not perfectly directed at viewing children, nor is it responsive to their gaze, verbalizations, or actions. A character on television does not use children's names, and bids to direct attention are not aligned with children's attention status. The absence of such cues may influence very young children to interpret prerecorded video formats as nonintentional situations (similar to the nonpedagogical comparison conditions used in research) and, thus, irrelevant for learning.

Indeed, video interventions designed to teach vocabulary have been largely unsuccessful (DeLoache et al., 2010; Krcmar, 2011; Robb, Richert, & Wartella, 2009; Vandewater, 2011). Toddlers can learn individual words from video when tasks are simple, such as when an object isolated in a close-up is repeatedly labeled in a voiceover (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Schafer & Plunkett, 1998; Scofield, Williams, & Behrend, 2007; Werker, Cohen, Lloyd, Casasola, & Stager, 1998), but they struggle with more challenging tasks requiring memory and the use of referential social cues (Strouse & Troseth, 2014; Troseth et al., 2017). The relative inefficiency of toddlers' language learning from video compared with face-to-face learning is especially striking (Krcmar, 2011; Krcmar, Grela, & Lin, 2007; Troseth et al., 2017).

Cues to pedagogy might be especially important for video-based learning because of young children's difficulty in understanding the symbolic (or representational) nature of two-dimensional screen images. In previous research, when toddlers were asked to apply information offered on video to a current real situation (e.g., imitation, object retrieval), they usually did not succeed (e.g., Barr, 2010; Barr & Hayne, 1999; Troseth, 2010; Troseth & DeLoache, 1998). For instance, after children watched as a person on the screen (whom they had met) hid a toy in a room where they had just played, they did not seem to realize that the on-screen event represented, and provided information about, the location of the hidden toy (Troseth & DeLoache, 1998).

Why do very young children lack "representational insight" (DeLoache, 1995) regarding this very iconic realistic type of image? One contributor may be the many ways in which video images can relate to reality (Troseth, 2010). Video can represent a real current event happening in the vicinity,

but it can also show a real event that is spatially distant or that occurred in the past. In addition, video often represents imaginary events—both ones that look very real (e.g., computer-generated images of dragons interacting with people) and ones that are clearly imaginary (e.g., cartoons of talking, clothes-wearing animals). Children observe cars and elephants on the screen apparently charging through the living room—with no effect on the real environment. As the youngest viewers try to make sense of how screen images relate to the real world, their conservative initial response may be to separate video from reality in the same way as they are thought to mark off pretend identities as separate from their real-world concepts (Harris & Kavanaugh, 1993; Leslie, 1987; Troseth, Pierroutsakos, & DeLoache, 2004). For this reason, very young children might not spontaneously realize that people on video can serve an information-providing role, and pedagogical cues may be important to help toddlers realize that a particular video is intended to communicate information.

Compared with watching a prerecorded television show, "video chat" over Skype or FaceTime (e.g., with distant grandparents or parents on business trips) offers two distinct opportunities for adults to provide social cues to support children's learning. An on-screen partner responds contingently to children's comments and actions, offering one set of supportive cues; often, a co-watching parent interacts with the person on the screen, offering other cues about the relevance of on-screen events. The in-person co-viewer may help children to interpret the relation between the on-screen events and real life (McClure & Barr, 2017).

Social cues from on-screen individuals and in-person co-viewers are paired not only in real life but in research as well. In several prior studies (Myers, LeWitt, Gallo, & Maselli, 2016; Nielsen, Simcock, & Jenkins, 2008; Roseberry, Hirsh-Pasek, & Golinkoff, 2013; Troseth, Saylor, & Archer, 2006), closed-circuit feeds or video chats have provided support for learning. All of these studies used a training phase to familiarize children with live video prior to presenting the information to be learned. Training interactions contained several important social cues; the on-screen actress used the child's name, had a personal conversation with the child (e.g., "I heard you had a birthday party yesterday . . ."), and referred to items in the child's environment (e.g., "There's a sticker under your chair!"). In two of the studies, parents sat alongside their children, conversing with the on-screen person and encouraging their children to participate (Nielsen et al., 2008; Troseth et al., 2006). Parents' presence and actions were intended to help establish the video's contingency for the children; however, they also had the potential to act as cues that the object of shared focus (the actress on video) was engaged in teaching. In the most recent studies (Myers et al., 2016; Roseberry et al., 2013), parents did not talk with the actress but were involved in other ways that may have influenced children's interpretation of the viewing situation toward learning rather than entertainment. In one, parents retrieved toys requested by the actress (Myers et al., 2016). In both, children sat on their parents' laps during the interaction (Roseberry et al., 2013). In the current study, we took a closer look at the individual and joint contribution of on-screen and in-person social cues to children's learning of information (a new object label) from video.

On-screen social cues

To establish that intentional teaching is taking place, contingent responsiveness between a teacher and learner is expected, including communicative cues such as making eye contact, using infant-directed speech, taking turns, calling the child by name, and reacting to what the child does (Gergely et al., 2007)—cues that are possible via FaceTime and similar video-chat programs but are not typically available in prerecorded video. These cues may help children to treat information provided by the on-screen adult as intentional and "to be learned" and motivate them to engage in processing and learning the presented content.

In contrast, interactive characters in online and tablet games and other media may offer social cues in a pseudocontingent way. In shows like *Blue's Clues,* an on-screen character addresses the audience in child-directed speech while looking squarely into the camera (the general direction of the child), often pausing as if listening for a response and commenting in a generic way but with no ability to respond to children personally. Infants are sensitive to the minor mismatches in interpersonal timing caused by a 1-s delay inserted in closed-circuit video (Striano, Henning, & Stahl, 2006), so for toddlers a lack of authentic contingency from prerecorded actors may serve as a cue *against* pedagogical intent.

Thus, in the current study, we compared learning after exposure with authentic live-feed video chat or a pause and wait–style prerecorded video. In the research reported here, we varied contingency but attempted to keep all other social aspects (e.g., apparent eye contact via the camera, smiling, use of infant-directed speech) the same across conditions.

Co-viewer social cues

An in-person co-viewer can also offer social-communicative cues to pedagogy. When watching prerecorded or live videos, co-viewing parents may provide social cues typical of everyday learning situations—looking from children to the screen (joint attention), verbal direction of attention, comments or questions about screen contents, and other behaviors. Thus, physically present adults may be able to establish a learning situation around video much like adults do with books.

Research on co-viewing and active mediation has focused on children aged 3 years and older. A number of studies demonstrate that adult mediation of older preschoolers' television viewing can enhance learning of both prosocial and educational content (Friedrich & Stein, 1975; Reiser, Tessmer, & Phelps, 1984; Reiser, Williamson, & Suzuki, 1988; Singer & Singer, 1998; Strouse et al., 2013; Watkins, Calvert, Huston-Stein, & Wright, 1980). In reporting preschool children's increased learning from storybook videos, Strouse et al. (2013) suggested three mechanisms (attention-directing, cognitive, and social feedback components) to explain the positive effects of adult co-viewing. However, a search for similar research with children under 3 years of age reveals relatively few studies.

For infants, parental cues during co-viewing do appear to be associated with changes in the way they *watch* videos; infants follow their parents' gaze to the screen, and infant looks preceded by parent looks tend to be longer in duration (Demers, Hanson, Kirkorian, Pempek, & Anderson, 2013). Infants also pay more attention to the screen when parents ask questions and provide labels (Barr, Zack, Garcia, & Muentener, 2008) or join in shared focus and turn taking (Fidler, Zack, & Barr, 2010). Evidence that co-viewing has an impact on infants' and toddlers' *learning* is less clear. When parents were instructed to repeatedly co-view videos with their babies, this supported some learning of the content in one study (Dayanim & Namy, 2015) but not in another (DeLoache et al., 2010). Slightly older children (2-year-old toddlers) did benefit from parents' support; children of this age struggled to learn individual words from researcher-created video clips, but they succeeded when parents provided a simple scaffold indicating that real objects in the room corresponded with those labeled on television (Strouse & Troseth, 2014). The inconsistent impact of parental co-viewing in these studies may indicate that optimal support varies with the age of the children or that *the way* in which parents co-view is of primary importance. In a study where infants were tasked with transferring information from a touch screen to a real toy, infants whose parents interacted in a well-structured and emotionally responsive manner while providing a variety of verbal information were more likely to transfer than infants whose parents interacted with lower levels of structure, variety, and responsivity (Zack & Barr, 2016).

A question raised in prior research is whether the key component of co-viewing is *what* parents say or something more implicit—that parents sharing attention and responding to information on the screen encourage children to put forth the effort to process the content (Strouse & Troseth, 2014; Strouse et al., 2013). One possibility is that nonverbal parental *social* cues (establishing joint attention) operate separately from verbal *cognitive* supports such as asking children questions (prompting them to retrieve, rehearse, and apply video-based content to new situations; see Strouse et al., 2013). Parents frequently use social cues while teaching (Gergely et al., 2007) such as when demonstrating a skill for their children or reading a picture book together; therefore, parents' use of social cues may signal to children that a co-viewed video is providing important relevant information that should be learned.

Study design and rationale

In the current study, we focused on the *social-pedagogical* role—how parents' modeling of attention, interest, and responsiveness to a video sets up a learning situation for young children. In our unique

manipulation, some parents were asked specifically to model participation *without* providing verbal prompts, thereby eliminating any explicit cognitive support that would be offered via verbal content. We hypothesized that parents modeling social responsiveness to the actress would provide sufficient cues to establish the video as pedagogically important to their children, supporting learning and transfer of the subsequent video content. We examined the effect of this aspect of co-viewing along with on-screen contingency as two separate social cue supports for toddlers using a factorial design to establish the role of each as well as their combined influence on learning. We predicted that the combination of on-screen contingency and parent modeling during training would establish a pedagogical situation, and children would learn and transfer subsequent new information presented on the screen. We also predicted that because both manipulations individually provide pedagogical cues, each would separately support children's acquisition of a new object label compared with the group who received neither support.

In prior studies (Myers et al., 2016; Nielsen et al., 2008; Roseberry et al., 2013; Troseth et al., 2006), video interactions were designed to establish for children that the on-screen actress was indeed responding contingently to them; she knew who they were (used their names); she was purposely communicating with them (made "eye contact" and used child-directed speech); she could see what they did (appropriately waited or responded to their actions); and she could give feedback accordingly. The video interaction in the current study was similarly designed to establish the contingency of the actress's responses to the children but differed in that she did not use the children's names or refer to prior events in the children's lives as pedagogical cues. This change helped us to develop a fairer test of the value of contingency; in previous studies, recordings of live-feed sessions were shown to other children in a "yoked" control group, who saw a video that, therefore, was both noncontingent (in response to another child's actions) and personally inaccurate (it included the other child's name and personal details). Removing these references avoided presenting inaccurate information to children in our noncontingent (prerecorded video) group, allowing us to center our comparison on the presence or absence of authentically contingent social cues.

In keeping with our focus on the importance of on-screen and in-person social cues for learning from video, we chose a challenging word-learning task requiring that children make use of the on-screen speaker's referential social cue to disambiguate the meaning of her novel label. When the label was offered, the target object was out of sight in an opaque bucket, whereas the distracter was visible in a transparent bin. The presence of a visible distracter required children to use the speaker's referential social cue—gaze into the opaque container at an absent referent—to infer the identity of the labeled object while not connecting the label with the visible object (see Baldwin, 1991, 1993). Therefore, it was vital that children recognize that the on-screen person was providing an intentional reference, an interpretation that might be helped by children viewing the labeling event as part of a pedagogical situation.

To avoid unintentional bias, another researcher who was blind to which object had been labeled tested children's word learning. In prior research, toddlers imitated from a display without pedagogical cues more for an adult present during the initial demonstration than for a person who had not been present (Vredenburgh, Kushnir, & Casasola, 2015). Thus, the actress who had labeled the objects on the screen was also present in the room during testing to encourage children to display their knowledge of the word she had demonstrated regardless of condition.

We tested children's word learning using three-dimensional objects, including a generalization set painted different colors from those depicted on the screen. Compared with having children point at an object pictured on the screen, asking them to choose between real objects is a more stringent test of word learning. To consistently succeed at our test, children needed to understand that the label *modi* applied not only to the on-screen object (simple association) but also to similar objects appearing in the real world (transfer and generalization). The ability to extend a novel word to a new instance of a referent indicates greater flexibility and understanding of the label (Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009; Werker et al., 1998).

Method

Participants

A sample of 88 children (44 boys and 44 girls) between 28.2 and 32.3 months of age (M = 30.3 months, SD = 0.95) and their parents were recruited from state birth records via telephone, drawn from a major metropolitan area in the southern United States. Children were excluded from participation if they were primarily exposed to a language other than English or had cognitive or developmental delays or hearing loss. Data from an additional 4 children were excluded from analysis due to experimenter error (n = 2) or uncooperativeness (n = 2).

Participants identified their race as White (92%), African American (3%), or multiple races (5%). They self-identified as Hispanic (6%) or non-Hispanic (94%). The average participating parent had a bachelor's degree, and participants were middle to upper-middle class (average household income in the \$50,000–\$75,000 bracket; area median \$46,686). There were no differences across conditions in parent education or income.

Materials

Questionnaires

Parents were asked to complete the short-form toddler version of the MacArthur–Bates Communicative Development Inventory (CDI; Fenson et al., 2000) by marking the words their children said on the 100-item checklist. Because our age range spanned the break between the Level 2 and Level 3 forms, we opted to use Level 2 raw scores for all children. Parents also were asked to complete a questionnaire on which they indicated their children's experience with video and video chat as well as demographic variables.

Training and labeling demonstration video

Our stimulus video had two main components: a video interaction (training), in which an on-screen actress engaged the child in songs and games, and a labeling demonstration, in which the actress followed a scripted procedure for introducing novel objects and a novel label. The video interaction was loosely based on the one used by Troseth et al. (2006) and adapted by others (Nielsen et al., 2008; Roseberry et al., 2013), but general questions in a pause-and-wait format were used instead of personally specific questions. In addition, the actress maintained focus on directly addressing the child only (not conversing with the parent) because parent social cues were manipulated separately. This interaction with the actress occurred on a live, closed-circuit (contingent) video chat for half of the children and on a prerecorded (noncontingent) video for the others, both shown on a 17-inch television monitor. The labeling demonstration that followed was identical to that used by Strouse and Troseth (2014) and Troseth et al. (2017) and was the same across conditions. Thus, cues to set up the situation as pedagogical were offered before the phase when the to-be-learned information was presented.

Familiar and novel objects

A stuffed Big Bird toy (24 cm tall) was used during the video interaction, and a set of familiar objects (a plastic turtle, frog, truck, and boat) was used during the familiar objects training. Two novel objects were labeled during the labeling demonstration: a hook and a mop holder (see Fig. 1); these were also used during the final testing along with an additional mop holder and hook that differed from the originals in color. A tube decorated with stickers and a PVC (polyvinyl chloride) pipe slide were available for children to put their selected objects in during testing so as to keep children engaged with the procedure.

Design and procedure

Children were assigned to view either a video-chat (contingent) version of the video interaction or a prerecorded (noncontingent) version. This *on-screen contingency* factor was crossed with the pres-
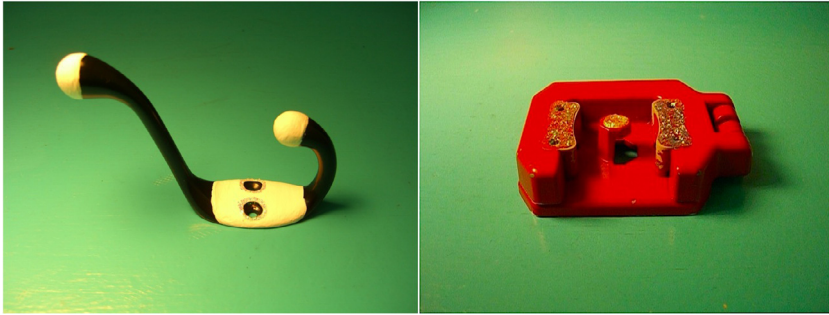
Fig. 1. Novel objects labeled by the actress. The object that was the target of labeling was counterbalanced across participants.

ence or absence of a *modeling* parent, yielding four groups of 22 children (11 boys and 11 girls). In the two modeling conditions, parents were asked to model the behaviors requested by the actress during the video chat or prerecorded training.

Warm-up

One researcher, the *tester*, warmed up with the child by playing on the floor with the novel and familiar objects, tube, and slide. During warm-up, the tester made sure that the child agreed with the names of all the familiar objects while not labeling the novel objects. The parent was asked to confirm that the child would not already have names for the novel objects. While the child was playing, a second researcher (the *actress*) explained the study procedures and paperwork to the parent, who was sitting near the child but did not interact directly with the child. A third researcher (the *assistant*) was also present in the room.

Familiar objects training

Once the child was comfortable, he or she was seated at a child-sized table across from the tester. The parent sat to the left of the child in a chair that faced away from the tester and was instructed to continue filling out the paperwork. The assistant sat to the right of the child to help keep him or her focused. The actress remained in the room but sat away from the table. As a "training" game, the tester held up a familiar object (frog or turtle) in each hand and exclaimed, "I have a frog! Show me the frog!" She then extended the objects toward the child and prompted, "Pick the frog!" Children who chose the frog were asked to put it in the tube or chute. If children chose the turtle, the tester held on to the objects, pulled back her hands, and prompted, "That's not the frog! Pick the frog!" If children attempted to grab both objects, she said, "Just pick one! Pick the frog!" This was repeated until the child chose the correct object. Once the child put the frog in the chute, he or she was offered the turtle. For the second trial, the tester switched the hand in which she held each of the familiar objects. She followed with 2 trials using the truck and boat. The child's behavior during the training and the rest of the procedure was filmed by a video camera for later coding.

Video training

After the child learned to play the game, the tester's chair was replaced with a television on a rolling cart. The actress left the room along with the tester, taking the novel objects with them. The assistant remained in the room with the parent and child. Parents in the *modeling* conditions were instructed to turn their chair to face the screen so that they could participate by modeling the behaviors called for by the on-screen actress (e.g., touching their shoulder, singing along). The parent was asked not to direct the child to participate but instead to set an example of participation so that the child's participation was not in response to direct instruction from the co-viewing parent but rather was in response to the video actress. The parent did not converse with the video actress, and any conversations with the child were generally limited to keeping the child in his or her chair (this was similar across conditions). Parents in the no-modeling conditions remained turned away from the

video, occupied with their paperwork. The parent was told that we wanted to see what the child would learn "on his/her own."

The actress appeared on video (either via video chat or prerecorded) and greeted the child without using his or her name by asking, "Hi! Here I am on TV! Can you see me?" The actress then introduced the Big Bird toy. As she put it on her own shoulder, arm, and head, she encouraged the child to play along by pointing to the corresponding place on his or her body. She then sang *If You're Happy and You Know It,* stopping after each line to encourage the child to clap and sing along. Finally, she led a game of *Simon Says,* asking the child to perform three different actions. In the parent modeling conditions, parents "played along" by following the actress's instructions themselves, thereby setting an example of listening and responding to the person on the screen.

One researcher served as the actress for 80% of the children across the four conditions, and three others portrayed the actress for a few children each. In all conditions, the actress followed the same script of modeled behaviors, prompts, and pauses for response, with the exception that as part of the manipulation of *on-screen contingency,* children watching video chat received corrective feedback and children watching prerecorded video did not. The contingent actress gave recognition of children's specific actions, adding, "I can see you patting your head!" or "I can't see you, can you [repeat request for response]?" Reprompts like the latter example were given in the contingent conditions up to two times (see "Scoring" section below). In contrast, following a prompt and pause, our prerecorded actress always gave either generic positive feedback (e.g., "Good job!") or neutral feedback (e.g., "OK!"), similar to that provided in many children's television shows. Any apparent pseudocontingency was a coincidence not based on the child's specific actions.

In addition, the on-screen actress gave children in the contingent conditions redirection for off-task behavior; if children got out of their chair and ran out of view of the chat, the actress called them back to their seat and waited for them to return. In the noncontingent conditions, the actress did not pause; instead, the assistant and/or parent present in the room needed to call children back to their chair.

Labeling demonstration

After the video interaction, the actress began the labeling demonstration, which was the same regardless of condition. Even in the contingent conditions, the actress no longer responded contingently to the child and instead proceeded according to the script, matching the prerecorded version of the demonstration. Because the actress now requested no actions from viewers, parents in the modeling conditions sat quietly and watched alongside their children. Thus, as in Sage and Baldwin's (2011) study, social cues were used to support children's interpretation of the situation as pedagogical *prior to* the teaching demonstration; the only co-viewer cue that remained was parents' gaze at the screen.

During the labeling demonstration, the actress looked at the child and smiled at scripted times. She said, "Now I have some things to show you," while placing two novel objects (a mop holder and hook) and two containers on the table. She held up and commented on each object, then placed the target object in an opaque bucket and the distracter in a clear bucket. Next she tilted each bucket, looked inside, and made four statements about the object (e.g., "I see a modi!", "I see this one"). Thus, when the label was offered, the target was out of sight in the opaque bucket, whereas the distracter was visible in the transparent bin. Finally, the on-screen actress took the objects out of the buckets, placing them briefly on the table so that the child could see them, before putting them away. All of the speaker's behavior toward the two objects, as well as her amount of talking about each object, was carefully matched. During the labeling demonstration, the tester waited in the hall, blind to which object was labeled the "modi."

Test

After the video ended, the actress and the tester (who remained blind to which object had been labeled) returned to the room. The tester sat across from the child. She gave the child the pair of novel objects to handle for about 30 s, then retrieved and held them up (one in each hand), saying, "I have a modi! Show me the modi!" She extended them forward and prompted, "Pick the modi!" If the child made a choice, he or she was prompted to put the object into the chute or tube. If the child attempted to grab both objects, he or she was reprompted to choose only one. As in training, the child was then

given the other object to put down the chute/slide. The trial was repeated with the objects on opposite sides. The final 2 trials (label generalization test) consisted of offering the same two objects painted different colors from the first set.

Scoring

Participation in video training interaction

Two coders reviewed children's videotapes for their verbal or nonverbal responses to the 8 specific prompts given by the actress during the interaction. Children in the noncontingent conditions received exactly these 8 prompts; however, children in the contingent conditions received the same 8 prompts plus a varying number of reprompts (up to 2 for each initial prompt) as a result of our manipulation. Thus, we coded two types of responsiveness: (a) an *initial response* score (1 point for each of the 8 prompts children responded to verbally or nonverbally immediately after hearing the initial prompt) and (b) a *total response* score (1 point for each prompt children responded to immediately or after reprompts). Significant differences in initial and total response scores were found only in the contingent conditions in which children heard additional prompts; however, small differences in the scores of the prerecorded video groups occurred due to occasional delayed responses children offered after the actress had moved on to the next prompt. The initial response score provides a measure of participation in response to comparable prompts across all four conditions, whereas the total response score indicates children's raw amount of participation across conditions. The intraclass correlation coefficient for the two raters was .92 for initial response and .98 for total response. Due to a videotape problem, 1 child's video was scored by only one coder and was not included in the reliability analysis.

Attention during labeling

Two coders reviewed the videotapes and scored children's attention during the labeling demonstration portion of the video, which varied in duration (ranging from 39 to 62 s; M = 47 s, SD = 4). Attention was recorded as the percentage of time children's eyes were focused on the television screen during the demonstration. Reliability for the two coders was measured using a mixed-effects model with participants as a random factor and raters as a fixed factor. This analysis yielded an intraclass correlation coefficient of .87.

Familiar and novel object tests

Children's object choices for each trial were recorded during the session by a researcher who was present in the room. Children received credit for the first item touched after they were prompted during each trial, for a total score out of 4 for both familiar and novel objects. If children touched both items simultaneously, they were reprompted until they chose only one item. A second coder who was blind to condition scored each child's novel object choices from videotape (except for the child whose tape was damaged). The intraclass correlation for the two coders was .98.

Results

There was a small but significant difference in mean age among the four conditions (prerecorded no modeling: M = 29.6 months, SD = 0.96; prerecorded modeling: M = 30.7 months, SD = 0.97; video chat no modeling: M = 30.4 months, SD = 0.76; video chat modeling: M = 30.6 months, SD = 0.71). As a result, age was entered as a covariate in our analyses.

Questionnaires

There were no group differences in children's raw CDI scores (M = 72.9, SD = 19.9) and no correlation between CDI score and performance on the word-learning task (r = −.02), so vocabulary score was dropped from further analyses.

There was a significant condition difference in the hours of television parents reported that their children watched; those in the video chat groups regularly watched more television per week (*Med*

= 6–10 h) than those in the prerecorded groups (*Med* = 1–5 h), Mann–Whitney *U* = 723.0, *p* = .035. The correlation between hours watched and word learning was nonsignificant (r_s = .17). Because group differences did exist, we included this factor as a covariate in our analyses. In addition, we ran our analysis of word learning both with and without viewing experience as a covariate (which did not change the significance of results).

Data on prior video-chat use were collected from 66 of the 88 parents in the sample. For those parents who did answer the question, there were no condition differences in the number of times children had used a webcam (e.g., video chat). According to parents' responses, 28% of the children had frequent prior experience (*n* = 19; had used more than 11 times), 38% had no prior experience (*n* = 25), and the remaining 33% had infrequent exposure. There was no correlation between video-chat frequency and children's word-learning performance (r_s = .01).

Participation in video interaction

Children in the contingent/modeling group (who received the most social cues from the responsive video actress and from parent modeling of requested behaviors) had the highest participation response scores, followed closely by children in the noncontingent/modeling group and the contingent/no-modeling group (Table 1).

A 2 (On-Screen Contingency) × 2 (Modeling) analysis of covariance (ANCOVA) with age and prior television viewing as covariates yielded a main effect of on-screen contingency, $F(1,81)$ = 4.93, *p* = .029, η^2 = .053, and a main effect of modeling, $F(1,81)$ = 4.87, *p* = .030, η^2 = .052, on children's *initial* responses to the requests of the person on video (see Table 1 for mean participation scores). There was no significant interaction and no significant effect of either covariate (age or television viewing) on children's initial participation. The same pattern of results was found in an analysis of children's *total* response score; significant main effects of on-screen contingency, $F(1,81)$ = 17.28, *p* < .001, η^2 = .160, and modeling, $F(1,81)$ = 5.27, *p* = .024, η^2 = .049. Bonferroni-corrected follow-up comparisons indicated that children in all three cue-supported conditions had significantly higher total response scores than children in the noncontingent/no-modeling group. There was no significant interaction and no significant effect of either covariate.

Although there were group differences in children's interactive responses, correlation analyses indicated that children's initial and total response scores were not related to their word learning (*r*s = .01 and −.01).

Attention to labeling demonstration

A 2 (Presentation Type: video chat vs. prerecorded) × 2 (Modeling: parent modeling vs. no-modeling) ANCOVA with age and prior viewing experience as covariates revealed no significant group differences in the amount of attention children paid to the labeling demonstration (group means ranged from 83% to 92%). However, there was a significant correlation between attention paid to the video and children's word-learning success, $r(88)$ = .22, *p* = .039. In addition, there was a small but significant difference in the total duration of the labeling demonstration, with the noncontingent (prerecorded) groups getting a slightly longer demonstration (*M* = 48 s, *SD* = 4) than the contingent (video chat) groups (*M* = 45 s, *SD* = 2), $F(1,81)$ = 7.57, *p* = .007, η^2 = .076. There was no correlation between labeling

Table 1
Means (and standard deviations) of children's participation scores.

Measure	Prerecorded		Video chat	
	No modeling	Modeling	No modeling	Modeling
Initial response	1.50 (2.56)	4.00 (2.91)	3.86 (2.48)	4.68 (2.61)
Overall response	1.73 (2.56)	4.64 (3.05)	5.64 (2.65)	6.18 (2.15)

Note. Scores were calculated as the number of experimenter prompts (out of 8) to which children responded. Initial response score: prompts responded to immediately. Overall response score: prompts answered initially and after up to 2 reprompts. In a repeated-measures analysis of variance, there was a significant difference between initial and overall response scores in the video chat conditions only, $F(1,12.71)$ = 15.93, *p* < .001, η^2 = .84.

Table 2

Means (and standard deviations) of children's scores on the familiar items training task and the novel items word-learning task.

Task	Prerecorded		Video chat	
	No modeling	Modeling	No modeling	Modeling
Familiar items score (of 4)	3.59*** (0.59)	3.77*** (0.75)	3.55*** (0.74)	3.50*** (0.51)
Word learning score (of 4)	2.41 (1.53)	2.73* (1.58)	1.41 (1.74)	3.45*** (0.80)

Note. Asterisks indicate scores significantly above chance (2 of 4).
** $p < .01$.
* $p < .05$.
*** $p < .05$.

duration and word learning ($r = .15$). To account for any potential effects of attention or duration in word learning, our analysis of word learning was done both with and without these covariates.

Familiar objects test

This test ensured that children understood how to respond to the comprehension questions. Only 4 children needed to be corrected on more than 1 familiar objects trial, and all groups responded to the familiar items prompts at above chance levels, indicating that children understood how to play the game (see Table 2 for mean scores). A 2 (On-Screen Contingency) × 2 (Modeling) ANCOVA with age as a covariate indicated that there were no group differences in children's selection of the correct familiar items (e.g., frog, boat).

Word learning test

Children's total scores across the 4 test trials are also included in Table 2. Children in the two groups with parent modeling performed significantly above chance [contingent/modeling: $t(21) = 8.52$, $p < .001$, $d = 1.81$; noncontingent/modeling: $t(21) = 2.16$, $p = .042$, $d = 0.46$]. Scores of the children in the two groups whose parents did not co-view did not differ from chance. Even when the on-screen actress responded contingently to the children but the parents did not model responding to the actress's prompts, children did not reliably learn the information she offered.

In addition to tests against chance, we tested for condition differences using an ordinal logistic generalized estimating equation (GEE). We entered children's total score on the 2 learning trials (in which children were tested using the exact items from the video) and on the 2 generalization trials (in which the objects differed in color) as a repeated measure. Because we expected that there might be more variability in children's generalization scores than in their learning scores (due to the generalization test objects' decreased similarity to the objects presented in the video), we used an autoregressive covariance structure for this analysis. On-screen contingency, modeling, the interaction between these factors, and test trial set were included as between-participant factors. There was a main effect of modeling, $\chi^2(1) = 14.56$, $p < .001$, with children in the parent modeling conditions outscoring those in the conditions without modeling. There was also a significant interaction between modeling and on-screen contingency, $\chi^2(1) = 6.55$, $p = .01$. Follow-up simple effects analyses indicated that in the presence of parent modeling, there was no effect of on-screen contingency, $\chi^2(1) = 2.64$, $p = .10$ (children generally scored well). However, in the absence of parent modeling, children who experienced on-screen contingency identified the correct object *less* often at test compared with those who watched the noncontingent video, $\chi^2(1) = 4.26$, $p = .039$. There was no main effect of on-screen contingency and no effect of test set; that is, children who did well on the initial word learning task also generalized the label to the object set that differed in color from the original objects. A model that included age and total video viewing hours as covariates provided similar results.

Discussion

We hypothesized that social cues provided by a contingently responsive on-screen actress and by a co-viewing parent both would support children's subsequent acquisition of a new object label pre-

sented on video. Our results indicated that both manipulations were effective at increasing children's *participation* with the on-screen actress during the training interaction, but only parent modeling was sufficient to support subsequent *word learning* in our brief, yet challenging, learning task. Parent modeling of responsiveness to the on-screen speaker seemed to legitimize the video as a source of new language and was particularly effective when social cues were also provided through on-screen contingency.

In this study, social cues either from a responsive person on video or from parent modeling incited children's participation. Children in the contingent/no-modeling group responded to the contingent actress's on-screen prompts during training at similar rates to children in the two parent modeling groups. We interpret this level of participation as resulting from the contingent social cues (including both positive and corrective feedback) introduced by the on-screen actress as part of our manipulation. This result is consistent with the findings of a recent study in which infants aged 6–24 months were observed interacting with grandparents using FaceTime. Infants whose on-screen grandparents were more sensitive (more responsive to the infants, with better timing and flexibility) and engaged in more instances of joint visual attention with them paid more overall attention during the video call (McClure, Chentsova-Dutton, Holochwost, Parrot, & Barr, 2017). Thus, contingent on-screen social cues were associated with increased child engagement. In addition, Crawley, Anderson, Wilder, Williams, and Santomero (1999) noted that repeated viewing of the same episode of the program *Blue's Clues* led to increases in preschoolers' participation with the program as well as increased comprehension of the episode. Increased participation over repeated viewings may give children more opportunity to learn from video if their attention would otherwise have been lost. However, in the current study, attention to the labeling demonstration was equivalent across groups, emphasizing that differences in learning resulted from processes other than attention alone.

Some research suggests that children "get to know" consistent characters featured in television programs and other media, forming a connection that has been termed a "parasocial" relationship (Horton & Wohl, 1956; Howard Gola, Richards, Lauricella, & Calvert, 2013). Children may perceive on-screen people or characters to be like themselves, increasing their motivation to learn from them (Lauricella, Howard, & Calvert, 2010). Increased participation with on-screen characters may help children to build these relationships over time, supporting children's learning from familiar characters even from new program episodes. Future research could address whether social cues given by co-viewing parents could "carry over" for children to other video-viewing situations, effectively establishing pedagogical intent for a particular on-screen person or ask and wait–style program.

In our study, the increased engagement that resulted from the presence of authentically contingent social cues translated into learning only when these cues were provided by an in-person co-viewer, not when they were provided on the screen. Children who had received the on-screen contingency cues in the absence of parent modeling learned *fewer* words than those who received no cues at all. One possible explanation for this pattern was the difficulty of our word-learning task; the distracter object was visible in the translucent bucket at the time of labeling, whereas the target object was not visible. The contingent videos attracted high levels of engagement, but perhaps when contingency was not paired with the co-viewer's cues to establish pedagogical intent, this engagement did not lead to deep processing. Children who did not learn the word may have instead selected the object with which they had more visual exposure (the distracter). However, because neither group demonstrated reliable above-chance learning in the absence of parent modeling, we do not believe that the difference between the two no-modeling conditions is particularly meaningful.

Despite the lack of *verbal* content from parent modelers (which we eliminated to control for cognitive support), the nonverbal social cues provided by parents, such as gaze that shifted between the children and the screen, and actions contingent with the actress's bids, were quite supportive of learning in our task. As in Sage and Baldwin's (2011) study, our social cue manipulations occurred prior to the teaching demonstration. As such, parent modeling did not simply highlight particular actions during the labeling (parents no longer modeled actions at this point) but instead acted to establish beforehand that the following demonstration was pedagogical in nature, enhancing subsequent learning. We believe that parents' participation cued children that the information the person on the screen subsequently presented was intended for them, similar to the effect of social cue interventions used by other researchers to establish an in-person adult's pedagogical intent (Sage & Baldwin, 2011; Topál

et al., 2008). It is also possible that the mere presence of parents facing the screen during the labeling demonstration helped to maintain children's attention in the modeling conditions. However, similar to other studies, the current study establishes that a set of co-viewer social cues supports learning from video; future research may tease apart the role of individual components of parents' modeling (e.g., joint attention, gaze direction, physical imitation).

Regardless of the type of *on-screen* cues (contingent or noncontingent) that children experienced, it was sufficient for parents to model for their children that they were attending to and participating with the actress on the screen, listening to her words and responding to her as an authority (i.e., doing what she requested), for children to learn from her. Children have been shown to learn in social situations they observe as onlookers. For example, O'Doherty et al., 2011 found that 2-year-olds learned a new word when they were part of a social interaction *or* when they watched a reciprocal social interaction occur between two other individuals, including individuals on prerecorded video. In the current research, children viewing their parents responding to the on-screen actress may have established the actress as a communicative social partner who could engage in reciprocal interaction, influencing children's perception of her viability as a potential teacher even when her contingency was actually inauthentic.

It is informative to compare our results with those of Roseberry et al. (2013). In their study, toddlers were taught a novel verb by a person either face to face, through a contingent video chat, or in the previously described yoked condition. Children were then tested on their verb knowledge using the intermodal preferential looking paradigm, in which two clips of an actor performing an action appear on opposite sides of a split screen and children need to look to the one that matches a stated verb. Toddlers displayed evidence of learning the verbs in the video-chat group as well as the face-to-face group. The type of information presented (actions vs. object labels), presentation style (visible actions vs. an object hidden in a bucket), and testing procedure (looking at a matching video vs. choosing a three-dimensional target object) all are important between-study differences that could have made our word-learning procedure comparatively difficult for toddlers in our contingent groups. In addition, our task relied on children's use of the actress's referential social cues of gaze direction into an opaque container at an absent referent rather than associating a word with an entity that was visible during labeling (cf. Baldwin, 1991, 1993). Thus, children needed to recall from memory the absent object or hold the label and location of reference in mind until the object was removed from the bucket. It is also possible that the inclusion of accurate personal details (references to children's names and prior events) in Roseberry et al. (2013) video-chat training gave children extra cues that the events on video were relevant and meaningful, leading to increased learning and transfer. Cues that emphasize relevance can support learning, such as when Henderson, Sabbagh, and Woodward's (2013) young Canadian participants were more likely to learn the names of toys special to children "around here" (Canada) than the names of toys special to children in Japan.

The effectiveness of the parent modeling and on-screen contingency manipulations when experienced together is compatible with the success of live, closed-circuit video interventions used in previous research with toddlers. Parents in three of the prior studies (Myers et al., 2016; Nielsen et al., 2008; Troseth et al., 2006) responded to the on-screen person, as did those in the current study's parent modeling conditions. In our study, parent modeling was also effective in the absence of on-screen contingency.

Also in line with prior studies, our noncontingent comparison group did not learn the demonstrated information when watching alone (Myers et al., 2016; Nielsen et al., 2008; Roseberry et al., 2013; Troseth et al., 2006). We purposely used pseudocontingency rather than inaccurate contingency for our comparison group; instead of yoked videos, children watched a prerecorded video of the on-screen actress asking for participation, pausing, and responding with generic feedback. In our study, children did not learn the novel word from the prerecorded pseudocontingent video alone, but they did succeed when they had previously observed a parent modeling responsiveness to the prerecorded actress. This finding highlights the importance of parental co-viewing when watching educational television at home, including programs using the ask-and-wait format.

It is important that toddlers in the prior and current studies were sensitive to the difference between pseudocontingency on prerecorded video and real closed-circuit contingency, indicated by their differential learning and participation. However, as technology changes to allow for increasing

levels of responsiveness between screen media and children (e.g., through touch screens and artificial intelligence), researchers must continue to explore the influence of various types of social and nonsocial contingency on children's learning.

A remaining question is how the presence of social cues available from video and from a co-viewer relates to the problems very young children have with symbolic media. Specifically, does the fact that children learn from a person on video chat or ask-and-wait pseudocontingent video mean that they "understand the symbolic nature of video" or "recognize the symbolic relation between the image and reality"? Some authors have reached this conclusion, but other interpretations are possible. The contingent responsiveness of a person on video might incite children to "look through" the surface features of the representational object—the small, flat, two-dimensional image on a screen—as if it were an unmediated event. Children who failed to represent what they were seeing as both "a surface in its own right and a display of information about something else" (Gibson, 1979, p. 282) would not be achieving "dual representation" (DeLoache, 1995) in the sense of mentally representing the symbolic or representational object (the two-dimensional image on the screen), its referent (the real absent person talking from elsewhere), and the "stands for" relation between the two. Rather, social responsiveness (with contingent video) or the illusion of social responsiveness (with pseudocontingent video) may be so compelling that children relate to the screen contents directly without achieving "representational insight."

How might these alternatives be tested? In previous research, children were helped to use video as a source of information after repeatedly seeing themselves "live" on their family television screen (Troseth, 2003; Troseth, Casey, Lawver, Walker, & Cole, 2007). When they came to the lab, they used video of the experimenter hiding a toy in a room to find the toy. Of note, on a later visit to the lab, children with the live video experience successfully used a different kind of image representation—small photographs of the hiding places—as clues in the hiding game, thereby showing a general insight into how images might relate to real events. In future studies, transfer tasks that change aspects of the learning situation previously supported by contingency (e.g., those that use another symbolic medium or have a different noncontingent person providing information on video) might be useful to clarify whether children have gained insight into the representational relation.

In the current research, parent co-viewing helped children to learn both from contingent video chat and from pseudocontingent prerecorded video. Children also generalized the novel label offered by the actress from the original object she labeled to a new example that differed in color. Parents' pedagogical cues (attention and responses to the actress's requests) may have helped children to detect the teaching intentions of the on-screen person. Parent modeling may have clarified for children that the actress was to be taken seriously as someone providing information that was *meaningful* and *intended to teach them.* In the absence of this parental support, children failed to learn the novel label following the on-screen person's referential cue.

The results from two earlier studies offer instructive comparisons. In a study using a scale model to represent a larger room, when the researcher highlighted the *intentional origins* and *intended function* of the model, 2.5-year-old children (who typically would not succeed) used the model to gain information about events in the larger room (Sharon, 2005). Similarly, in a recent study, 2-year-old children did not successfully use printed photos or iPhone photos of hiding places as clues to find a hidden toy unless they first gained experience that highlighted the *representational intention* behind taking iPhone photos of the toy's hiding places and showing them to another person who successfully used them to search for the toy (Russo Johnson, 2017). Thus, it is possible that in the current research parents taking the on-screen person seriously as someone to be listened to highlighted for children that they could learn from this person on video. Future studies will help to clarify the degree of representational insight that young children have when they learn from video chat.

Besides highlighting the importance of social cues from co-viewing parents, this study provides additional evidence that interactivity built into digital media products cannot easily replace the role of supportive co-viewers. It has been suggested that on-screen interactive features, such as contingent responsiveness to children, may make using new digital media platforms (e.g., tablets, video chat) better for children's learning and development than using "passive" media such as television (Christakis, 2014; Kirkorian, Choi, & Pempek, 2016; Lauricella, Pempek, Barr, & Calvert, 2010; Troseth, Russo, & Strouse, 2016). Although this may be the case, a growing body of research indicates that adult support

enhances learning from digital media even when contingency is built in (Lauricella, Barr, & Calvert, 2014; Strouse & Ganea, 2016).

Notably, there were important differences between our study and the use of video chat by families in the world outside the lab. Learning the name of a new object from a stranger is a very different task from communicating with family and friends through Skype or FaceTime. In our study, video chat was used in a highly manipulated experimental setting to study the effect of contingency isolated from other supportive factors. A familiar on-screen person (e.g., parent, grandparent) who directly speaks to a child, uses the child's name, and knows about the child's life provides many communicative cues not used in this study. Such interactions need to be studied using other research paradigms. For example, young children do recognize and receive comfort from their mothers on Skype when in a stressful situation (Tarasuik, Galligan, & Kaufman, 2011), and family members often help children to navigate technical disruptions and delays (McClure & Barr, 2017). Nevertheless, the effectiveness of parent modeling in helping children to learn from the contingent actress suggests that extra help from a co-participant in video chat may support very young children's recognition that on-screen events can relate to real life.

In summary, we explored parent modeling and on-screen contingency as social supports for language acquisition from video. These supports were intended to provide cues to children that the information on the video was pedagogical in nature and was useful and worth processing. Both supports increased participation with simultaneous content, but on-screen cues alone did not provide sufficient support for subsequent learning in this task. Cues given by a co-viewing parent supported word learning across the two video types and were most effective when paired with live, contingent closed-circuit video chat. We conclude that authentic social supports for learning from video, especially the robust host of cues provided by an in-person co-viewer, are effective because they emphasize the pedagogical relevance of its contents to young viewers.

Acknowledgments

References

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875–890.
Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418.
Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review, 30*, 128–154.
Barr, R., & Hayne, H. (1999). Developmental changes in imitation from television during infancy. *Child Development, 70*, 1067–1081.
Barr, R., Zack, E., Garcia, A., & Muentener, P. (2008). Infants' attention and responsiveness to television increases with prior exposure and parental interaction. *Infancy, 13*, 30–56.
Butler, L. P., & Markman, E. M. (2013). Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition, 130*, 116–127.
Christakis, D. A. (2014). Interactive media use at younger than the age of 2 years. *JAMA Pediatrics, 168*, 399–400.
Crawley, A. M., Anderson, D. R., Wilder, A., Williams, M., & Santomero, A. (1999). 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, 91*, 630–637.
Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 249–274). Oxford, UK: Oxford University Press.
Dayanim, S., & Namy, L. L. (2015). Infants learn baby signs from video. *Child Development, 86*, 800–811.

DeLoache, J. S. (1995). Early understanding and use of symbols: The model model. *Current Directions in Psychological Science, 4*, 109–113.

DeLoache, J. S., Chiong, C., Vanderborght, M., Sherman, K., Islam, N., Troseth, G. L., ... O'Doherty, K. (2010). Do babies learn from baby media? *Psychological Science, 21*, 1570–1574.

Demers, L. B., Hanson, K. G., Kirkorian, H. L., Pempek, T. A., & Anderson, D. R. (2013). Infant gaze following during parent–infant coviewing of baby videos *Child Development, 84*, 591–603.

Diesendruck, G., & Markson, L. (2011). Children's assumptions of the conventionality of culture. *Child Development Perspectives, 5*, 189–195.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics, 21*, 95–116.

Fidler, A. E., Zack, E., & Barr, R. (2010). Television viewing patterns in 6- to 18-month-olds: The role of caregiver–infant interactional quality. *Infancy, 15*, 1–21.

Friedrich, L. K., & Stein, A. H. (1975). Prosocial television and young children: The effects of verbal labeling and role playing on learning and behavior. *Child Development, 46*, 27–38.

Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science, 10*, 139–146.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*, 23–46.

Harris, P. L., & Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, 58(1, Series No. 231).

Henderson, A. M. E., Sabbagh, M. A., & Woodward, A. L. (2013). Preschoolers' selective learning is guided by the principle of relevance. *Cognition, 126*, 246–257.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction. *Psychiatry: Journal for the Study of Interpersonal Processes*, 19, 215–229.

Howard Gola, A. A., Richards, M. N., Lauricella, A. R., & Calvert, S. L. (2013). Building meaningful parasocial relationships between toddlers and media characters to teach early mathematical skills. *Media Psychology, 16*, 390–411.

Kirkorian, H. L., Choi, K., & Pempek, T. A. (2016). Toddlers' word learning from contingent and noncontingent video on touchscreens. *Child Development, 87*, 405–413.

Krcmar, M. (2011). Word learning in very young children from infant-directed DVDs. *Journal of Communication, 61*, 780–794.

Krcmar, M., Grela, B. G., & Lin, Y.-J. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology, 10*, 41–63.

Lauricella, A. R., Barr, R., & Calvert, S. L. (2014). Parent–child interactions during traditional and computer storybook reading for children's comprehension: Implications for electronic storybook design. *International Journal of Child-Computer Interaction, 2*, 17–25.

Lauricella, A. R., Howard, A. A. H., & Calvert, S. L. (2010). Toddlers' learning from socially meaningful video characters. *Media Psychology, 14*, 216–232.

Lauricella, A. R., Pempek, T. A., Barr, R., & Calvert, S. L. (2010). Contingent computer interactions for young children's object retrieval success. *Journal of Applied Developmental Psychology, 31*, 362–369.

Leslie, A. M. (1987). Pretense and representations: The origins of "theory of mind". *Psychological Review, 94*, 412–426.

McClure, E., & Barr, R. (2017). Building family relationships from a distance: Supporting connections with babies and toddlers using video and video chat. In R. Barr & D. N. Linebarger (Eds.), *Media exposure during infancy and early childhood: The effect of content and context on learning and development* (pp. 227–248). New York: Springer.

McClure, E. R., Chentsova-Dutton, Y. E., Holochwost, S. J., Parrot, W. G., & Barr, R. (2017). Look at that! Video chat and joint visual attention development among babies and toddlers. *Child Development*. https://doi.org/10.1111/cdev.12833.

Myers, L. J., LeWitt, R. B., Gallo, R. E., & Maselli, N. M. (2016). Baby FaceTime: Can toddlers learn from online video chat? *Developmental Science* 20. https://doi.org/10.1111/desc.12430.

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science, 11*, 722–731.

O'Doherty, K., Troseth, G. L., Shimpi, P. M., Goldenberg, E., Akhtar, N., & Saylor, M. M. (2011). Third-party social interaction and word learning from video. *Child Development, 82*, 902–915.

Reiser, R. A., Tessmer, M. A., & Phelps, P. C. (1984). Adult–child interaction in children's learning from "Sesame Street". *Educational Communication and Technology Journal, 32*, 217–223.

Reiser, R. A., Williamson, N., & Suzuki, K. (1988). Using Sesame Street to facilitate children's recognition of letters and numbers. *Educational Communication and Technology Journal, 36*, 15–21.

Richert, R. A., Robb, M. B., & Smith, E. I. (2011). Media as social partners: The social nature of young children's learning from screen media. *Child Development, 82*, 82–95.

Robb, M. B., Richert, R. A., & Wartella, E. A. (2009). Just a talking book? Word learning from watching baby videos. *British Journal of Developmental Psychology, 27*, 27–45.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Skype me! Socially contingent interactions help toddlers learn language. *Child Development, 85*, 956–970.

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development, 80*, 1360–1375.

Russo Johnson, C. (2017). *Young children's representational understanding of digital photos*. Unpublished doctoral dissertation, Vanderbilt University.

Sage, K. D., & Baldwin, D. (2011). Disentangling the social and the pedagogical in infants' learning about tool-use *Social Development, 20*, 825–840.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development, 69*, 309–320.

Scofield, J., Williams, A., & Behrend, D. A. (2007). Word learning in the absence of a speaker. *First Language, 27*, 297–311.

Shafto, P., & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 1632–1637). Austin, TX: Cognitive Science Society.

Sharon, T. (2005). Made to symbolize: Intentionality and children's early understanding of symbols. *Journal of Cognition and Development, 6,* 163–178.

Singer, J. L., & Singer, D. G. (1998). *Barney & Friends* as entertainment and education: Evaluating the quality and effectiveness of a television series for preschool children. In J. K. Asamen & G. L. Berry (Eds.), *Research paradigms, television, and social behavior* (pp. 305–367). Thousand Oaks, CA: Sage.

Striano, T., Henning, A., & Stahl, D. (2006). Sensitivity to interpersonal timing at 3 and 6 months of age. *Interaction Studies, 7,* 251–271.

Strouse, G. A., & Ganea, P. A. (2016). Are prompts provided by electronic books as effective for teaching preschoolers a biological concept as those provided by adults? *Early Education and Development, 27,* 1190–1204.

Strouse, G. A., O'Doherty, K., & Troseth, G. L. (2013). Effective co-viewing: Preschoolers' learning from video after a dialogic questioning intervention *Developmental Psychology, 49,* 2368–2382.

Strouse, G. A., & Troseth, G. L. (2014). Supporting toddlers' transfer of word learning from video. *Cognitive Development, 30,* 47–64.

Tarasuik, J. C., Galligan, R., & Kaufman, J. (2011). Almost being there: Video communication with young children. *PLoS ONE, 6*(2), e17129.

Topál, J., Gergely, G., Miklósi, Á., Erdohegyi, Á., & Csibra, G. (2008). Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science, 321,* 1831–1834.

Troseth, G. L. (2010). Is it life or is it Memorex? Video as a representation of reality. *Developmental Review, 30,* 155–175.

Troseth, G. L. (2003). TV guide: Two-year-old children learn to use video as a source of information. *Developmental Psychology, 39,* 140–150.

Troseth, G. L., Casey, A. M., Lawver, K. A., Walker, J. M., & Cole, D. A. (2007). Naturalistic experience and the early use of symbolic artifacts. *Journal of Cognition and Development, 8,* 309–331.

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development, 69,* 950–965.

Troseth, G. L., Pierroutsakos, S. L., & DeLoache, J. S. (2004). From the innocent to the intelligent eye: The early development of pictorial competence. *Advances in Child Development and Behavior, 32,* 1–35.

Troseth, G. L., Russo, C. E., & Strouse, G. A. (2016). What's next for research on young children's interactive media? *Journal of Children and Media, 10,* 54–62.

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development, 77,* 786–799.

Troseth, G. L., Strouse, G. A., Verdine, B. N., & Saylor, M. M. (2017). Responsiveness and relevance: Is social information from video sufficient for word learning?

Vandewater, E. A. (2011). Infant word learning from commercially available video in the U.S. *Journal of Children and Media, 5,* 1748–2801.

Vredenburgh, C., Kushnir, T., & Casasola, M. (2015). Pedagogical cues encourage toddlers' transmission of recently demonstrated functions to unfamiliar adults. *Developmental Science, 18,* 245–254.

Watkins, B., Calvert, S. L., Huston-Stein, A. C., & Wright, J. C. (1980). Children's recall of television material: Effects of presentation mode and adult labeling. *Developmental Psychology, 16,* 672–674.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-olds. *Developmental Psychology, 34,* 1289–1309.

Zack, E., & Barr, R. (2016). Maternal scaffolding enhances learning from touch screens during infancy. *Frontiers in Psychology, 30.* https://doi.org/10.3389/fpsyg.2016.01264.