The Ghost in the Touchscreen: Social Scaffolds Promote Learning by Toddlers

Article *in* Child Development · December 2016

DOI: 10.1111/cdev.12683

5 authors, including:

Laura Zimmermann
University of Delaware

Alecia Moser
Binghamton University

Peter Gerhardstein
Binghamton University

Rachel Barr
Georgetown University

CHILD DEVELOPMENT

Child Development, xxxx 2016, Volume 00, Number 0, Pages 1–13

The Ghost in the Touchscreen: Social Scaffolds Promote Learning by Toddlers

Laura Zimmermann
Georgetown University

Alecia Moser
Binghamton University

Herietta Lee
Georgetown University

Peter Gerhardstein
Binghamton University

Rachel Barr
Georgetown University

This study examined the effect of a "ghost" demonstration on toddlers' imitation. In the *ghost* condition, virtual pieces moved to make a fish or boat puzzle. Fifty-two 2.5- and 3-year-olds were tested on a touchscreen (no transfer) or with 3D pieces (transfer); children tested with 3D pieces scored above a no demonstration baseline, but children tested on the touchscreen did not. Practice on the touchscreen (n = 23) by 2.5- and 3-year-olds prior to the ghost demonstration did not improve performance. Finally, children who learned the puzzle task via a social demonstration and were tested on the touchscreen (n = 26) performed better than the ghost conditions. Taken together, these studies demonstrate that social demonstrations enhance learning from novel touchscreen tools during early childhood.

Use of touchscreen devices by 2- to 4-year-olds in the United States increased from 39% to 80% between 2011 and 2013 (Rideout, 2013). Touchscreens have advantages over other media because they can provide interactive and contingent feedback, but also have limitations; children may not receive sufficient scaffolding to learn effectively from such platforms (Chiong, Takeuchi, & Erickson, 2012; Parish-Morris, Mahajan, Hirsh-Pasek, Golinkoff, & Collins, 2013). Many educational apps do not include social demonstrators, and instead display on-screen objects that move by themselves. Young children demonstrate a *transfer deficit*, consistently learning fewer actions when they have to transfer learning from television and touchscreens to real-world objects, compared to when no transfer is required (Dickerson, Gerhardstein, Zack, & Barr, 2013; Moser, Zimmermann, Grenell, Barr, & Gerhardstein, 2015; Zack, Barr, Gerhardstein, Dickerson, & Meltzoff, 2009; Zack, Gerhardstein, Meltzoff, & Barr, 2013; Zimmermann et al., 2015). Barnett and Ceci (2002) define transfer as the application of information to a situation distinctly different from the condition in which it was learned and can be applied to conditions changing from 2D to 3D. The *transfer deficit* may be partly accounted for by the differences in dimensionality between the time of encoding and the point of retrieval; that is, children experience difficulty in generalizing their learning across changes in the physical context (Barr, 2010, 2013).

Other researchers have focused on the lack of social contingency to account for the *transfer deficit* (Nielsen, 2006; Troseth, 2010). Ghost displays (a term coined by Fawcett, Skinner, & Goldsmith, 2002) provide a unique opportunity to examine the role of social contingency in learning from media. In a ghost condition, an apparatus is moved without obvious influence from the demonstrator,

We thank Jim Tse, MS for program development, Debbie Simon for audio recordings, and Sarah Olsen for editing the stimuli. Thanks to Amanda Grenell for her help with data collection and Arianna Gorkowitz for providing writing assistance. Special thanks to the families and children who made this research possible. This research was supported by a National Science Foundation (NSF) grant to Peter Gerhardstein and Rachel Barr (1023772).

Correspondence concerning this article should be addressed to Laura Zimmermann, Psychology, Georgetown University, 306A WGR Hall, 3700 O St. NW Washington, DC 20057. Electronic mail may be sent to ljz7@georgetown.edu.

DOI: 10.1111/cdev.12683

usually on a screen or by fishing line (for review, see Hopper, 2010). The use of the ghost display assesses whether or not the experimenter's behavior is necessary for the subject to learn. If learning occurs from the ghost display, it is inferred that it was done through emulation. An enhanced ghost control has also been used (Fawcett et al., 2002). This method is the same as the original ghost condition, but an experimenter is positioned near the apparatus and directs visual attention to the task. Using an enhanced ghost condition controls for the effect of a social presence during the encoding period (Hopper, 2010).

Researchers have used ghost displays to manipulate the conditions of social learning by infants (Huang & Charman, 2005; Tennie, Call, & Tomasello, 2006), toddlers (Thompson & Russell, 2004), preschoolers (Subiaul, Vonk, & Rutherford, 2011; Subiaul et al., 2007), and adults (Jacobson & Sisemore, 1976), as well as nonhuman species (Groesbeck & Duerfeldt, 1971; Klein & Zentall, 2003) and in comparative studies (e.g., Hopper, Lambeth, Schapiro, & Whiten, 2008). Findings have been mixed with some studies showing no differences between ghost and social demonstrations (e.g., Hopper et al., 2008; Jacobson & Sisemore, 1976; Klein & Zentall, 2003), while other studies demonstrating significantly poorer performance following a ghost relative to the social demonstration (e.g., Groesbeck & Duerfeldt, 1971; Hopper, Lambeth, Schapiro, & Whiten, 2015; Subiaul et al., 2011; Tennie et al., 2006). When the task is simple and has some clear object affordances, performance following a ghost and social demonstration does not differ. For example, Hopper et al. (2008) compared performance by chimpanzees to performance by 3- to 4-year-old children on a simple one-step bidirectional door sliding task. In the push demo group, the subject would see another child slide the door and get a reward, while in the enhanced ghost control group, the door seemed to slide without human control while a child sat in front of the box. The authors reported no significant difference between live demonstration and enhanced ghost control groups. Both chimpanzees and children matched the direction of the door movement to retrieve a reward in the enhanced ghost condition on their first trial (see also Klein & Zentall, 2003). There were, however, age-related changes in the ability to learn from an enhanced ghost demonstration using this task. Using the same one-step bidirectional door sliding task, Tennie et al. (2006) tested infants (12, 18, and 24 months), as well as four great ape species. Twelve-month-old children were unable to perform the task in either the enhanced ghost or live demonstration conditions, while 18-month-olds were only successful in the live condition. However, 24-month-olds were able to succeed in both the enhanced ghost and live demonstrated conditions. Taken together, results from these studies indicate that by 24 months, children can learn a single goal-directed action in the absence of a demonstrator.

In other instances, where the task is more complex, or the affordances are less obvious and outside the animal's typical repertoire, it appears that the social demonstration provides scaffolding needed to facilitate learning over and above the ghost demonstration (Hopper, Flynn, Wood, & Whiten, 2010). For example, Thompson and Russell (2004) examined learning of a hidden pulley system by 14- to 26-month-olds. The conditions included two counterintuitive demonstrations (a) simple, where a mat was pushed backward resulting in a toy moving forward ("single mat"), and (b) complex, where pulling the first mat without a toy enabled the second mat with the toy to move forward ("double mat"). Infants in the ghost demonstration performed more actions in the simple single-mat task, but less in the complex double-mat task compared to the social demonstration.

A few studies have also taken advantage of screens to examine learning from ghost demonstrations. Screens have an advantage in that they do not have hidden wires or pulley systems, but instead use edited video footage. Huang and Charman (2005) randomly assigned 17-month-olds to one of three demonstration conditions: (a) full demonstration on video, (b) ghost display (objects moved but experimenter was digitally removed from video), (c) body movement display (experimenter's movement shown but objects digitally removed from video), or a baseline control condition. They reported that 17-month-olds were able to imitate the one-step target actions (e.g., pull apart a dumbbell) significantly more in the full demonstration and ghost display relative to body movement or control conditions. This suggests that object movement can provide infants with salient information about the capacity of a particular object to be used to achieve a goal. Potentially prior knowledge of the objects themselves, coupled with object movement, allowed for affordance learning.

Finally, Subiaul et al. (2007, 2011) compared performance between typically developing preschoolers and children diagnosed with autism spectrum disorders on sequence learning tasks presented on touchscreens via a ghost or a social demonstrator. They found that performance on a three-step cognitive imitation sequence (press three images in the correct order, e.g., apple–boy–cat) did not differ as a function of demonstration condition or as a function of clinical diagnosis. All children imitated the sequence significantly better than individual trial-and-error baseline condition. In a subsequent study, Subiaul et al. (2011) examined whether children's attribution of agency to the computer or training on the task influenced learning via a ghost or social demonstrator. They found that performance was enhanced in the social demonstration conditions over the ghost demonstration conditions when there was not a training phase even though children attributed agency to the computer. That is, when the task was complex, the affordances were not obvious and in the absence of training, the social demonstration facilitated task performance.

A recent study exposed chimpanzees to an escalating series of social demonstrations to investigate the role of social support when observing ghost displays (Hopper et al., 2015). The task was a two-action Pan-pipes tool-use task that has been used previously with children (Hopper et al., 2010) and chimpanzees. Chimpanzees presented with various displays in their home social groups showed no success in solving the task in baseline, trial-and-error, and ghost display conditions. Next, Hopper et al., 2015 presented chimpanzees with a series of five sequential tests to assess learning (baseline, ghost display, human model, chimpanzee video, live chimpanzee). Most chimpanzees were only successful after watching a live chimpanzee demonstrate the task suggesting live demonstrations and the species demonstrating impact learning from novel imitation tasks.

Although cognitive factors associated with the transfer deficit have been examined, social manipulations have not been completely explored. In the present study, we further examine learning by toddlers from ghost demonstrations using a complex touchscreen-based puzzle imitation task (Dickerson et al., 2013). Moser et al. (2015) used this task to compare 2.5- and 3-year-olds' transfer of learning. Children were shown either a live touchscreen demonstration (2D) or a live (3D) demonstration of the magnet puzzle task in which an experimenter made a fish or boat on the touchscreen or with real puzzle pieces on a magnetized puzzle board. Children were subsequently tested on the touchscreen or the puzzle board. This allowed for a fully crossed design, where children were tested on a no transfer task (e.g., demonstration on the 2D touchscreen and test on 2D touchscreen) or a transfer task (e.g., demonstration on the 2D touchscreen and test on the 3D magnet board). Moser et al. reported that although older children outperformed younger children, there was a bidirectional transfer deficit; no transfer groups produced higher imitation performance than transfer groups (see also Zack et al., 2009; Zack et al., 2013 for the same pattern of results with 15-month-old infants). Children failed to transfer learning across dimensions regardless of transfer direction. Moser et al. concluded that the poor memory flexibility was a barrier to transfer learning that persisted until children were 3 years of age or older.

The present study extends that of Moser et al. (2015) to test the ability of 2.5- and 3.0-year-old children to learn how to assemble a puzzle from a demonstration on a touchscreen in which the pieces move on their own across a virtual board surface while an experimenter is present. This enhanced ghost demonstration is referred to as "ghost" throughout as we did not include a ghost condition where an adult experimenter was absent. Children were subsequently tested with either the 2D touchscreen puzzle or with the 3D magnetic puzzle. Based on the transfer deficit literature, we predicted that, following a ghost demonstration on the touchscreen, children would perform better when tested on a touchscreen than a magnet board. We also predicted that 3-year-olds would perform significantly better than 2.5-year-olds, as reported in previous studies using this task (Dickerson et al., 2013; Moser et al., 2015; Zimmermann et al., 2015).

Experiment 1a: Transfer of Learning From Ghost Conditions

In Experiment 1a, we examined whether children could transfer learning from a 2D ghost demonstration to a 3D puzzle test relative to a 2D touchscreen test. Children were randomly assigned to one of two conditions: *no transfer ghost 2D or transfer ghost 3D*. In the *no transfer ghost 2D* condition, children viewed a (2D) ghost demonstration followed by a 2D touchscreen test. In the *transfer ghost 3D* condition, children viewed a (2D) ghost demonstration followed by a 3D test on the magnet puzzle board (ghost 3D). We predicted that children in the no transfer group (ghost 2D) would perform better than those in the transfer group (ghost 3D) based on transfer findings from Moser et al. (2015).

Method

Participants

The study included 52 typically developing children (23 boys) from two metropolitan areas. Data were collected from May 2014 to February 2016. Independent groups of children were tested at 2.5 years (N = 27, M_{age} = 30 months 16.9 days, SD = 12.85 days, range = 29.77–31.54 months) and 3 years (N = 25, M_{age} = 36 months 14.19 days, SD = 14 days, range = 35.68–37.22 months). Participants were primarily Caucasian (73.1%) and from college-educated families ($M_{years\ of\ education}$ = 17.56, SD = 1.30). The remaining 27% of the sample included the following races: mixed (17.3%), African American (1.9%), and Asian (7.7%). Additionally, 9.6% of the sample was Latino. The mean rank of socioeconomic index (SEI; Nakao & Treas, 1992) was 77.54 (SD = 14.02) based on 36 (69%) families. Additional children were excluded from the analysis for the following reasons: six due to parental interference, four due to technical error, and two for interacting with the stimuli prior to test.

Baseline data from Moser et al. (2015) was included (N = 50, 28 boys) as a comparison sample. Additionally, we partially replicated the previously collected baseline data (N = 24, 13 boys) using the same methods as the current study and Moser et al. (2015). The partial replication sample included 2.5-year-olds (N = 12, M_{age} = 30 months 14.60 days, SD = 12.67 days) and 3-year-olds (N = 12, M_{age} = 36 months 11.86 days, SD = 13.18 days). As with the above sample, the participants were primarily Caucasian (87.5%) but included those of mixed race (0.04%) and Asian (0.04%). Of the replication baseline sample, 12.5% was Latino. The average years of education was 17.42 (SD = 0.93) and the mean rank of SEI was 72.62 (SD = 19.45) based on 19 (79.2%) families.

Apparatus

The Planar touchscreen apparatus, model LA1710RTC (17 in. diagonal touchscreen display) with capacitive touch hardware was set on a small table, measuring 13 in. high × 24 in. wide × 18 in. deep. The child was seated on a stool in front of it within reaching distance. The apparatus was made of black plastic and was 35.5 cm tall, 42 cm wide, and 23.3 cm deep. For the 3D tests, the apparatus contained a removable school bus yellow magnetic board. For the 2D tests, the touchscreen display with 1,280 × 1,024 resolution was used throughout and had the same colored background as the 3D magnetic board (Figure 1).

3D magnetic stimuli. The two types of puzzles for the 3D test were the same as those for the 2D test, representing a "boat" and a "fish." The two starting positions at the beginning of the test were the same as those used in the 2D demonstrations. The boat, when fully assembled, measured 15.2 cm high at the center and 12.1 cm long at the widest point. The fish, when fully assembled, measured 8.9 cm high at the widest point and 15.2 cm long from the end of the green piece to the end of the blue piece (from Dickerson et al., 2013). The 3D magnetic puzzle pieces (0.5 cm depth) were located in three corners of the apparatus. The pieces were removable and could be easily slid around on the magnet board.

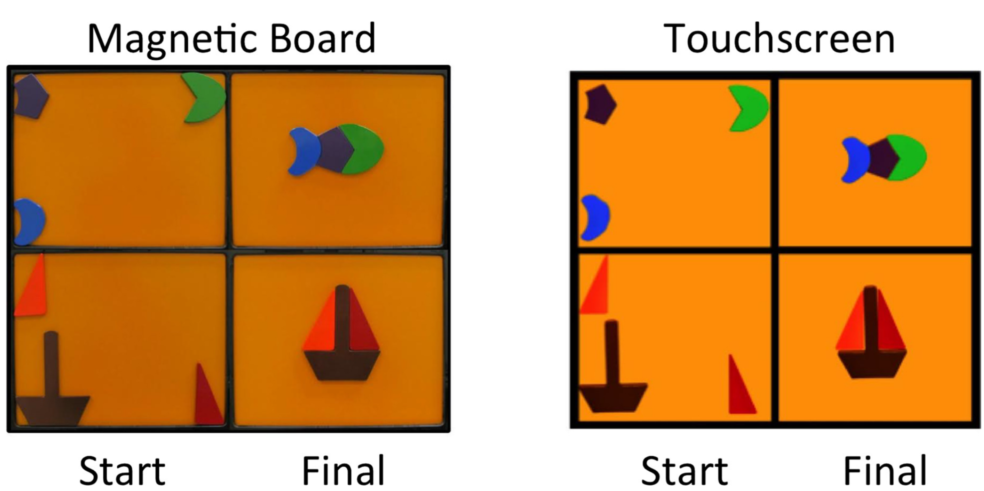

Figure 1. Stimuli. Left: Magnetic board (3D) puzzle start/end configurations. Right: Touchscreen (2D) start/end configurations. [Color figure can be viewed at wileyonlinelibrary.com].

2D touchscreen stimuli. At the beginning of the test, the images of puzzle pieces were located in three corners of the apparatus and there were two starting positions. The 2D stimuli were generated from high-resolution photographs of the 3D pieces in order to match color and size. The ghost demonstration was made by video-capturing the movement of the puzzle pieces and corresponding verbal cues made by an adult demonstrator. In this way the movement of the puzzle pieces were identical to those of a social demonstration.

Procedure

All protocols were approved by the Georgetown University and Binghamton University Internal Review Boards. Participants were randomly assigned to the ghost 2D or ghost 3D conditions. The session began with a brief play session, to make the child comfortable and minimize distraction during the test.

Demonstration phase. The demonstration began with the touchscreen covered with a black cloth. During the ghost demonstration (see Figure 2), the experimenter lifted the cloth and allowed the ghost video to run. The ghost video consisted of three repetitions of the same recording. The first piece moved while a female voice said, "Look at this." The second piece then moved to fit with the first piece, and the voice said, "What was that?" The last component moved, and the voice said, "Isn't that fun?" The recorded female voice was different from the live experimenters. In the ghost demonstration the shapes slide across the screen making the 2D object motion analogous to a 3D social demonstration (Dickerson et al., 2013) except no gesture was present. The video then repeated twice for a total of three demonstrations lasting approximately 60 s (M = 61 s, SD = 4 s). The voice concluded with, "Now it's your turn." Once the puzzle was complete, the experimenter covered the display with the black curtain to obscure the child's view while the pieces were moved back to their starting positions by pressing the spacebar on the touchscreen (2D).

Test phase. Half the participants were tested on a touchscreen, ghost 2D, and half using a real magnetic puzzle board, ghost 3D. Once the demonstration concluded, the experimenter reset the touchscreen magnet program for the no transfer ghost 2D group or added the magnet board and magnet pieces for the transfer ghost 3D group. A short delay (no transfer: M = 19.67 s, SD = 8.14 s; transfer: M = 25.55 s, SD = 5.43 s) separated the start of the test phase from the end of the demonstration phase. The test phase began when the experimenter lifted the black cloth to reveal either the touchscreen or the magnetic board, with pieces in their starting positions and said "Now it's your turn." If the child did not approach the apparatus soon after the test phase began, the experimenter encouraged the child to interact with the puzzle using nonspecific prompts (i.e., "It's your turn" or "It's ok, you can touch it"). Each child was given 60 s from their first contact with the touchscreen or magnetic board to interact with the apparatus. The ghost 2D and ghost 3D groups were compared to baseline 2D and baseline 3D groups, respectively, that had been collected by Moser et al. (2015), obtained using the same sampling methods from the same geographic regions. Children in the baseline conditions participated only in the test phase to estimate the rate of spontaneous production of the demonstrated actions during puzzle play and completed the manipulation check. For these

Figure 2. The ghost demonstration phase (left) and test phase (right) on the touchscreen using the virtual 2D pieces to make the fish puzzle. Note presence of experimenter (not demonstrating) during the ghost demonstration. [Color figure can be viewed at wileyonlinelibrary.com].

participants, the test phase was identical to the ghost 2D and ghost 3D conditions except that they did not view a ghost demonstration. Following the test phase, the experimenter demonstrated the target actions once, and gave the child the opportunity to interact with the apparatus. The manipulation check was performed to ensure that children were physically capable of sliding and connecting the puzzle pieces on the magnetic board and the touchscreen. All passed the manipulation check. When tests were completed the child received a small toy such as a ball and a certificate for participating.

Coding

Each session was video-recorded for later coding in Datavyu (Datavyu Team, 2014), a software program designed for flexible behavioral coding. All behaviors were time-stamped and coded goals. The coding scheme is identical to that used by Moser et al. (2015).

On-task behaviors. Each contact with a puzzle piece (beginning when a piece was touched and ending when the touch ended) was coded along two dimensions, gesture and goal. The combined behavior was considered an on-task behavior. On-task behaviors excluded exploratory play (interactions where the piece was removed from the board for more than 3 s) and microgestures (a piece was "nudged," meaning that it was moved < 1/6 of the board) not resulting in any connection.

Goal coding. Coded actions that connected puzzle pieces included the following categories of goals: *correct connection, target–error connection, and connect other* (Table 1). Based on 30% of all test sessions collected in Experiments 1a and 1b that were rescored by a second coder, interrater reliability was above the acceptable level of .70 (Landis & Koch, 1977); $\kappa_{goal} = .921$.

Goal imitation score. Following Dickerson et al. (2013), children received one point for each correct connection (max = 2 correct connections). The goal imitation score was then converted to a proportion (out of two). Although we had previously established that we could calculate a goal efficiency measure using these same codes (Moser et al., 2015; Zimmermann et al., 2015), we only report the goal imitation scores in the present study as the pattern of results for goal efficiency was similar.

Results

Both daily puzzle play (*M* = 23 min, *SD* = 22 min), and the frequency of daily touchscreen usage (*M* = 20 min, *SD* = 26 min) were common daily activities for children in the present study. A 2 (age: 2.5, 3.0 years) × 2 (condition: ghost 2D, ghost 3D) × 2 (sex: male, female) analysis of variance (ANOVA) yielded a main effect of condition $F(1, 44) = 20.84$, $p < .001$, $\eta_p^2 = .32$, and a main effect of age, $F(1, 44) = 4.43$, $p < .05$, $\eta_p^2 = .09$. Post hoc Tukey's honestly significant difference (HSD) tests showed that children performed significantly better ($p < .05$) on the ghost 3D (*M* = 0.56, *SD* = 0.50) than the ghost 2D condition *M* = 0.17, *SD* = 0.31) and that older children *M* = 0.52, *SD* = 0.49) outperformed younger children (*M* = 0.22, *SD* = 0.38; Table 2). A Sex × Age interaction was present, $F(1, 44) = 11.16$, $p < .01$, $\eta_p^2 = .20$, as well as a Sex × Condition interaction, $F(1, 44) = 4.89$, $p < .05$, $\eta_p^2 = .10$. Both of these interactions were unexpected and our design was not powered to

Table 1
Definitions for the Goal Codes and the Goal Imitation Measure

Goal codes	Definitions
Correct connection	Puzzle piece is connected to the central piece in the correct location and orientation within approximately 2 mm of touching.
Target–error connection	Pieces connected either (a) in the correct location but incorrect orientation (e.g., one piece turned sideways), (b) not within the 2 mm threshold, or (c) disconnected before the end of the behavior.
Connect other	Connections made to nontarget pieces (e.g., connecting the two sails together in the boat puzzle).
No connection	An on-task behavior that did not result in any type of connection.

Measure	Calculation
Goal imitation	Number of correct connections/2 (max. possible correct connections)

Table 2
Means (SD) *of Male and Female 2.5- and 3-Year-Olds in the Ghost 2D and Ghost 3D Conditions*

	2.5-year-olds		3-year-olds	
	Male	Female	Male	Female
Ghost 2D	.17 (.26)	.06 (.18)	.08 (.20)	.42 (.49)
Ghost 3D	.88 (.25)	.11 (.33)	.71 (.49)	.83 (.41)

systematically examine sex differences. An examination of the means did reveal an interesting pattern of results that require further empirical attention. Younger (2.5-year-old) females performed worst on the ghost 3D condition, but older (3-year-old) females performed best on the ghost 3D condition (see Table 2). The Sex × Age interaction occurs due to the fact that the 2.5-year-old boys in the ghost 3D group are doing surprisingly well, but otherwise overall that 3-year-olds olds are outperforming 2.5-year-olds.

Wilcoxon nonparametric tests were used to compare test performance to baseline 2D and baseline 3D groups collected by Moser et al. (2015) combined with partial replication data not previous reported. Children of both ages in the ghost 3D conditions produced significantly more correct goals than age- and test dimension-matched baseline controls (*p*s < .05). In contrast, children in the ghost 2D condition did not perform above 2D baseline at either age.

Experiment 1b: How Does Practice Influence Learning From a Ghost Demonstration?

The outcome of Experiment 1a was surprising. Although children performed significantly above baseline in the ghost 3D condition, performance plummeted to baseline levels in the ghost 2D condition. It is possible that without a social demonstration, children were simply unaware that the image was presented on a functional touchscreen. Even with encouragement, this lack of awareness may have limited goal imitation in the ghost 2D condition. To rule out this possibility, in Experiment 1b we added a touchscreen practice activity prior to demonstration, to ensure that all children were familiarized with the touchscreen apparatus. Children practiced interacting with the touchscreen using a commercially available drawing program, followed by a ghost demonstration and 2D touchscreen test (practice + ghost 2D).

Given that prior studies had shown that training enhanced learning via a ghost demonstration (Subiaul et al., 2011), presumably by increasing the child's understanding of the object affordances of the touchscreen, we predicted that giving children practice with the touchscreen would elevate their performance. Specifically, we predicted children in the practice + ghost 2D group would perform better than those in the ghost 2D condition from Experiment 1a because they should benefit from the interactive knowledge of the touchscreen apparatus provided by the practice session prior to the demonstration.

Method

Participants

The study included 23 typically developing children (11 boys) from two metropolitan areas. Independent groups of children were tested at 2.5 years ($N = 12$, $M_{age} = 30$ months 16.9 days, $SD = 9.37$ days, range = 30.19–31.04 months) and 3 years ($N = 11$, $M_{age} = 36$ months 17.36 days, $SD = 13.59$ days, range = 35.81–37.22 months). Participants were primarily Caucasian (91.3%) and from college-educated families ($M_{years\ of\ education} = 15.52$, $SD = 2.182$). The remaining 8.7% of the sample included the following races: mixed (4.3%) and African American (4.3%). The mean rank of SEI (Nakao & Treas, 1992) was 64.85 ($SD = 20.669$) based on 20 (87%) families. One additional child was excluded due to parental interference.

Apparatus and Procedure

The touchscreen apparatus and demonstration and test procedures for the 2D touchscreen were identical to those used in Experiment 1a except that children in the practice 2D-ghost condition participated in a brief practice session prior to the ghost demonstration. Following the warm-up, the child participated in a touchscreen practice session using a commercially available drawing program. Practice began after the experimenter removed the black cloth that covered the touchscreen. A square (10.16 × 9.53 cm) and circle (10.80 cm diameter) were outlined on the screen against a white background (see left panel Figure 3). The experimenter prompted the child to touch the shapes with phrases like, "look at the circle," "you can touch it," "now touch the square," and "wow, wasn't that fun?" When the child touched either of the shapes or the background, they filled in with a teal color. The practice phase ($M = 40$ s, $SD = 52$ s) ended when the child successfully touched and filled in the circle and square. The practice phase was followed by the identical ghost demonstration from Experiment 1a and the 2D test phase protocol.

Results

As in Experiment 1a, daily puzzle play ($M = 30$ min, $SD = 39$ min) and touchscreen usage ($M = 37$ min, $SD = 35$ min) were common daily

Figure 3. The touchscreen practice session was standardized across children and all children were instructed to touch items on the screen. [Color figure can be viewed at wileyonlinelibrary.com].

events for children in the practice ghost 2D group. Coding was identical to Experiment 1a. A 2 (age: 2.5, 3.0 years) × 2 (condition: ghost 2D, practice + ghost 2D) × 2 (sex of the child: male, female) ANOVA on goal imitation yielded no main effect of age; $F(1, 42) < 1$, *ns*, no effect of experimental condition, $F(1, 42) = 1.50$, $p = .23$, but a main effect of sex of the child emerged, $F(1, 42) = 5.16$, $p < .05$, $\eta_p^2 = .11$. There were no significant interactions. A post hoc Tukey's HSD test revealed that females ($M = 0.33$, $SD = 0.39$) performed significantly better ($p < .05$) than males ($M = 0.13$, $SD = 0.27$). Once again, Wilcoxon nonparametric tests revealed that 3-year-olds in the practice ghost 2D conditions ($M = 0.21$, $SD = 0.33$) did not perform significantly above age-matched baseline 2D control group ($M = 0.13$, $SD = 0.23$). The 2.5-year-olds ($M = 0.42$, $SD = 0.42$) performed significantly above ($p < .05$) the age-matched baseline 2D group ($M = 0.08$, $SD = 0.25$), showing some benefit of the practice session. Although females performed significantly better than males, the practice group did not outperform the no practice group. That is, poor performance on the ghost 2D condition could not be accounted for by poor understanding of the functionality of the touchscreen apparatus.

Does Performance Following a Ghost Demonstration Differ From a Social Demonstration?

We also elected to partially replicate our Moser et al. (2015) on the social demonstration performed with the same virtual touchscreen pieces as in the current study. We could then compare whether the demonstrator influenced learning from a touchscreen or transfer to 3D objects. By comparing these 2D touchscreen demonstration conditions (ghost 2D vs. social 2D), we may examine how social scaffolds influence learning on a transfer task. We predicted that performance in the social demonstration group would be greater than the ghost demonstration group.

Participants

The participants for social demonstration conditions collected by Moser et al. (2015) were drawn from the same geographical location using the same recruitment methods and their demographic profile did not differ for participants in the present study. Partial replication of the social demonstration conditions were combined with Moser et al. (2015) to form the comparison groups. This replication data included both 2.5-year-olds ($N = 15$, $M_{age} =$ 30 months 14.30 days, $SD = 12.03$ days) and 3-year-olds ($N = 11$, $M_{age} = 36$ months 17.95 days, $SD = 11.69$ days) from college-educated families ($M_{years\ of\ education} = 17.46$, $SD = 1.07$). The majority of participants were Caucasian (80.8%)

Table 3
Description of the Demonstration and Test Conditions and Experimenter Actions for the Social (Moser et al., 2015) and Ghost Groups (Present Study) for the Cross Experiment Comparison

Group name	Demo	Experimenter action	Test
Ghost 2D	G	Virtual pieces automatically move	TS
Ghost 3D	G	Virtual pieces automatically move	MB
Social 2D	S	Experimenter slides virtual pieces on the touchscreen	TS
Social 3D	S	Experimenter slides virtual pieces on the touchscreen	MB

Note. All groups saw a 2D demonstration, followed by either a 2D or 3D test. Group name denotes demonstration test type. G = ghost; S = social; TS = touchscreen; MB = magnet board.

and the mean rank of SEI was 78.62 (SD = 11.19) based on 22 (84.62%) families.

Apparatus and Procedure

The apparatus, procedure, and coding protocols for the social demonstration groups were identical to the ghost demonstration conditions, except that an experimenter moved the puzzle pieces on the touchscreen instead of the pieces moving automatically (see Table 3).

Results

A three-factor ANOVA on Model (ghost, social) × Transfer (2D, 3D) × Age (2.5, 3.0 years) on goal imitation scores yielded a main effect of age, $F(1, 133) = 24.31$, $p < .001$, $\eta_p^2 = .15$, a main effect of transfer $F(1, 133) = 5.77$, $p < .05$, $\eta_p^2 = .04$, and a main effect of model $F(1, 133) = 19.25$, $p < .001$, $\eta_p^2 = .13$. Post hoc Tukey's HSD tests showed that overall, children learned significantly better ($p < .05$) from the social model (M = 0.65, SD = 0.44) than the ghost model (M = 0.37, SD = 0.45) and that 3.0-year-old children (M = 0.73, SD = 0.41) performed significantly better than 2.5-year-olds (M = 0.38, SD = 0.44; Figure 4). Additionally, children performed significantly better in the transfer 3D test group (M = 0.61, SD = 0.48) than the no transfer 2D test group (M = 0.48, SD = 0.44). The effect of the model were qualified by a significant Model × Transfer interaction, $F(1, 133) = 9.63$, $p < .01$, $\eta_p^2 = .07$. Post hoc Tukey's HSD tests revealed that performance was significantly worse on the ghost 2D condition (M = 0.17, SD = 0.31) than all other conditions, which did not significantly differ from one another. In other words, the social demonstrator elevated performance in the social 2D condition (M = 0.68, SD = 0.39), but performance did not differ as a function of demonstrator when children were tested in the transfer 3D test, ghost 3D (M = 0.56, SD = 0.50) and social 3D (M = 0.63, SD = 0.47). This interaction reflects a poorer imitation in the ghost than in the social group when children were tested on the touchscreen, and no difference between the ghost and social groups' transfer performance on the magnet board (Figure 4).

For confirmation purposes, Wilcoxon nonparametric tests contrasted social 2D to baseline 2D and social 3D to baseline 3D groups and revealed that 3-year-olds in the social 2D conditions (M = 0.89, SD = 0.21) performed significantly above age-matched baseline 2D control group (M = 0.13, SD = 0.23) and the social 3D conditions (M = 0.85, SD = 0.35) performed significantly above age-matched baseline 3D control group (M = 0.14,

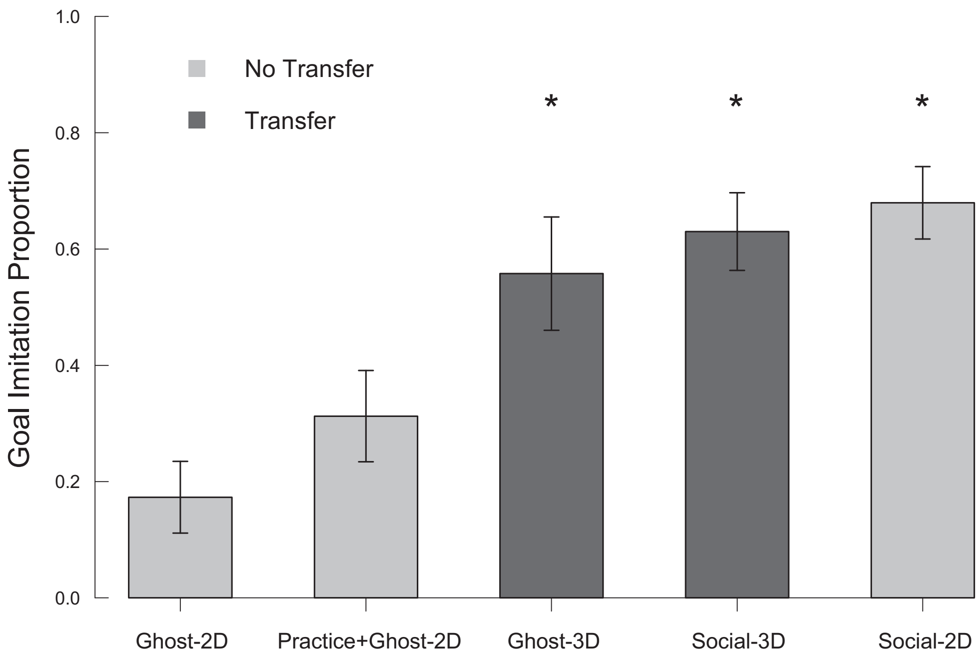

Figure 4. Goal imitation scores following either a ghost or social demonstration on the touchscreen. Asterisks (*) indicate groups significantly above the ghost 2D group. Children were tested either on the 2D touchscreen or the 3D puzzle board.

SD = 0.29). As expected, the pattern of results was the same for the 2.5-year-olds, social 2D (*M* = 0.50, *SD* = 0.42) performed significantly above age-matched baseline 2D control group (*M* = 0.08, *SD* = 0.25) and social 3D (*M* = 0.44, *SD* = 0.49) performed significantly above age-matched baseline 3D control group (*M* = 0.06, *SD* = 0.24).

Discussion

Overall, the present study showed that imitation of a complex puzzle task varied as a function of both the type of demonstration and whether or not the test involved transfer of learning to 3D magnetic puzzle pieces. Recall that in Experiment 1a children performed significantly above baseline in the ghost 3D condition; however, performance was equivalent to baseline levels in the ghost 2D condition. In Experiment 1b we provided children with a practice session prior to the demonstration. Performance did not improve, suggesting that poor performance in the touchscreen test was not due to a failure to understand that the touchscreen apparatus was interactive. The ghost demonstration condition, in which no demonstrator was visible, was as effective as a social demonstrator in promoting transfer of learning to real-world objects. Only the social demonstrator, however, was effective when children were tested on a novel touchscreen tool. Overall, we conclude that the social demonstration facilitated performance under conditions where the affordances of the virtual world were less well known, but did not facilitate performance under the difficult transfer of learning condition. This surprising pattern of results highlights the importance of the social model in rapid learning during early childhood and suggests that social scaffolding increases children's perception of seemingly ambiguous events across time (Thompson & Russell, 2004). That is, by initiating a sequence of events or connecting actions in a sequence, the demonstrator contingently links actions to goals. This set of findings has important practical implications for educational and other app designers who typically orient children to games using a "ghost-like" tutorial.

The findings are consistent with others that have examined learning from ghost models. Overall, on this complex type of task, learning in a ghost condition did not exceed baseline levels. These findings replicate and extend studies conducted with infants using simple tasks (Tennie et al., 2006) and complex tasks with toddlers (Subiaul et al., 2011; Thompson & Russell, 2004), demonstrating tasks that are high in complexity and have less obvious affordances result in little overall learning from a ghost model (Acerbi, Tennie, & Nunn, 2011; Hopper et al., 2010). Thus, there appears to be interplay between task complexity and social learning. The current study also extends previous research, showing differences in how effective the demonstrator was in different settings, with social model facilitation in the 2D touchscreen condition but not in the 3D magnet board condition. Although this finding requires additional empirical attention, we interpret this to be due to child experiences with the affordances of the 3D puzzles allowing them to utilize the ghost demonstration under real-world test conditions. This may also help to reconcile some of the mixed findings with ghost conditions. Many studies conducted with toddlers and preschoolers have shown no differences between imitation from ghost and social models, perhaps due to lower task complexity (e.g., Hopper et al., 2008, 2010; Huang & Charman, 2005) or the introduction of training to enhance understanding of affordances (Subiaul et al., 2011).

Toddlers are increasingly being exposed to tablets and the need to understand how the social learning context impacts learning from this technology is paramount (Hipp et al., 2016). The present study specifically illustrates the facilitation effect of the social cues and context in learning from apps over and above familiarization with the touchscreen alone. Children who viewed a ghost demonstration, in which pieces moved on a digital screen by themselves, were unable to replicate the target actions on a screen; those who were shown a social demonstration of the moving pieces, however, performed successfully. That is, learning to make a puzzle on the touchscreen was dramatically impaired in the ghost demonstration compared to the social demonstration. The fact that social scaffolding during a 2D demonstration results in high levels of performance compared to the ghost demonstrator suggests that learning can be boosted when young children are unfamiliar with the learning context (e.g., making a puzzle on a touchscreen), as they often are when it comes to digital technology. This finding is particularly important in the context of apps intended for education and young children, as many current applications, including some written for the toddler and preschool population, use exactly this type of "ghost" demonstration. The present results suggest that young children may take much longer to master an app under these learning conditions. This may lead parents and educators to draw false conclusions. Rather than believing that children are having difficulty learning from the ghost demonstration per se, they may conclude that children cannot master the content or apply the content to other situations. Based on the current findings both of these assumptions may be wrong. These findings suggest that app developers should move away from ghost demonstrations of features or app interfaces aimed at children under 3 unless the ghost demonstrations are showing very simple and well understood virtual manipulations. Early educators and parents must gauge the child's capability of learning from apps and consider supplementing with additional verbal instruction or hands-on demonstrations that may better allow children to learn information in the app.

We had anticipated that applying learning outside of the digital context may be even more difficult. But surprisingly in our study, the transfer condition performance was not impacted by whether there was a ghost or a social demonstration. This may be due to the fact that for children, puzzle play is highly social and collaborative (Levine, Ratliff, Huttenlocher, & Cannon, 2012) and children have a great deal, more experience with puzzle play using real objects than on a touchscreen interface. Consistent with this conclusion, studies with older children have shown that this transfer deficit may diminish with age. Huber et al. (2015) reported with 4- to 6-year-olds on the Tower of Hanoi task; with three practice trials on the touchscreen children were able to transfer to the 3D Tower of Hanoi apparatus as well as those who had practice with the 3D version. It is possible that additional scaffolding may boost transfer of learning and this question warrants further empirical investigation. At this point, it is also unclear how prior touchscreen exposure might influence learning across different apps. Mere touchscreen exposure should not be equated with digital competency as this engagement may be passive (i.e., viewing videos on Youtube) or active (i.e., collaborating with a peer on a game). Our results from a brief practice phase with the touchscreen did not facilitate better imitation performance from the ghost demonstration, and touchscreen exposure in the home did not correlate with performance on the novel touchscreen task.

Given research on the prevalence of educational apps, app developers need to consider principles of science of learning in the design and use of apps during early childhood (see Hirsh-Pasek et al., 2015 for review). Recent work has identified four key features that make apps educational: active involvement, engagement, meaningfulness, and social interaction (Hirsh-Pasek et al., 2015). As predicted based on these learning principles, the current study demonstrates that the presence of social interaction or scaffolding is critical when toddlers are learning in a novel virtual context. Understanding social factors has practical importance, as educational contexts often require children to transfer knowledge from a task they learned in one context (e.g., online or in a video) to another (a real-life or hands-on test context). For example, a child may learn basic addition and subtraction with blocks from their teacher in the classroom and later be given a math activity on a touchscreen for a follow-up assignment. To better support children's understanding of the content, app developers can build repetition to facilitate encoding and comprehension (Barr et al., 2007; Crawley et al., 1999), adaptive play to facilitate engagement, contingency or direct feedback to allow children to learn from their correct and incorrect responses to improve performance, or levels/scaling to increase task complexity based on the child's individual performance.

The present study was not without limitations. For the 2.5-year-olds, in some instances, performance was very low and it would be useful to vary task complexity as well as the model (ghost, social) in order to disentangle the relative contribution of task complexity and model. The puzzle task and the touchscreen measurement strategy does however lend itself well to these further investigations. Future studies could examine the role of touchscreen feedback (auditory or other sensory information such as haptic cues) that would provide more contingent responses to the child. The unexpected finding of sex differences with a female advantage to learning following a ghost demonstration requires additional empirical investigation. It is possible that this was experiential, although we did not show any clear differences in parent reports of touchscreen usage between boys and girls. It could be also that girls were performing at levels that were more representative of older children on the ghost condition. We interpret these findings very cautiously however, given that performance was low on ghost 2D conditions and did not exceed baseline levels overall.

This study contributes to our understanding of the novel touchscreen tool. Taken together, the present study and the studies leading up to it show that young children's learning is affected by social scaffolding, prior knowledge of the objects, the degree of transfer of learning, task complexity, and other factors (potentially including the gender of the child, an issue that requires further future investigation). Due to the ease of programming and the interactive interface, touchscreen technology holds significant promise for use in educational settings. The present study underscores, however, that despite the frequency and ease with which young children interact with touchscreen devices, it is the case that they require social scaffolding in order to maximize its educational potential. The issue of touchscreen-specific interactive affordances does not appear to be a primary factor in limiting children's use of the technology, but additional research will be necessary to examine how knowledge and skills transfer within the touchscreen environment to better understand how and when social scaffolds are necessary for learning to occur.

References

Acerbi, A., Tennie, C., & Nunn, C. L. (2011). Modeling imitation and emulation in constrained search spaces. *Learning & Behavior*, *39*, 104–114. doi:10.3758/s13420-010-0009-z

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637. doi:10.1037/0033-2909.128.4.612

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review*, *30*, 128–154. doi:10.1016/j.dr.2010.03.001

Barr, R. (2013). Memory constraints on infant learning from picture books, television, and touchscreens. *Child Development Perspectives*, *7*, 205–210. doi:10.1111/cdep.12041

Barr, R., Muentener, P., Garcia, A., Fujimoto, M., & Chávez, V. (2007). The effect of repetition on imitation from television during infancy. *Developmental Psychobiology*, *49*, 196–207. doi:10.1002/dev.20208

Chiong, C., Takeuchi, L., & Erickson, I. (2012). *Print books vs. e-books, comparing parent–child co-reading on print, basic, and enhanced e-book platforms* [press release]. Retrieved from www.joanganzcooneycenter.org/wp-content/uploads/2012/07/jgcc_ebooks_quickreport.pdf

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. doi:10.1207/S1532785XMEP0202_4

Datavyu Team. (2014). Retrieved from www.datavyu.org

Dickerson, K., Gerhardstein, P., Zack, E., & Barr, R. (2013). Age-related changes in learning across early childhood: A new imitation task. *Developmental Psychobiology*, *55*, 719–732. doi:10.1002/dev.21068

Fawcett, T. W., Skinner, A. M., & Goldsmith, A. R. (2002). A test of imitative learning in starlings using a two-action method with an enhanced ghost control. *Animal Behaviour*, *64*, 547–556. doi:10.1006/anbe.2002.3092

Groesbeck, R. W., & Duerfeldt, P. H. (1971). Some relevant variables in observational learning of the rat. *Psychonomic Science*, *22*, 41–43. doi:10.3758/BF03335928

Hipp, D., Gerhardstein, P., Zimmermann, L., Moser, A., Taylor, G., & Barr, R. (2016). The dimensional divide: Learning from TV and touchscreens during early childhood. 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 33–54). New York, NY: Springer.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps lessons from the science of learning. *Psychological Science in the Public Interest*, *16*, 3–34. doi:10.1177/1529100615569721

Hopper, L. M. (2010). "Ghost" experiments and the dissection of social learning in humans and animals. *Biological Reviews*, *85*, 685–701. doi:10.1111/j.1469-185X.2010.00120.x

Hopper, L. M., Flynn, E. G., Wood, L. A. N., & Whiten, A. (2010). Observational learning of tool use in children: Investigating cultural spread through diffusion chains and learning mechanisms through ghost displays. *Journal of Experimental Child Psychology*, *106*, 82–97. doi:10.1016/j.jecp.2009.12.001

Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Whiten, A. (2008). Observational learning in chimpanzees and children studied through "ghost" conditions. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*, 835–840. doi:10.1098/rspb.2007.1542

Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Whiten, A. (2015). The importance of witnessed agency in chimpanzee social learning of tool use. *Behavioural Processes*, *112*, 120–129. doi:10.1016/j.beproc.2014.10.009

Huang, T. C., & Charman, C. (2005). Gradations of emulation learning in infant's imitation of actions on objects. *Journal of Experimental Child Psychology*, *92*, 276–302. doi:10.1016/j.jecp.2005.06.003

Huber, B., Tarasuik, J., Antoniou, M. N., Carrett, C., Bowe, S. J., Kaufman, J.; The Swinburne Babylab Team. (2015). Young children's transfer of learning from a touchscreen device. *Computers in Human Behavior*, *56*, 56–64. doi:10.1016/j.chb.2015.11.010

Jacobson, J. M., & Sisemore, D. A. (1976). Observational learning of a lever pressing task. *Southern Journal of Educational Research*, *10*, 59–73.

Klein, E. D., & Zentall, T. R. (2003). Imitation and affordance learning by pigeons (*Columba livia*). *Journal of Comparative Psychology*, *117*, 414. doi:10.1037/0735-7036.117.4.414

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: a predictor of preschoolers' spatial transformation skill. *Developmental psychology*, *48*, 530. doi:10.1037/a0025913

Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., & Gerhardstein, P. (2015). They can interact but can they learn? Toddlers' transfer learning from touchscreens and television. *Journal of Experimental Child Psychology*, *137*, 137–155. doi:10.1016/j.jecp.2015.04.002

Nakao, K., & Treas, J. (1992). *The 1989 socioeconomic index of occupations: Construction from the 1989 occupational prestige scores* (General Social Survey Methodological Report No. 74). Chicago: University of Chicago, National Opinion Research Center.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, *42*, 555–565. doi:10.1037/0012-1649.42.3.555

Parish-Morris, J., Mahajan, N., Hirsh-Pasek, K., Golinkoff, R. M., & Collins, M. F. (2013). Once upon a time: Parent–child dialogue and storybook reading in the electronic era. *Mind, Brain, and Education*, *7*, 200–211. doi:10.1111/mbe.12028

Rideout, V. (2013). *Zero to eight: Children's media use in America 2013*. San Francisco, CA: Common Sense Media.

Subiaul, F., Lurie, H., Romansky, K., Klein, T., Holmes, D., & Terrace, H. (2007). Cognitive imitation in typically-developing 3-and 4-year olds and individuals with autism. *Cognitive Development*, *22*, 230–243. doi:10.1016/j.cogdev.2006.10.003

Subiaul, F., Vonk, J., & Rutherford, M. (2011). The ghosts in the computer: The role of agency and animacy attributions in "ghost controls". *PLoS ONE*, *6*, e26429. doi:10.1371/journal.pone.0026429

Tennie, C., Call, J., & Tomasello, M. (2006). Push or pull: Imitation vs. emulation in great apes and human children. *Ethology*, *112*, 1159–1169. doi:10.1111/j.1439-0310.2006.01269.x

Thompson, D. E., & Russell, J. (2004). The ghost condition: Imitation versus emulation in young children's observational learning. *Developmental Psychology*, *40*, 882. doi:10.1037/0012-1649.40.5.882

Troseth, G. L. (2010). Is it life or is it Memorex? Video as a representation of reality. *Developmental Review*, *30*, 155–175. doi:10.1016/j.dr.2010.03.007

Zack, E., Barr, R., Gerhardstein, P., Dickerson, K., & Meltzoff, A. N. (2009). Infant imitation from television using novel touch screen technology. *British Journal of Developmental Psychology*, *27*(1), 13–26. doi:10.1348/026151008X334700

Zack, E., Gerhardstein, P., Meltzoff, A. N., & Barr, R. (2013). 15-Month-olds' transfer of learning between touch screen and real-world displays: Language cues and cognitive loads. *Scandinavian Journal of Psychology*, *54*, 20–25. doi:10.1111/sjop.12001

Zimmermann, L., Gerhardstein, P., Moser, A., Grenell, A., Dickerson, K., Yao, C., & Barr, R. (2015). Do semantic contextual cues facilitate transfer learning from video in toddlers? *Frontiers in Psychology*, *6*, 561. doi:10.3389/fpsyg.2015.00561

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Video S1. Ghost Demonstration: Boat Puzzle
Video S2. Ghost Demonstration: Fish Puzzle