Journal of Cognition and Development

ISSN: 1524-8372 (Print) 1532-7647 (Online) Journal homepage: http://www.tandfonline.com/loi/hjcd20

An Attentional Goldilocks Effect: An Optimal Amount of Social Interactivity Promotes Word Learning From Video

Kate Nussenbaum & Dima Amso

To cite this article: Kate Nussenbaum & Dima Amso (2016) An Attentional Goldilocks Effect: An Optimal Amount of Social Interactivity Promotes Word Learning From Video, Journal of Cognition and Development, 17:1, 30-40, DOI: 10.1080/15248372.2015.1034316

To link to this article: http://dx.doi.org/10.1080/15248372.2015.1034316

JOURNAL OF COGNITION AND DEVELOPMENT, 17(1):30–40

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2015.1034316

An Attentional Goldilocks Effect: An Optimal Amount of Social Interactivity Promotes Word Learning From Video

Kate Nussenbaum and Dima Amso
Brown University

Television can be a powerful education tool; however, content makers must understand the factors that engage attention and promote learning from screen media. Prior research has suggested that social engagement is critical for learning and that interactivity may enhance the educational quality of children's media. The present study examined the effects of increasing the social interactivity of television on children's visual attention and word learning. Three- to 5-year-old (M_{age} = 4;5, *SD* = 9 months) children completed a task in which they viewed videos of an actress teaching them the Swahili label for an on-screen image. Each child viewed these video clips in 4 conditions that parametrically manipulated social engagement and interactivity. We then tested whether each child had successfully learned the Swahili labels. Though 5-year-old children were able to learn words in all conditions, we found that there was an optimal level of social engagement that best supported learning for all participants, defined by engaging the child but not distracting from word labeling. Our eye-tracking data indicated that children in this condition spent more time looking at the target image and less time looking at the actress's face as compared with the most interactive condition. These findings suggest that social interactivity is critical to engaging attention and promoting learning from screen media up until a certain point, after which social stimuli may draw attention away from target images and impair children's word learning.

Since the premiere of *Sesame Street* in 1969, many stations have aired programming that aims to both entertain and educate preschool-aged audiences. But the extent to which children can learn through television remains unclear. Research has suggested that social interactivity may enhance the educational quality of children's television (Rice, Huston, Truglio, & Wright, 1990; Richert, Robb, & Smith, 2011). However, the direct effects of increasing television interactivity on children's learning have not been demonstrated.

Previous studies on learning through media have focused on the role of social information. Many studies have revealed a "video deficit"—a gap between the amount of information children learn from real people compared with the amount they learn from people talking to them via television (Anderson & Pempek, 2005; Kuhl, 2007; Troseth, 2003; Troseth, Saylor, & Archer, 2006). In one study, 9- and 10-month-olds exposed to Mandarin, through reading children's books and playing with toys with a native Mandarin speaker, performed significantly better on tests of Mandarin speech sound discrimination compared with peers exposed only to English (Kuhl, Tsao, & Liu, 2003). However, infants exposed to movies of Mandarin speakers

Correspondence should be sent to Kate Nussenbaum, Cognitive, Linguistic and Psychological Sciences, Brown University, 190 Thayer St., Providence, RI 02912, USA. E-mail: katenuss@gmail.com

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hjcd.

reading books and playing with toys for the same amount of time did not demonstrate phonetic learning. In another study, 2-year-olds were able to locate a toy after watching an experimenter hide it in person but not when they watched it being hidden through a video feed (Troseth & DeLoache, 1998).

Younger children may face a particularly strong need for live social interaction to learn language. In a study of verb learning, children aged 2 to 4 years old were able to learn novel verb–action pairs through a combination of educational television clips and live teaching demonstrations by an experimenter (Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009). However, when the live interaction component was removed from the learning phase, only older children learned the new verbs. The researchers hypothesized that the benefits of live interaction were derived from children's heightened arousal in social settings and the experimenter's ability to establish joint gaze to better direct children's attention toward the named action. This idea is supported by the work of Krcmar, Grela, and Lin (2007), who examined 15- to 24-month-olds' learning of object names under different conditions, including among others, a joint reference condition in which an experimenter labeled the object when the child attended to it and a children's programming condition in which children learned an object label from an edited *Teletubbies* clip. Children learned most successfully in their joint reference condition, thereby highlighting the important role of social engagement in directing attention toward relevant stimuli to promote learning. Taken together, this work supports the hypothesis that learning may be "gated" by live social interaction (Kuhl, 2007).

Still, research suggests real-world digital media can promote word learning. Allen and Scofield (2010) found, for example, that 2-year-olds could learn object labels from videos and then transfer those labels to real-world settings. Additionally, children who spent more time watching age-appropriate television programming demonstrated greater vocabulary growth than their peers who watched fewer hours of these shows (Linebarger & Walker, 2005; Rice et al., 1990;). In one study, 3-year-olds who watched *Sesame Street* developed larger vocabularies than their nonviewing peers (Rice et al., 1990). In another study, time exposed to certain children's programs was correlated with larger vocabularies and higher expressive language scores in 2.5-year-olds (Linebarger & Walker, 2005).

One explanation that may reconcile these conflicting findings is that children learn from television only when it effectively captures their attention and elicits active participation during viewing, an idea supported by the finding that infants were more responsive to television when parent coviewers asked questions, provided labels, or made abstractions about video content (Barr, Zack, Garcia, & Muentener, 2008). Similarly, children were more capable of finding a toy after watching it being hidden through a live video feed when experimenters interacted with them through the feed prior to object retrieval (Troseth et al., 2006). Three- and 4-year-olds were also able to learn more letters and numbers from *Sesame Street* clips when adults sat with them, asked them reinforcing questions (e.g., "What letter is that?"), and provided contingent feedback during viewing (Reiser, Tessmer, & Phelps, 1984).

Furthermore, integrating contingent social interaction into screen-based learning may reduce, or even erase, the "video deficit" (Lauricella, Pempek, Barr, & Calvert, 2010; Nielsen, Simcock, & Jenkins, 2008; Roseberry, Hirsch-Pasek, & Golinkoff, 2014). For example, 2- to 3-year-old children who interacted with an experimenter over Skype prior to engaging in a Skype-based word-learning task learned novel verbs equally as well as children who engaged in the learning task with a live experimenter (Roseberry et al., 2014). But children who watched a video recording of an experimenter's interactions with a different child prior to the word-learning task failed to learn the novel verbs. Nonetheless, as of yet, television programs cannot broadcast contingent social interactions into children's homes. Thus, it is important to examine strategies to promote word learning that can more feasibly be implemented in existing children's programs. To our knowledge, the effects of parametrically increasing the social interactivity of prerecorded digital media on word learning have not been tested directly.

Here we examined the effects of increasing the interactivity of a television clip on children's word learning. To do so, we designed video clips in which a television actress taught children Swahili words across four different conditions that parametrically manipulated social engagement and interactivity. The majority of the reviewed literature, as well as children's programming, is focused on preschool-aged children. As such, we tested 3-, 4-, and 5-year-old children on our novel paradigm. Based on the findings of Roseberry et al. (2009), we reasoned that the effects of social interactivity may impact younger word learners differently relative to older, more language-experienced preschoolers. Specifically, we expected the more interactive conditions to elicit greater social engagement and enhanced attention to the video clips for all participants in our sample. But we hypothesized that younger children may benefit more from this greater attentional engagement during word learning and thus would demonstrate a greater difference in their ability to learn from video clips in more interactive versus less interactive conditions.

METHOD

Participants

Sixty-four children (M_{age} = 4;5, *SD* = 9 months, range = 3;2–5;11; 40 boys) participated in the study. Of these, 22 were 3-year-olds (*M* = 3;7, *SD* = 2.8 months), 25 were 4-year-olds (*M* = 4;6 *SD* = 3 months), and 17 were 5-year-olds (*M* = 5;4, *SD* = 3 months). An additional 3 children were tested but excluded from analyses for failure to complete the task. Participants were screened via parental report to ensure that they had no previous or current known learning disabilities, uncorrected visual or auditory impairments, and/or preterm birth. Participants were recruited from the community and through the Department of Health. Parents gave consent for their children's participation, and families were compensated for their time. According to parental report, 78.1% of participants lived in homes in which only English was spoken and 15.6% of participants lived in homes in which English and one other language were spoken. No participants had any exposure to Swahili. The parents of four children did not provide language information.

Stimuli and Script

The stimuli for this study included four different video clips each filmed in four different ways, for a total of 16 unique clips. All clips were 30 seconds in duration and featured the same actress. The background of every video was the same and featured a cartoon sun, cloud, giraffe, and tree spread across the four corners (Figure 1). The goal of each video was to teach participants to name one of the objects in Swahili. All of the Swahili words were two syllables

FIGURE 1 Example of a video task screen. Each 30-s video clip featured the same actress and background. In each video, the actress taught participants the name of one of the background images in Swahili: *wingu* (cloud), *jua* (sun), *miti* (tree), or *twiga* (giraffe).

with the emphasis placed on the first syllable. All videos began and ended with the actress on a safari and taking pictures of her surroundings.

In the very low-interactivity condition, the actress did not make eye contact with the camera and said, for example, "I know how to say *'giraffe'* in Swahili," and then she took a picture of the relevant giraffe. When the actress clicked her camera, a red circle appeared around the giraffe. The actress then said the word in Swahili twice and then said, "That's how you say *'giraffe'* in Swahili." In the low-interactivity condition, the script remained the same as in the very low-interactivity condition, except the actress made eye contact with the camera after repeating the Swahili word twice, and while looking out at the participant, she added, "Do *you* know how to say *giraffe* in Swahili?" The actress then immediately resumed taking pictures. Medium-interactivity condition videos were nearly identical to those of low-interactivity condition videos, except that before the actress took pictures at the end, she paused for 4 s and looked at the camera presumably waiting for the participant's response. High-interactivity condition videos were nearly identical to those of medium-interactivity condition videos, except that after the pause, the actress said, "That's right! You're speaking Swahili, just like me." Sixteen videos were created so that each of the four words appeared in each of the four conditions. All videos were identical in length.

Each word was presented once per child for a total of four trials per participant. Condition order was counterbalanced across participants. Test trials displayed four images in a row in the center of the touch screen (Figure 2). These four images included either all four pictures that corresponded to the Swahili words taught in video clips or three pictures that corresponded to the familiar Swahili words and one novel catch image, which was either a leaf or flower, designed to increase task difficulty. Novel catch images appeared equally across conditions. Each test screen was accompanied by an audio recording of the actress's voice that said, "Touch the *Swahili word*." The location of the correct image was counterbalanced across participants.

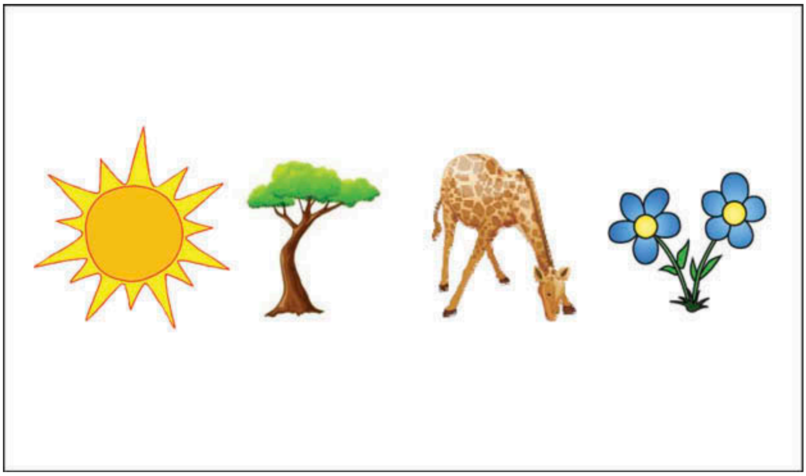

FIGURE 2 Example of a test screen. Test screens included four images. While participants viewed the screen, a recording of the actress's voice instructed them to touch one of the images. For example, they might hear, "Touch the *twiga*," to which the correct response would be pressing the image of the giraffe.

Video Task Procedure

Participants were randomly assigned to one of four video groups. Each video group comprised four different videos, each featuring a different word in a different condition. They then watched the experimenter do a demonstration trial, which consisted of a high-interactivity condition video that taught the Swahili word for truck: *lori*. During the demonstration, the experimenter said the Swahili word out loud during the actresses pause to demonstrate the interactive nature of the videos. After the video, a test screen came up, accompanied by an audio recording of, "Touch the *lori*." The experimenter touched the truck and explained that she was doing so because *lori* means truck in Swahili. Participants were then told they had an opportunity to practice what they had just seen. They then watched another practice video that used the Swahili word for elephant: *tembo*. When the test screen came up, the experimenter did not advance forward until the participant had correctly pressed the image of the elephant. Participants then saw a screen with images of the four words they were going to hear in Swahili, and the experimenter ensured the participants knew the English names of all four images. Participants were told they were going to watch more videos and that they would learn two words in a row and should try to remember both of them.

Participants then watched a set of two 30-s videos, each presenting a different word in a different condition. A 1-s-long blank screen followed each video to separate them from each other. After participants viewed two videos, they saw two test screens, each asking them to touch the picture of one of the words they had just learned. Participants had an unlimited amount of time to respond and were told to "make a guess" if they were unsure. After participants responded to the second test screen, they were told they were going to watch two more videos. The testing procedure was then repeated with a different set of two videos, which included the remaining two words in the remaining two conditions. The pairing of the videos was counter-balanced within each video group, across participants. The order of the videos within each video pair was randomized for each participant. The order of the test screens within each testing pair was also randomized for each participant. The touch screen recorded the accuracy per trial.

Eye-Tracking Procedure

Of the 64 participants who completed the study, the final subset of 25 (M_{age} = 4;3, *SD* = 9 months) were eye-tracked while they completed the task. The eye-tracking condition was added after the initial behavior-only data collection revealed that word learning was better in the medium-interactivity condition than in the high-interactivity condition. We added an additional 25 participants with eye tracking to better understand the reasons for these results and to specifically investigate how patterns of visual attention in the medium- and high-interactivity conditions may have been contributing to differential word learning.

A remote eye tracker (SensoMotoric Instruments REDSystem) was used to record participants' eye movements. Prior to testing, each participant's point of gaze was calibrated using a five-point protocol provided by the SensoMotoric Instruments (SMI) Experiment Center software and was then validated using a four-point procedure. Based on the data from validation, we computed the deviation between each participant's point of gaze and the location of a screen target. We used the tools provided by the SMI BeGaze analysis software to draw areas of interest (AOI) around the four background target images and the actress's face. When the face AOI was closest to one of the background AOIs, the visual angle between them was 2.5º. Thus, we used the deviations computed from the validation procedure to identify participants whose calibrations were not sufficiently accurate (< 2.5º) to determine whether they were looking at the actress's face or one of the background AOIs (*N* = 10 participants). For the remaining participants included in our analyses (*N* = 15, M_{age} = 4;0, *SD* = 8 months), average horizontal deviation was 0.89º (*SD* = 0.40º) and average vertical deviation was 0.97º (*SD* = 0.37º), thereby enabling assessments of eye movements to the face and background images.

Language Measure

Participants were administered Tests 1 through 7 and Test 9 from the Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock et al., 2001). Test 1 yielded a raw verbal comprehension score for 62 participants, which was calculated by summing participants' scores on the Picture Vocabulary, Synonyms, Antonyms, and Verbal Analogies subtests.

RESULTS

Effects of Condition by Age

Our primary aims were a) to determine whether increasing the level of interactivity in a video clip would enhance children's word learning, and b) to determine whether this effect differed across our age range. Trials were coded categorically as either correct or incorrect and were treated with statistical methods appropriate for binary variables. A Cochran's Q test identified significant differences between the interactive conditions, $\chi^2(3, N = 64) = 10.43$, $p = .015$. Planned McNemar's tests revealed significantly better accuracy for medium-interactivity videos relative to very low-interactivity videos, $M_3 = 0.63$ (CI_3 [0.50, 0.74]); $M_1 = 0.38$ (95% CI_1 [0.26, 0.50]); χ^2

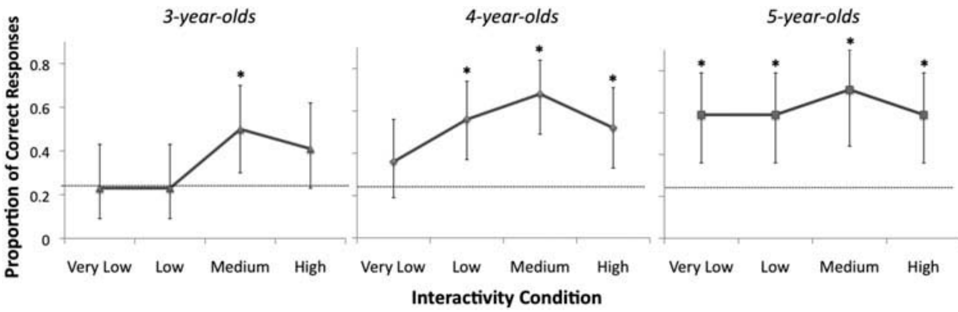

FIGURE 3 Condition accuracy by age group. Error bars indicate 95% confidence intervals. Asterisks indicate performance that was significantly above chance (.25). All *p*s ≤ .01.

(1, $N = 64$) = 8.65, $p = .003$, and a trend-level improvement between medium-interactivity videos and low-interactivity videos, $M = 0.45$ (95% CI [0.34, 0.58]), $\chi^2(1, N = 64) = 3.23$, $p = .072$. No other accuracy differences between very low-, low-, or high-interactivity conditions were significant (all *p*s > .12).

We further examined the effects of interactive condition as a function of age (Figure 3). Using one-sample binomial tests, we compared participants' performance on the task to chance-level accuracy (four alternatives on each test screen results in a chance level accuracy of .25). Three-year-olds performed above chance only on medium-interactivity videos ($p = .01$). In contrast, 4-year-olds performed above chance on low-, medium-, and high-interactivity videos (all *p*s < .005). Five-year-olds performed above chance on all interactive conditions (all *p*s < .005). These findings suggest that older children were able to learn from the video stimuli regardless of their level of social interactivity. However, increasing social interactivity helped learning in younger age groups. In support of this interpretation, we found that the three age groups performed significantly differently from each other in the low-interactivity condition, Pearson χ^2 (2, $N = 64$) = 6.93, $p = .03$, and trended toward performing significantly differently from each other in the very low-interactivity condition, Pearson $\chi^2(2, N = 64) = 5.37$, $p = .068$, with accuracy in these conditions improving with age. There were no group differences in the medium-interactivity condition, $\chi^2(2, N = 64) = 2.26$, $p = .32$, and the high-interactivity condition, $\chi^2(2, N = 64) = 1.30$, $p = .523$.

Visual Attention

An unexpected pattern emerged in the data. Performance in the medium-interactivity condition surpassed performance in the very low-interactivity condition, while performance on the high-interactivity condition, which was designed to be more interactive due to its inclusion of non-contingent positive feedback, did not. We thus examined the extent to which differences between the medium-interactivity condition and the high-interactivity condition specifically elicited different distributions of visual attention. We eye-tracked a subset of participants across ages. For each interactive condition, we calculated the proportion of time participants spent looking at the actress's face and the target image (Face or Target Image/Total Time Spent Looking at the Screen for that trial). A two-way within-subjects analysis of variance (ANOVA) was conducted with the within-subjects factors of interactive condition (medium-interactivity condition and high-interactivity condition) and attention distribution (object and face).[1] We found a main effect of attention distribution, $F(1, 14) = 96.63$, $p < .001$, as well as an Interactive Condition × Attention Distribution interaction effect, $F(1, 14) = 4.47$, $p = .05$. Figure 4 shows that while watching high-interactivity videos, children demonstrated an increased proportion of looking to the face relative to the object. These results suggest that increased social interactivity of high-interactivity videos compared with medium-interactivity videos may have impaired word learning by drawing children's attention away from the target image during encoding.

Role of Language Ability

Finally, we examined the agents of change in performance across age groups. Specifically, older children may have performed better than younger children on the less interactive conditions due to their advanced language expertise. We divided our sample into three groups based on raw verbal comprehension scores from the Woodcock-Johnson III Test of Cognitive Abilities: a) weak verbal scorers ($N = 22$), who correctly answered 8 to 13 items ($M = 11.3$, $SD = 1.42$); b) average verbal scorers ($N = 20$), who correctly answered 14 to 21 items ($M = 17.75$, $SD = 2.2$); and c) strong verbal scorers ($N = 20$), who correctly answered 22 to 29 items ($M = 24.45$, $SD = 2.46$). Though raw verbal comprehension scores correlated with age, $r(60) = .67$, $p < .001$, the weak-, average-, and strong-scoring groups contained children from different age groups, with respective mean ages for weak verbal scorers of 3;10 (SD = 7 months, range = 3;2–5;1), for average verbal scorers of 4;5 (SD = 7 months, range = 3;4–5;8), and for strong verbal scorers of 5;0 (SD = 6 months, range = 4;3–5;10).

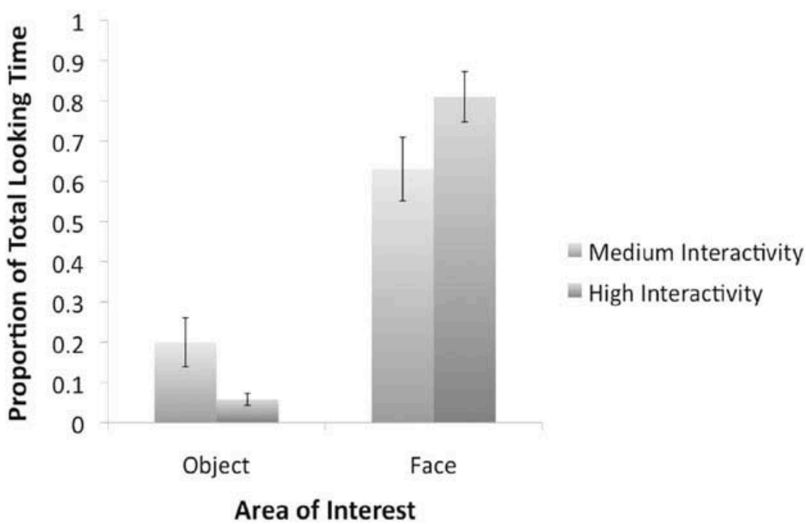

FIGURE 4 Proportion of total looking time to the face and object of interest in the medium-interactivity condition versus the high-interactivity condition.

[1] An ANOVA examining Interactive Condition (very low, low, medium, high) × Attention Distribution (object and face) resulted in only a main effect of attention distribution, $F(1,11) = 141.20$, $p = .000$, with participants spending a greater proportion of time looking at the face than at the object in all conditions.

We compared group performance in each condition to chance level (.25) using one-sample binomial tests. Weak verbal scorers performed above chance in the most interactive conditions, the medium-interactivity condition (p = .03) and the high-interactivity condition (p = .003). Average verbal scorers performed above chance in the low-interactivity condition (p = .035), the medium-interactivity condition (p < .001), and the high-interactivity condition (p = .01), while strong verbal scorers performed above chance on all conditions (all ps < .005). This pattern suggests older children's ability to learn words under all levels of social interactivity was in part driven by their greater experience and facility with language.

DISCUSSION

Our findings indicated that children are able to learn novel words through digital media created in the style of popular children's programming. Though 5-year-olds were able to learn under all conditions, increasing the social interactivity of the clips promoted word learning in 3- and 4-year-olds. Interestingly, participants only benefited from increasing interactivity to a point, after which the social elements of the stimuli seemed to draw visual attention away from the object of interest. Thus, there seems to be an attentional Goldilocks effect for learning from digital media—an optimal amount of social interactivity that engages children without distracting them from target information. The term "Goldilocks effect" refers to data from Kidd, Piantadosi, and Aslin (2012), who found that infants spent the most time looking at sequences of images that were neither too simple nor too complex.

The idea that an optimal amount of social interactivity promotes learning from digital media is supported by previous work showing that social stimuli are useful for learning in so far as they cue attention toward relevant information (Wu & Kirkham, 2010; Yurovsky, Wade, & Frank, 2013). In Yurovsky et al.'s (2013) word-learning paradigm, attention to the face was a negative predictor of learning—children who followed social cues toward their referents learned new word mappings, while those who remained focused on the speaker's face did not. Other work has shown that shared gaze and joint attention, in which gaze is mutually focused on a target object, promote word learning in both typical and atypical development (Brooks & Meltzoff, 2005, 2008; Mundy, Sigman, & Kasari, 1990). Similarly, Roseberry et al. (2009) found that social engagement was especially critical to promote younger children's learning from media. Our results are consistent with this work and suggest that an optimal level of interactivity may be especially important to support learning from videos in young preschool children, where word learning is still developing rapidly (Gathercole, Willis, Emslie, & Baddeley, 1992). Our own data indicated that age may be a proxy for the number of words or amount of language experience children have. We found that children with lower verbal comprehension scores benefitted the most from increasing levels of social interactivity during word learning.

Krcmar et al. (2007) helped explain the "video deficit" by arguing that children's programs do not successfully draw attention toward relevant information. However, our findings provide an alternative view. We suggest that any general claims about the attentional patterns elicited by children's programming should be based in the specific contents and structure of a children's television clip. The medium-interactivity and high-interactivity conditions elicited different patterns of visual attention that differentially affected word learning even as they were identical until the positive feedback was given. The idea that subtle differences in the design of children's television can affect word learning is supported by previous research that suggests that differences between children's programs can affect their ability to support vocabulary growth, with shows like *Blue's Clues* and *Dora the Explorer* leading to greater vocabulary growth than shows like *Teletubbies* and *Barney & Friends* (Linebarger & Walker, 2005).

The idea that social engagement affects learning from digital media through modulating attention is further supported by previous research on the effects of sound on children's imitation from television. This work specifically points to the ways in which programs' educational efficacy can be helped or hindered by subtle changes in design that may influence attention. For example, 12- and 18-month-olds were able to imitate actions demonstrated via video clips when they included matched sound effects, but they were unable to do so when the sound effects were out of sync with the visual demonstration (Barr, Wyss, & Somander, 2009). Additionally, though infants had difficulty learning from videos with cartoon soundtracks, adding sound effects matched with actions shown on screen improved their ability to imitate actions from the videos 24 hrs later (Barr, Shuck, Salerno, Atkinson, & Linebarger, 2010). In this case, the soundtrack seemed to draw attention away from relevant information, while the added matched sound effects successfully directed attention back to the target stimuli. These data also suggest that our finding that social interactivity is beneficial for learning to the extent that it draws attention toward target information may not be specific to the social domain—any elements of media that aim to increase overall engagement may detract from learning unless they are specifically designed to draw attention toward relevant information.

Our study did not include a live-person condition and so we cannot make any definitive claims about whether increasing interactivity can completely close the "video deficit" noted in previous research (Anderson & Pempek, 2005; Kuhl, 2007; Troseth, 2003; Troseth et al., 2006). Additionally, in our design, children's learning may have benefitted from the demonstration trial. In real-world settings, children may not watch television alongside adult coviewers who explicitly tell them to try to learn words from screen media. Of course, the use of a demonstration trial could not have influenced the pattern of results observed *across* conditions. Children's better performance in the medium-interactivity condition suggests the social content of television clips themselves played a critical role in driving the attentional patterns that supported word learning. Our results indicate that the beneficial attentional effects of live social interaction can be replicated on screen to at least diminish, if not completely close, the digital learning gap.

FUNDING

The authors gratefully acknowledge the National Institutes of Health (NIH R01 MH099078 to DA) and Brown University (Solsbery Summer Research Fellowship) for their generous support of this work.

ORCID

Kate Nussenbaum 0000-0002-7185-6880

REFERENCES

Allen, R., & Scofield, J. (2010). Word learning from videos: More evidence from 2-year-olds. *Infant and Child Development*, *19*, 649–660.

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist*, *48*, 505–522.
Barr, R., Shuck, L., Salerno, K., Atkinson, E., & Linebarger, D. L. (2010). Music interferes with learning from television during infancy. *Infant and Child Development*, *19*, 313–331.
Barr, R., Wyss, N., & Somander, M. (2009). The influence of electronic sound effects on learning from televised and live models. *Journal of Experimental Child Psychology*, *103*, 1–16.
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.
Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*, 535–543.
Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*, 207–220.
Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*, 887–898.
Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, *7*(5), e36399.
Krcmar, M., Grela, B., & Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology*, *10*, 41–63.
Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*, 110–120.
Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*, 9096–9101.
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.
Linebarger, D. L., & Walker, D. (2005). Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist*, *48*, 624–645.
Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*, 115–128.
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.
Reiser, R. A., Tessmer, M. A., & Phelps, P. C. (1984). Adult–child interaction in children's Sesame Street. *Educational Communication and Technology Journal*, *32*, 217–223.
Rice, M. L., Huston, A. C., Truglio, R., & Wright, J. C. (1990). Words from 'Sesame Street': Learning vocabulary while viewing. *Developmental Psychology*, *26*, 421–428.
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.
Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). 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.
Troseth, G. L. (2003). 2-year-olds learn to use video as a source of information. *Developmental Psychology*, *39*, 140–150.
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., 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.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside..
Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*, 118–136.
Yurovsky, D. Y., Wade, A., & Frank, M. C. (2013). Online processing of speech information in early word learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. (pp. 1641–1646). Austin, TX: Cognitive Science Society).