Investigating the Use of Interactive Features for Children's Television Programming

Elizabeth J. Carter
Disney Research
Pittsburgh, PA, USA
liz.carter@disneyresearch.com

Jennifer Hyde
Disney Research
Pittsburgh, PA, USA
jdtam@cs.cmu.edu

Jessica K. Hodgins
Disney Research
Pittsburgh, PA, USA
jkh@cs.cmu.edu

ABSTRACT
As children begin to watch more television programming on systems that allow for interaction, such as tablets and videogame systems, there are different opportunities to engage them. For example, the traditional pseudo-interactive features that cue young children's participation in television viewing (e.g., asking a question and pausing for two seconds to allow for an answer) can be restructured to include correct response timing by the program or eventually even feedback. We performed three studies to examine the effects of accurate program response times, repeating unanswered questions, and providing feedback on the children's likelihood of response. We find that three- to five-year-old children are more likely to verbally engage with programs that wait for their response and repeat unanswered questions. However, providing feedback did not affect response rates for children in this age range.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Child-computer interaction; interactive multimedia for children; children's television; pseudo-interaction.

INTRODUCTION
The introduction of *Blue's Clues* in 1996 ushered in an expansion of pseudo-interaction in children's television, in which characters appear to converse with their viewers directly [5]. These conversational features typically consist of a character asking a question, pausing for a set amount of time, and then responding either in a way that does not acknowledge the child's answer specifically (e.g., presenting the correct answer and then moving on) or responding only positively (e.g., saying, "Great!" and presenting the correct answer, regardless of whether the viewer's response was correct). The massive popularity of shows like *Blue's Clues* and *Dora the Explorer* indicates that children enjoy the pseudo-interactive features and become very engaged with the shows and their characters. Moreover, the format has been shown to have positive effects on viewers (e.g., [3, 9, 16], helping ensure that this style of programming has remained on the air for two decades.

With the advent of computers and tablets, contingent reactions by the media to children's responses are now possible. Broadcast television is limited by its inability to provide flexible timelines for response, prompt a second time to attempt to get an answer, or contribute feedback. Tablets provide an opportunity to improve upon standard television techniques because they have embedded microphones that can be used to detect user speech. Moreover, children as young as two years old use touchscreen tablets to perform simple tasks [10]. While accurate speech recognition is not yet developed for preschool-aged children, detecting speech (i.e., sound) is possible.

We designed a set of three experiments to investigate the effects of using temporal contingency (i.e., having the character respond quickly after the child finishes rather than after a set amount of time), repeating unanswered questions, and providing feedback on children's engagement. In the first study, we compared an unaltered television episode containing prompts followed by a set amount of silence to a similar episode that we divided into video clips to create a program in which the amount of silence following each prompt was contingent on the child's response. After a prompt, an experimenter waited for the child to provide a verbal response before advancing to the next clip. If no response was given after 10 s, the next clip would automatically play. In the second study, we compared the temporally contingent version to a version that was temporally contingent with "reprompts". In the version with reprompts, the child would be prompted with the same question again if s/he had not responded within 8 s. These two studies mimic the use of a microphone in an app to detect speech. In the third study, we compared the effects of using speech recognition to provide feedback for correct and incorrect answers to doing only basic speech detection. We specifically focused on behaviors reflecting engagement (e.g., answering questions) because they are an important base for more complex research questions.

RELATED WORK
It has long been known that young children engage with television. Much of the previous research on program structure probed pseudo-interactive features and their effects on learning and memory in order to inform the development of educational programming. For example, adding different questions to a televised story determined that learning was facilitated

IDC 2017, June 27–30, 2017, Stanford, CA, USA.

ACM ISBN 978-1-4503-4921-5/17/06 ...$15.00.
http://dx.doi.org/10.1145/3078072.3079717

most by asking personalized, viewer-directed questions, followed by asking rhetorical questions and asking no questions [16]. Long-term viewing of programs with pseudo-interactive features has been found to be positively correlated with vocabulary size and expressive language production in toddlers [9]. Moreover, rates of responsiveness to these features during an episode correlate with story comprehension [2].

In addition to the presence of the pseudo-interactive features themselves, various aspects of the programs can affect the likelihood of response. For example, toddlers were better able to learn a task from a pre-recorded video when it was presented by a familiar, socially meaningful character rather than an unfamiliar character [6, 8]. Additionally, use of a familiar actor (their mother) has been found to help infants and toddlers imitate videorecorded actions [7]. Moreover, multiple viewings of a single episode of a program with pseudo-interactive features elicited increased verbal and nonverbal interactions from preschool children [3]. Program familiarity has also been demonstrated to affect the relationship between pseudo-interactive features and comprehension such that children who are already familiar with programs show greater comprehension benefits from pseudo-interaction cues [13].

Finally, perceived contingency matters. Contingency is a key component of interpersonal interaction. Infants younger than six months of age can distinguish and attend differently to temporally contingent interactions over videoconference than to prerecorded presentations [4]. Truly social interactions are inherently temporally contingent, to the degree that adults are sensitive to audiovisual interaction delays of 500 ms and audio delays of 400 ms when using telecommunication, beyond which conversations are considered unnatural [15]. Even very young children show influences of social contingency. Prior research found that two-year-olds who were given instructions via passive video viewing were not able to follow them as well as those given instructions in person; however, having a brief, contingent interaction over videoconference that demonstrated the instructor's ability to share attentional focus mitigated the effect [17]. Social contingency improves imitation [12], novel verb learning [14], and verbal pattern learning [11]. These results suggest that social familiarity and temporal contingency can affect engagement in and recall of interactions.

EXPERIMENT 1

In our first experiment, we manipulated the timing of a television program to increase perceived contingency and investigated the effects on verbal and nonverbal responses.

Participants

Seventeen children between the ages of 3 and 5 years (10 girls, 7 boys, age range 3.003 to 5.537 years, mean(SD) age = 4.240(0.833)) participated successfully in this research, which was approved by our Institutional Review Board. We recruited children from the local community via posted advertisements and email lists. An additional boy was enrolled in the study but not included in the analysis due to equipment failure. Participants had normal or corrected-to-normal hearing and vision. Families were compensated for their time.

Episode	Total Prompts	Repeat Phrase	Yes/No	Free Response
Egg	26	7	6	13
Birdhouse	24	7	9	8

Table 1. Number of prompts by episode.

Materials

We selected a television program, *Mickey Mouse Clubhouse*, that was directed at a preschool-aged audience (ages three to five years), contained prompts to elicit verbal responses from the children, and included characters that are familiar to most children in our culture. By using familiar characters, we hoped to maximize the children's verbal responses to the program, similar to previous research (e.g., [6, 8]). We watched several episodes of the program and annotated various categories of prompts (i.e., phrase repetition, yes-or-no questions, and free response questions) to identify two similar episodes. (See Table 1.) In one episode, the characters built a birdhouse ("Goofy the Homemaker", henceforth referred to as Birdhouse episode); in the other, they took care of an egg ("Donald Hatches an Egg", or Egg episode). In both episodes, viewers are asked directly to help the characters select tools to solve problems. The episodes contained 24 and 26 total prompts, 7 of which asked the child to repeat what the character said, with the pauses typically lasting less than 2 s (mean(stdev) $= 1.790(0.962)$, range = 0 to 5.471 s). To make the episodes short enough that a within-subjects design was feasible within one hour, we edited them down from 24 min. to approximately 20 min. while ensuring that only content irrelevant to the overarching plots was removed (e.g., the theme song) and all prompts were unaltered. The Birdhouse and Egg episodes then lasted 19 min 49 s and 20 min 14 s, respectively.

For each episode, we created two versions. For the Original version, we only edited the episodes for length, so the children saw the episodes in a similar fashion to watching them on television at home. For the Contingent version, we edited the videos into individual clips that ended after each prompt. (See Figure 1.) The clips ranged in length from 3 s to 2 min. 34 s, and the two episodes did not differ significantly in clip length (Birdhouse mean(stdev) clip length $= 41.4(44.6)$ s, Egg $= 45.9(41.1)$, $t = 0.553$, $p = 0.583$). We created software that could play the Original and Contingent versions of the episodes. When playing the Contingent versions of the episodes, the program would play a clip until its end and then start a timer for 10 s. A manual command could be sent to interrupt the timer and play the next clip immediately. If no command was sent, the next clip would automatically play at the end of the timer. We used a Wizard-of-Oz method wherein an experimenter listened to the participant and only sent the manual command to play the next clip after each prompt had received a verbal response. (Nonverbal responses were ignored, as they would be by a microphone.) If a participant failed to respond to a prompt, the experimenter let the software advance automatically to the next clip after the waiting period of 10 s. This experimenter remained out of the child's view.

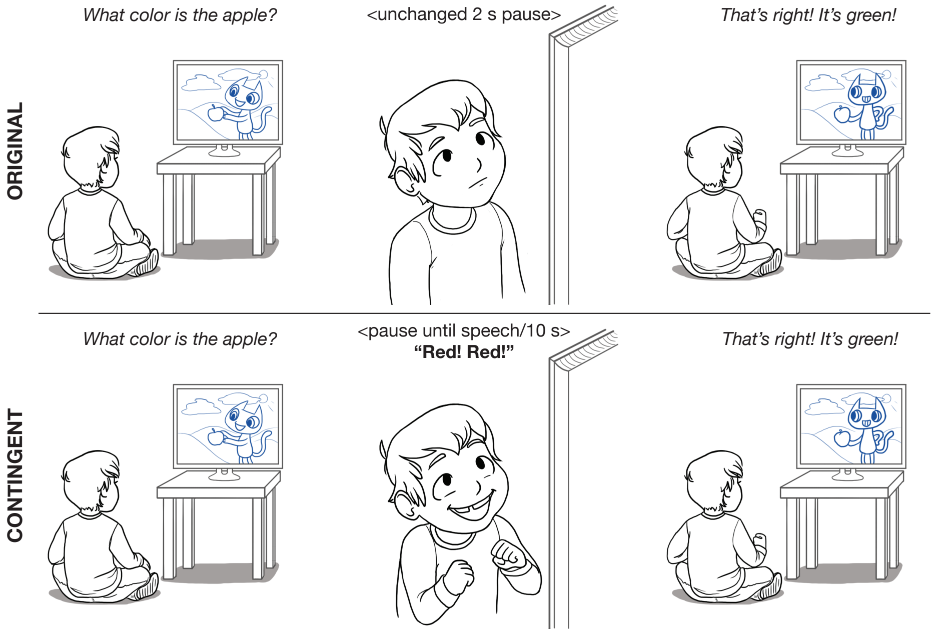

Figure 1. Illustrations of the Original and Contingent conditions.

To assess the children's memory for different characters and events throughout the episodes, we designed a quiz. After each episode, an experimenter asked the participant to put images of four scenes from the episode in chronological order. Additionally, participants were asked five questions to see if they remembered various plot points. The experimenter also asked if they thought the characters could hear them, what their favorite part was, and if they had seen the episode before.

For each participant, a parent completed a survey to assess the child's familiarity with the program: how often the child watched the show; whether the child had watched these episodes; and whether the child was familiar with each character. We also asked whether the child watched similar programs (*Dora the Explorer*, *Blue's Clues*) or used tablet applications that provide responses to prompts (e.g., ToyTalk/PullString™ apps; Disney Junior Appisodes); the child's favorite shows; if the child typically responded to televised prompts; and if the child behaved similarly to how s/he would at home.

Procedure

Each participant came into the laboratory individually for the experiment. While participants viewed the episodes, their parents completed the surveys about media habits. Parents were allowed into the experiment room, but we requested that they remain quiet and at a small distance to ensure that they did not interfere with the children's responses. Participants were recorded while viewing each episode on an Apple 27-inch monitor with an integrated camera and microphone system to ensure corresponding timestamps on the episode and the participant video. Each participant sat on a chair approximately 0.85 m from the monitor, alongside an experimenter.

Each episode of the selected program started with the same scene and two prompts ("Want to come inside my clubhouse?" and "Say the magic words...") prior to the theme song. Thus, the experimenter sitting near the child was able to indicate to children who were not responding at the beginning of either episode that they needed to respond without affecting memory for the unique content of the episode (e.g., "<Character> can't hear you, you need to speak up."). This experimenter also informed participants if they were not speaking loudly enough to be accurately recorded and heard by the wizard experimenter. Other than these instructions, the experimenter kept her own speech to a minimum by not initiating conversation and providing only minimal responses when the child initiated interactions (e.g., saying only yes or no and then attending to the monitor).

A repeated-measures design was performed to maximize statistical power. Each participant viewed two episodes of the program, one each in the Original and Contingent conditions. Episode and condition were both counterbalanced, and combinations were randomly assigned. After each episode, they completed the corresponding quiz.

Analysis

An annotator reviewed the videos for all participants and used ELAN annotation software [1] to denote: the beginning and ending of each utterance made by the child, whether each utterance was or was not a response to a prompt from the TV show, and whether each utterance in response to a prompt was a valid (expected) or unexpected response. Similarly, they denoted the same features for nonverbal behaviors (e.g., pointing or other gestures) and behaviors where the children combined verbal and nonverbal responses (e.g., pointing while saying, "That one!"). Responses including both verbal and nonverbal content were marked as combined. In some cases, children responded to an anticipated prompt before the prompt actually occurred. In cases where this response began more

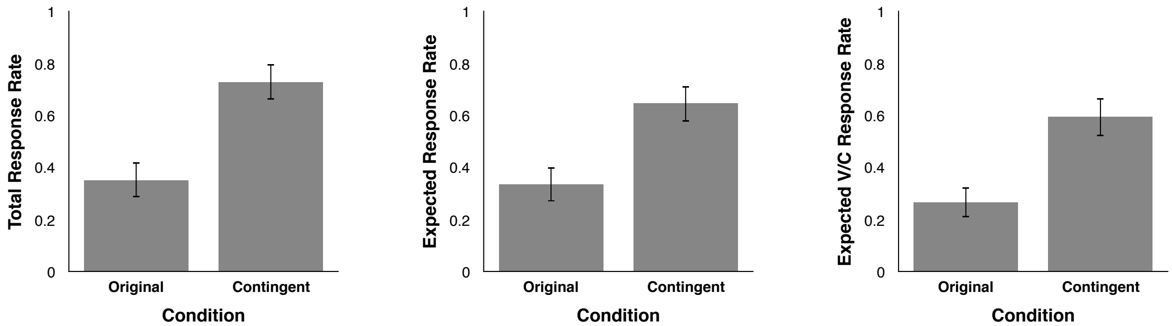

Figure 2. Response rates by condition for (a) all possible responses, (b) expected responses, and (c) expected verbal or combined responses. Error bars represent standard error in all figures.

than 800 ms before the end of the prompt, we classified it as an Early Response and considered it separately in analyses. We did not score whether any answer was correct because we did not want to penalize incorrect but valid answers (e.g., saying no instead of yes, suggesting an incorrect tool to solve a problem when it was also listed as a possibility). Additionally, in cases where a participant pointed at an object on the screen and said, "That one!", it was often impossible for the experimenter or annotator to tell where they were pointing.

We looked at the overall rate of response, rate of expected response, and response time for all categories of behavior (verbal, nonverbal, combined) together. To analyze the combined behaviors, we merged them with verbal behaviors because they included the verbal output required for contingency. This categorization simulated the use of a microphone by an app.

We analyzed the data using the Restricted Maximum Likelihood (REML) approach to ensure proper p-values for multiple comparisons and account for non-normal distribution.

Results

We examined both response rates and response times in our analyses. We omitted data from the first two prompts because they served as practice with experimenter intervention.

Response Rates

First, we examined the overall response rate, calculated as the ratio of prompts that received any response (verbal, nonverbal, or combined; expected or unexpected). The children responded significantly more to the prompts in the Contingent condition (Original mean(stdev) response rate = 0.354(0.266), Contingent = 0.730(0.272); $F = 75.335$, $p < 0.0001$). (See Figure 2 and Table 2.) In fact, higher response rates for the Contingent condition were found for 16 of the 17 children. Additionally, they responded more to the video that they saw second (First = 0.489(0.382), Second = 0.594(0.261); $F = 14.570$, $p = 0.0030$) and more to one episode of the program than the other (Egg = 0.640(0.300), Birdhouse = 0.444(0.331); $F = 15.617$, $p = 0.0023$). There was no significant effect of whether the children reported having seen an episode before (which occurred for 12 children for the Egg episode and 9 for the Birdhouse episode). When we limited our analysis to only expected (verbal, nonverbal, or combined) responses, we found the same pattern of effects with a greater response to Contingent than Original videos (Original mean(stdev) response rate = 0.336(0.259), Contingent = 0.647(0.279); $F = 53.779$, $p < 0.0001$), to the video they saw second (First = 0.467(0.360), Second = 0.526(0.263); $F = 9.401$, $p = 0.0108$), and to the Egg episode than the Birdhouse episode (Egg = 0.570(0.291), Birdhouse = 0.414(0.314); $F = 10.384$, $p = 0.0081$). Again, prior viewing had no significant effect.

We then looked at effects in the various response categories. For these analyses, the combined responses were analyzed with the verbal responses because they contained verbal content that a microphone could hear in a system similar to current apps doing speech detection. We saw the same pattern in the verbal/combined responses such that the rate was higher for the Contingent condition (Original mean(stdev) response rate = 0.283(0.245), Contingent = 0.675(0.295); $F = 69.625$, $p < 0.0001$), to the video they saw second (First = 0.450(0.378), Second = 0.508(0.290); $F = 7.150$, $p = 0.0222$), and to the Egg episode (Egg = 0.605(0.307), Birdhouse = 0.353(0.317); $F = 24.572$, $p = 0.0005$). Prior viewing had no effect. The same pattern was found for expected verbal/combined responses (Original mean(stdev) response rate = 0.268(0.236), Contingent = 0.595(0.300); $F = 62.462$, $p < 0.0001$; First = 0.418(0.356), Second = 0.445(0.274); $F = 5.816$, $p = 0.0347$; Egg = 0.534(0.300), Birdhouse = 0.329(0.301); $F = 20.347$, $p = 0.0009$).

For the 64 observed nonverbal responses, we did not see a similar pattern. There were no significant effects for condition, order, episode, or prior viewing, but there was a significant interaction between condition and order such that the children responded nonverbally more often when they saw the Contingent version first than second ($F = 6.437$, $p = 0.0282$).

	Original	Contingent	F	p
Overall Response Rate	0.354	0.730	75.335	< 0.0001
Expected Response Rate	0.336	0.647	53.779	< 0.0001
Expected V/C Response Rate	0.283	0.595	69.625	< 0.0001
Overall Response Time	0.584	2.243	25.668	0.0004
V/C Response Time	0.673	2.110	18.754	0.0001

Table 2. Response data by condition.

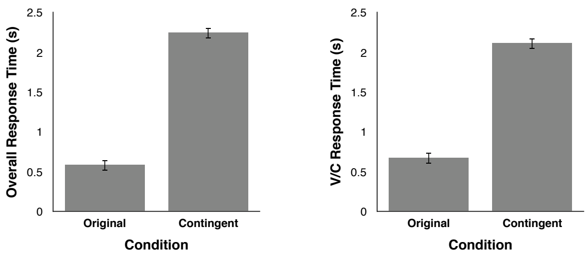

Figure 3. Response times by condition for (a) all responses and (b) verbal and combined responses only.

The rate of Early Response (37 occurrences) was not affected by condition, order, episode, prior viewing, or the interaction between condition and order.

Response Times

When we examined the time that elapsed between the end of a prompt and the beginning of a response, we found a significant effect of condition that indicated slower responses in the Contingent condition (Original mean(stdev) response time (s) $= 0.584(0.305)$, Contingent $= 2.243(0.962)$; $F = 25.668$, $p = 0.0004$). (See Figure 3.) There were no other significant effects. When we examined verbal/combined responses alone, we similarly found only an effect of condition (Original $= 0.673(0.431)$, Contingent $= 2.110(1.053)$; $F = 18.754$, $p = 0.0001$). In the Original condition, the maximum response time to any prompt was 3.667 s. For the Contingent condition, the maximum response time to any prompt was 10.317 s. We found that 94.4% of responses in the Contingent condition occurred in under 8 s. When we compared the response times in the Contingent condition to the pauses provided in the Original condition, we found that 48.9% of the responses in the Contingent condition began after the pause in the Original condition would have ended.

Response Trends

To determine whether the children understood or were aware that the Contingent episode required a verbal response to continue, we examined longitudinal patterns in the children's responses. The response rates to the first two prompts were similar and very high (Original mean response rate $= 0.618$, Contingent $= 0.688$) because the children who did not speak on their own during these two prompts in both conditions were encouraged to do so by the experimenter. We paired the subsequent ten prompts in order to examine approximately half of each episode, assuming that it would take fewer than 12 prompts for children to realize that the video was advancing when they responded. The results are displayed in Figure 4. An analysis of variance (ANOVA) revealed significant main effects of condition ($F = 65.342$, $p < 0.0001$), prompt pair ($F = 2.665$, $p = 0.0240$), episode ($F = 8.300$, $p = 0.0046$), and the interaction between prompt and condition ($F = 2.473$, $p = 0.0343$). This interaction shows that children become significantly less likely to respond to a question in the Original condition as time passes relative to the Contingent condition.

Questionnaire Results

In addition to using the responses on the parent questionnaires and children's quizzes to determine which children were familiar with the episodes, we examined them to put the results in context. Three children provided no expected verbal, nonverbal, or combined responses while viewing either condition. From the parent questionnaires, we learned that two of those children had no prior experience with pseudo-interactive television (or, in one case, any television). The third child was shy and the parent reported that she was not behaving and speaking as she would at home. Additionally, we found no significant correlations between correct answers to the memory questions in the children's quiz (mean(SD) $= 2.42(1.12)$) and whether the parents or children reported that they had seen the episodes before. Finally, there were no significant differences in the number of correct quiz questions by condition or episode. However, we did find a significant correlation between age and total score ($r = 0.654, p = 0.002$), which may have overwhelmed any other potential effects.

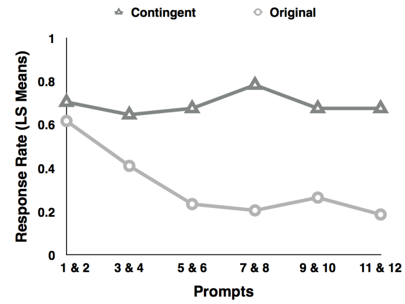

Figure 4. Response rates by prompt for the first twelve prompts in the Contingent and Original conditions.

Discussion

Overall, whether children were viewing an episode in the Contingent versus Original condition had many effects. Children had higher response rates, expected response rates, verbal/combined response rates, and expected verbal/combined response rates in the Contingent condition. Additionally, they had longer response times (overall and verbal/combined) during the Contingent condition, nearly half of which occurred after the original pause lengths. Together, these findings suggest that children are far more likely to speak to the television when the episode is waiting for them and they have time to do so. Moreover, they realize that the episode is or is not waiting for their responses within the first several prompts. Additionally, the pauses in typical television programs may preempt responses by providing insufficient time for children to formulate and speak their answers before proceeding, although future research should investigate this question more directly. Typically, the programs wait only up to two seconds for responses. Our findings suggest that even when children answer, it is often not within this time window, and contingent programs should wait up to 8 s.

The children were also more likely to provide any response, expected responses, and verbal/combined responses for the second episode that they saw in the experiment relative to the first. These findings suggest that there may be a warm up period in which children are learning about and adjusting to the technology, after which response rates increase.

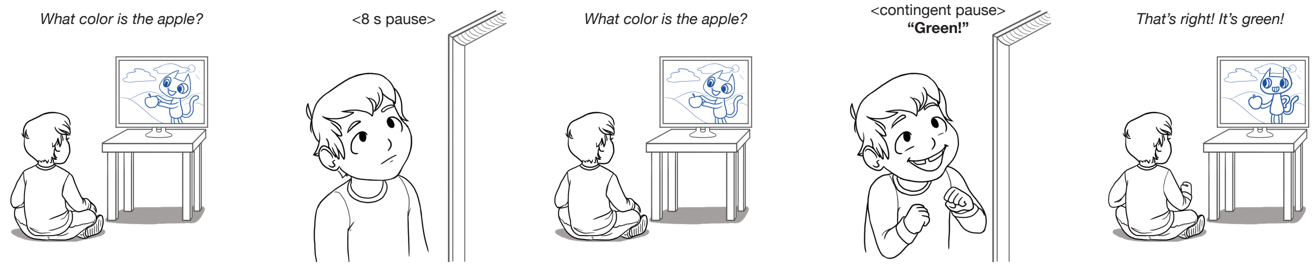

Figure 5. An example of what happened if a child did not respond to a prompt in the Reprompt condition.

Our findings also suggest that one episode contained more challenging prompts than the other. The response rate, expected response rate, verbal/combined response rate, and expected verbal/combined response rate were all higher for the Egg episode than the Birdhouse episode. These results suggest that the Birdhouse episode was more difficult.

Finally, some variables had no effect on response types, rates, or timing. Prior viewing of an episode did not seem to be reflected in behavior, and nonverbal behaviors were generally unaffected by experimental manipulations. Given that there was no positive reinforcement for nonverbal behaviors in either condition, the lack of a significant effect was unsurprising and also indicated that the children only adjusted to the reinforced (i.e., verbal) behavior.

EXPERIMENT 2

Our second experiment investigated whether repeating questions that did not originally receive a response affected the likelihood of eliciting verbal responses from children.

Participants

We enrolled 16 children between the ages of 3 and 5 years for this study (9 girls, 7 boys, age range 3.485 to 5.638 years, mean age $= 4.583$, SD $= 0.655$). Five additional children were not included in analyses due to equipment failure. Eleven of the children participated in Experiment 1, and five children previously participated in Experiment 3. Participants waited at least one month between visits (mean(stdev) $= 65(25)$ days, range 37 to 113 days). Recruitment and consent procedures were identical to Experiment 1.

Materials

We used the same selection methodology to choose the two episodes included in Experiment 2, "Pete's Beach Blanket Luau" (referred to as Luau episode) and "Mickey's Roundup" (Roundup episode). These videos were matched on the number and structure of prompts (see Table 3) and had similar clip lengths (Luau $= 37.8(51.769)$ s, Roundup $= 43.8(46.346)$ s, $t = 0.452$, $p = 0.653$). The Contingent condition remained the same; however, we replaced the Original condition with a Reprompt condition that had additional clips. Thus, participants viewed two episodes each, one Contingent and one Reprompt. For the Reprompt condition, the software would replay a clip containing only the prompt from the end of the previous clip if the participant remained silent for 8 s (Figure 5). We decreased this interval from 10 s to 8 s based on the response time results from Experiment 1. The experimenter would play the next clip

Episode	Total Prompts	Repeat Phrase	Yes/No	Free Response
Luau	28	5	5	18
Roundup	27	6	8	13

Table 3. Number of prompts by episode.

as soon as a verbal response was produced; alternatively, the software automatically advanced in the absence of a verbal response after an additional 8 s.

Procedure

We followed the same procedure as in Experiment 1 for viewing. The episodes and conditions were again counterbalanced, and participants completed a similar quiz after each episode. We administered the parent questionnaire to new participants.

Analysis

We used an identical coding scheme, analysis method, and statistical model as in Experiment 1, with the addition of noting which clips consisted of reprompts where the previous prompt was repeated.

Results

In order to get a full picture of the effects of reprompting, we specifically examined individual instances of reprompts after performing our initial analyses.

Response Rates

Across both the Contingent and Reprompt conditions, participants showed a high overall response rate (mean(SD) $=$ $90.8(0.162)\%$) to the first instances of the prompts. There was a similarly high verbal/combined response rate ($89.1(0.174)\%$) and expected verbal/combined response rate ($85.727(0.177)\%$). We examined the total response rates, verbal/combined response rates, and expected verbal/combined response rates to the prompts that all children saw (i.e., without reprompt clips) and found no significant effects of condition, episode, order, or their interactions (all $p > 0.05$).

Response to Reprompts

Eleven of the 16 participants saw at least one reprompting clip. The other six children provided a verbal response to all of the original prompts during the Reprompt condition. Forty-one instances of reprompting occurred throughout the entire experiment, for which there were 440 opportunities in the Reprompt condition. For the 41 instances, two had an interval

Participant	Instances	Responses	Expected V/C
1	1	1	0
2	5	3	3
3	8	8	8
4	2	2	1
5	4	4	4
6	2	2	2
7	7	4	4
8	2	2	1
9	2	1	0
10	5	5	5
11	1	0	0
Total	39	32	28

Table 4. For the participants who saw at least one reprompt clip, the number they saw (Instances) and the number to which they subsequently responded in any way (Responses) and in an expected verbal or combined way (Expected V/C).

between the reprompt clip and the following clip of under 0.2 s, indicating that they answered during the reprompt clip before it noticeably played. Thus, only 39 reprompt clips were analyzed (Table 4). Of the 11 children who saw reprompts, the number of reprompts ranged from 1 to 8 (mean(stdev) = 3.545(2.423)). In 32 of these cases, the participants responded in the interval following the reprompt clip; 28 times with an expected verbal or combined response. Participants did not respond to the reprompt clip in the remaining six cases. Thus, 32 of 39 (82.1%) of the experienced reprompts elicited responses, 28 (71.8%) of which were appropriate and included verbal input.

Because only a subset of the children ever saw a Reprompt clip, we analyzed their data separately. These 11 children did not show significant differences in their overall response rates, verbal/combined response rates, and expected verbal/combined response rates by condition. However, we note for future research that the means for all response rates were greater in the Reprompt condition than for the Contingent condition (90.693 vs. 85.606, 88.396 vs. 83.297, and 85.750 vs. 78.692).

Questionnaire Results

Because the parent questionnaire and children's quiz responses for Experiment 1 did not affect results, we did not integrate questionnaires into these analyses. We do note that every child in this experiment spoke in both conditions, including the three who had not provided expected responses in Experiment 1.

Discussion

In Experiment 2, we showed the children two video conditions, both of which waited for them to provide a verbal response before continuing. In the Contingent condition, the episode would continue after eight seconds regardless of whether the child provided a response; in the Reprompt condition, the episode would repeat the prompt in the absence of a verbal response and wait up to eight additional seconds for a response before continuing. The overall rates of verbal response were quite high in this experiment, likely because all but two of the children had previously experienced contingent programming in the laboratory. The children who did see reprompts provided responses to the majority of these instances, and most of those responses were expected verbal or combined responses. These results suggest that building reprompts into pseudo-interactive programming helps maintain engagement.

Figure 6. Zoe and Bigbot.

EXPERIMENT 3

After determining that children are sensitive to temporal contingency, we wanted to examine the effects of providing feedback (as contingency of content) on children's responses. We created a new animated short film with clip playback controlled by a WoZ system. After completing an interactive task with or without feedback, participants took a quiz to assess their memory, generalization of material, and experiences.

Participants

Participants were recruited using online and physical bulletin boards in the local community. Twenty-four children between the ages of four and five years (12 boys, 12 girls, $M = 6.997, SD = 1.744$) successfully completed this experiment, four of whom previously completed Experiment 1 and/or 2. Additional participants were not included in our analyses because of nine technical failures, one case of parent interference, and one behavior issue. This research was approved by our Institutional Review Board, and families were compensated.

Children participated in this study as part of a larger series of studies that additionally included two experiments on interaction with robots and one experiment on storytelling. We chose to combine these studies into a single research session because they were each very brief and one of the child-robot interaction studies required repeated, separate interactions with the robot. The Bigbot activity was not integrated with or similar to any previous experience in the session. Together, the entire session took approximately 50 minutes, up to 10 of which was spent in the Bigbot study.

Animations

In collaboration with a team of animators, we created an animated program for children starring a human character named Zoe and her two robot friends, Smallbot and Bigbot. The program started by showing Zoe, Smallbot, and Bigbot in their treehouse. The viewer had to help Zoe and Smallbot identify a problem with Bigbot (a dead battery) and then get him out of the treehouse, across a river, and up the stairs into a laboratory

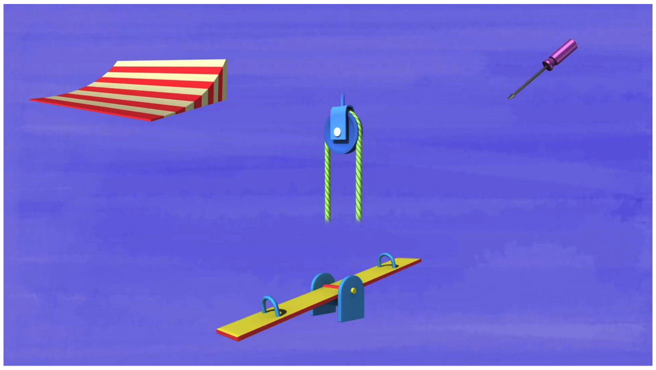

Figure 7. The toolbox shown to the children

building so that they could install a new battery. To achieve these goals, children were asked what they believed the original problem was and then selected tools from a virtual toolbox belonging to Smallbot to solve subsequent problems. All children saw the same base clips between the question and answer segments; these clips lasted between 10 and 53 s, totaling 3 min 18 s. The children saw a toolbox with only the four tools they would need during the episode. The questions themselves were therefore phrased such that the children, who only had one item left in their toolbox for the last question, were asked if that specific tool would work to solve the problem instead of being asked which tool would solve that problem. In total, all children saw five prompts, one about identifying Bigbot's problem and four about choosing tools to help him on his way.

No Feedback. For the *No Feedback* condition (11 children), we created a set of short clips that would play after each question was posed to the children. All children in this condition saw the same clips, regardless of their responses. In these clips, Zoe neutrally presented the correct answer to the question (e.g.,"Hmm, I think we should try the pulley."). These clips lasted between 1 and 5 s.

Feedback. For the *Feedback* condition (13 children), we created a larger set of short clips that were played after the children were given the opportunity to respond. This set included four types of responses for Zoe to present:

1. Repeating the question (played if the child did not respond within 8 s)
2. Positive feedback (played for correct answers–e.g., "I think you're right, the see-saw could work if we use it to launch him across.")
3. Negative feedback (played for incorrect answers–e.g., "I think we should try something else. What might work better?")
4. Neutral feedback (played in the absence of a verbal response even after the question was repeated a second time–identical to the clips used in the No Feedback condition)

These clips were 1 to 3 s in length.

Questionnaires

Upon completion of the experiment, the experimenter asked the children ten questions. The children were given a sheet of paper showing the contents of Smallbot's toolbox that they had seen earlier. The first four of these questions were about which tools were used to solve specific problems during the program. The subsequent four questions asked which tool the participants would use to solve new tasks that were similar to those shown in the program. Finally, the participants were asked whether they thought Zoe could hear them and what their favorite part was.

Procedure

Participants entered the lab and were seated on a chair approximately 0.85 m in front of a 27-inch Apple monitor with an integrated camera and microphone system. The camera and microphone were used to record the participants throughout the experiment. An experimenter sat next to the participant and ensured that the child was comfortable and ready before proceeding. The experimenter gave participants the following instructions, "You will be watching a cartoon about Zoe and her robots and you're going to help her. In order to help her, you are going to need to use your big voice. Do you have a big voice? [Pause for response] Great! Let's start."

At that point, the wizard would play the first clip of the cartoon. If the child spoke too quietly for the wizard to hear clearly, the experimenter would remind the child that s/he needed to use a big voice. Otherwise, the experimenter strove not to speak during the cartoon and would direct attention toward the screen if the child looked at her. In cases where speech was unavoidable (e.g., a participant directly asked her a task-related question), she responded as minimally as possible (e.g, supplying a word for an item) and then redirected the child's attention to the screen.

In the *No Feedback* condition, the wizard advanced to the neutral version of the next clip after the child spoke; if no speech occurred, a reprompt occurred after 8 s of silence. For the *Feedback* condition, she selected the appropriate version of the following clip (positive, negative) or let the reprompt and neutral response occur based on the child's reaction. Because the wizard was replacing a speech recognition module, gestures (such as tapping the monitor) were ignored.

After the final clip of the animation, the experimenter administered the questionnaire to the child and displayed the image of the toolbox for all questions.

Analysis

An annotator watched all of the recordings of the children completing the experiment to note the timing and occurrence of multiple features. For verbal features, she transcribed the children's speech, noted to whom it was directed, and marked whether each instance was a correct response to a prompt, an incorrect response to a prompt, or unrelated speech. She also annotated nonverbal behaviors, including gestures (e.g., shrugs, points, etc.), the person or thing at whom they were directed (e.g., Zoe, experimenter), and whether the behavior was an accurate response, an inaccurate response, or unrelated to a prompt. Finally, the annotator tracked when the experimenter spoke and whether the speech was to reiterate instructions.

Additionally, we analyzed the questionnaire data to determine the number of correct answers for the memory questions, the number of correct answers for the generalization questions, and whether the participants thought Zoe could hear them.

Results

We compared the response patterns for the Feedback and No Feedback conditions as well as examining responses within the Feedback condition to determine the effects of feedback.

Response Rates

The response rates were high for both conditions (total mean(SD) = 99.17%(4.08), Feedback mean(SD) = 98.33(5.77), No Feedback mean(SD) = 1.00(0)). REML analysis showed no significant difference in response rates between conditions (p = 0.385). The verbal/combined response rates were also similar across conditions (Feedback mean(SD) = 88.33(1.80), No Feedback mean(SD) = 90.00(13.48), $F = 0.021$, $p = 0.886$). Condition also did not impact correct verbal/combined response rates (Feedback mean(SD) = 68.33(15.86), No Feedback mean(SD) = 73.33(23.09), $F = 0.328$, $p = 0.573$).

Response to Reprompts

There were 13 instances in which children failed to respond and saw reprompts (mean(SD) = 0.542(0.588) reprompts per child). Eleven children saw one reprompt, one child saw two reprompts, and the remaining twelve children saw no reprompts. For the 13 instances of reprompts, 12 (92.31 %) garnered responses, 11 (84.62 %) elicited verbal responses, and 5 (38 %) received correct verbal responses. There were no differences between conditions for the number of reprompts seen or the likelihood of a response to a reprompt or a verbal response to a reprompt ($p > 0.20$). All five correct verbal responses to a reprompt were given by children in the Feedback condition; however, the number of instances is too small to be informative.

Reactions to Negative Feedback

The 13 children in the Feedback condition saw a total of 13 instances of Negative Feedback (two saw 0 instances, eight saw 1 instance, one saw 2 instances, and one saw 3 instances). In 10 of 13 instances (76.92%), children verbally provided the correct answer after receiving negative feedback.

Questionnaire Results

On the four recall questions, overall accuracy was high (mean(SD) = 3.88(0.34)), and a t-test revealed no difference between conditions. Similarly, the responses generalization questions were largely accurate (mean(SD) = 3.38(1.05)) and there was no significant difference between conditions.

Discussion

In Experiment 3, temporal contingency elicited high overall, verbal/combined, and correct response rates. These rates were unaffected by whether the children received feedback on the accuracy of their answers. Once again, reprompts elicited high rates of response, although the responses were often incorrect, potentially suggesting that children failed to respond to the first prompt because they did not know the answer. Additionally, children provided accurate responses to quiz questions.

GENERAL DISCUSSION

We performed three Wizard-of-Oz experiments to determine how increasing contingency affects preschool children's responses while viewing pseudo-interactive television programs. In Experiment 1, we found that adding temporal contingency–waiting for the child to provide a response before continuing the episode–approximately doubled the response rates relative to typical viewing, in which the episode moves on automatically after a set amount of time. Our second experiment revealed that repeating a prompt when a child fails to answer elicits a response the majority of the time in a contingent condition. In Experiment 3, the presence or absence of feedback did not affect any response rates, but we confirmed that prompt repetition is effective at eliciting responses.

Design Implications

Our findings inform the design of pseudo-interactive television programs and apps for preschool-aged children. First, we recommend that any available microphone be used to detect the presence or absence of speech. Contingent timing of media increased children's verbal responses, which in turn may increase engagement and enjoyment of programming. This feature could be integrated into tablet applications and apps on video game consoles, such as the XBox Kinect, both of which have access to microphones to capture speech input. Although considering the use of varied devices will introduce a challenge for content producers, the content will likely reach a broader audience. Given that a few children in Experiment 1 who had no previous exposure to pseudo-interactive programming showed low levels of verbal responses even after the experimenter gave introductory instructions, it may also be beneficial to integrate instructions during early exposure to this programming and repeat those instructions as needed so that children learn to respond. Additionally, we recommend repetition of questions in the absence of responses to re-engage children who might be distracted. However, it might be wise to limit this repetition if children persistently do not answer to avoid making the experience unpleasant. It could be limited to only key questions and plot points or a subset of missed responses.

We also recommend the use of this technology in a few specific contexts, specifically for speech elicitation. For example, encouraging young children's speech in a fun, engaging context could be used to improve therapeutic interventions. An app with temporal contingency could be designed to boost rehearsal of specific sounds and words for children in speech therapy. Additionally, adding temporal contingency to interactions could be useful for researchers who need children's speech data, such as by being integrated into prompts to elicit speech to build datasets and train speech recognition technology for this age group. The quantity of data that can be attained in an app far exceeds that recorded in person.

Limitations and Future Work

In this research, we aimed to increase children's responses to pseudo-interactive media by manipulating temporal contingency and media responses. Because prior work demonstrated that children who respond more to educational media have better comprehension in some cases [2], future work should explicitly investigate whether the increased level of response specifically resulting from contingency can also improve comprehension. It is possible that this method of increasing engagement could result in additional learning from and memory for media. Although we assessed memory, we were unable to do so in a way that accounted for prior knowledge (e.g., already knowing what "aloha" means) in this research. Moreover, we likely had a ceiling effect for the quiz results due to a combination of straightforward questions and prior knowledge. Future work should use novel media to which children have not had prior access and unusual content, along with a pre/posttest design, to examine this point.

Future research also could examine other aspects of media that may influence children's responsiveness and engagement. For example, characters in typical pseudo-interactive programs often assume that the viewer answered a question correctly and compliment them, even when this is not the case. This process could result in children being reinforced for incorrect answers and learning inaccurate information. Subsequent experiments could assess whether it is feasible to provide feedback for answers. While speech recognition for very young children currently has very low accuracy rates, it is possible that even occasional feedback (e.g., for a yes/no question where a speech recognizer might plausibly be developed) could positively affect the perception of contingency. Moreover, expectations of accurate understanding by the system could be influenced by age, so a larger age range of children should be examined.

Additionally, we relied on a Wizard-of-Oz procedure because we wanted to quickly determine whether adding these various types of contingency would be a worthwhile endeavor. Future research should test whether technological solutions can provide similar response times and effects as a human listener.

A final issue in the line of research is whether children's behavior in the laboratory is comparable to their behavior at home. Future work could include more in-depth interactions with parents to ascertain this information, and standalone apps developed in the future could be sent home for testing over longer time periods and more naturalistic settings.

ACKNOWLEDGMENTS

We thank our participants. We are grateful to A. Rivera, K. McIntosh, M. Landlord, M. Mahler, J. Krahe, B. Karp, J. Lehman, and I. Leite for their assistance.

REFERENCES

1. Hennie Brugman and Albert Russel. 2004. Annotating Multi-media/Multi-modal Resources with ELAN. In *LREC, Fourth International Conference on Language Resources and Evaluation*.

2. Sandra L Calvert, Bonnie L Strong, Eliza L Jacobs, and Emily E Conger. 2007. Interaction and participation for young Hispanic and Caucasian girls' and boys' learning of media content. *Media Psychology* 9, 2 (2007), 431–445.

3. Alisha M Crawley, Daniel R Anderson, Alice Wilder, Marsha Williams, and Angela Santomero. 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, 4 (1999), 630.

4. Sylvia MJ Hains and Darwin W Muir. 1996. Effects of stimulus contingency in infant-adult interactions. *Infant Behavior and Development* 19, 1 (1996), 49–61.

5. Dade Hayes. 2008. *Anytime Playdate: Inside the Preschool Entertainment Boom, or, How Television Became My Baby's Best Friend*. Simon and Schuster.

6. Alice Ann Howard Gola, Melissa N Richards, Alexis R Lauricella, and Sandra L Calvert. 2013. Building meaningful parasocial relationships between toddlers and media characters to teach early mathematical skills. *Media Psychology* 16, 4 (2013), 390–411.

7. Marina Krcmar. 2010. Can social meaningfulness and repeat exposure help infants and toddlers overcome the video deficit? *Media Psychology* 13, 1 (2010), 31–53.

8. Alexis R Lauricella, Alice Ann Howard Gola, and Sandra L Calvert. 2011. Toddlers' learning from socially meaningful video characters. *Media Psychology* 14, 2 (2011), 216–232.

9. Deborah L Linebarger and Dale Walker. 2005. Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist* 48, 5 (2005), 624–645.

10. Michael Cohen Group and USDOE. 2011. *Young Children, Apps and iPad*. Michael Cohen Group.

11. Lauren J Myers, Rachel B LeWitt, Renee E Gallo, and Nicole M Maselli. 2016. Baby FaceTime: Can toddlers learn from online video chat? *Developmental Science* (2016).

12. Mark Nielsen, Gabrielle Simcock, and Linda Jenkins. 2008. The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science* 11, 5 (2008), 722–731.

13. Jessica Taylor Piotrowski. 2014. Participatory cues and program familiarity predict young children's learning from educational television. *Media Psychology* (2014), 311–331.

14. Sarah Roseberry, Kathy Hirsh-Pasek, and Roberta M Golinkoff. 2014. Skype me! Socially contingent interactions help toddlers learn language. *Child Development* 85, 3 (2014), 956–970.

15. Jennifer Tam, Elizbeth Carter, Sara Kiesler, and Jessica Hodgins. 2012. Video increases the perception of naturalness during remote interactions with latency. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2045–2050.

16. Ron Tamborini and Dolf Zillmann. 1985. Effects of questions, personalized communication style, and pauses for reflection in children's educational programs. *The Journal of Educational Research* 79, 1 (1985), 19–26.

17. Georgene L Troseth, Megan M Saylor, and Allison H Archer. 2006. Young children's use of video as a source of socially relevant information. *Child Development* 77, 3 (2006), 786–799.