Conversing with Children: Cartoon and Video People Elicit Similar Conversational Behaviors

Jennifer Hyde[1], Sara Kiesler[2], Jessica K. Hodgins[1,3,4], and Elizabeth J. Carter[3]

[1]Computer Science Department,
[2]HCI Institute, [3]Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{jdtam, kiesler, ejcarter}@cs.cmu.edu

[4]Disney Research, Pittsburgh
4720 Forbes Avenue
Lower Level, Suite 110
Pittsburgh, PA 15213
jkh@disneyresearch.com

ABSTRACT

Interactive animated characters have the potential to engage and educate children, but there is little research on children's interactions with animated characters and real people. We conducted an experiment with 69 children between the ages of 4 and 10 years to investigate how they might engage in conversation differently if their interactive partner appeared as a cartoon character or as a person. A subset of the participants interacted with characters that displayed exaggerated and damped facial motion. The children completed two conversations with an adult confederate who appeared once as herself through video and once as a cartoon character. We measured how much the children spoke and compared their gaze and gesture patterns. We asked them to rate their conversations and indicate their preferred partner. There was no difference in children's conversation behavior with the cartoon character and the person on video, even among those who preferred the person and when the cartoon exhibited altered motion. These results suggest that children will interact with animated characters as they would another person.

Author Keywords

Children; conversation; behavior; avatar; agent; animated character; facial motion

ACM Classification Keywords

H.4.3. Communications Applications: Computer conferencing, teleconferencing, and videoconferencing; H.5.1. Multimedia Information Systems: Evaluation/methodology

INTRODUCTION

Over the past decade, the use of animated human characters has increased dramatically in education, entertainment, and therapy. They assist customers on shopping websites, occupy virtual worlds on behalf of users in games such as *The Sims* and *World of Warcraft*, serve as teachers or virtual peers in educational software, and act as mock job interviewers who provide feedback to users. Animated characters can provide corporeal representations when none are available, impart anonymity, and add entertainment value. People's interactions are complex, with verbal and nonverbal cues that can often be very subtle and meaningful. The benchmark of a successful conversational agent, as proposed by Cassell and Tartaro [12], is if an agent can interact with a people similarly to how people interact with one another.

Extensive research has focused on how additions of human-like behaviors to agents create more successful human-agent interactions. These studies have incrementally tested how adding various behaviors, such as smiles, emotional facial expressions, gaze, and mirroring, can influence human behavior [1, 8, 18, 20]. Although researchers have shown that these behaviors improve human-agent interaction, they have not shown that these interactions are similar to human-human interaction. These studies also focused on how to improve adult human interactions with agents; however, many animated characters are created for children.

Much of children's entertainment and educational programming features interactive or pseudo-interactive animated characters. As of 2003, approximately 70% of children under two had watched television, and over 90% of childrn have done so by age six [31]. The characters often wear bright colors and have simple, exaggerated features. The bright colors are supposed to grab children's attention, and the simple, exaggerated features (e.g., large eyes) are supposed to help children focus on particular parts of the screen [21]. These characters have also been given human-like behaviors (e.g., blinking) to make them seem more "alive" and interactive. It is unclear whether these characters are actually more appealing or engaging than real people. Is Steve from *Blue's Clues* a better host than Dora from *Dora the Explorer*? What about Fred Rogers and Daniel Tiger from *Mr. Rogers'/Daniel Tiger's Neighborhood*?

Some previous research has compared how children speak differently to real and animated people. The researchers identified possible differences in children's language patterns, but the researchers could not explain whether the differences were due to different language patterns in the real and animated partners or appearance differences [4, 29]. We ran an experiment to examine children's preferences, attention, and language when conversing with an adult partner who appeared via videoconference as herself and as a cartoon character. The cartoon characters' facial motion was driven by

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.

ACM 978-1-4503-2473-1/14/04..$15.00.
http://dx.doi.org/10.1145/2556288.2557280

our adult confederates in realtime; thus, we removed the issue of natural language generation, allowing us to focus on the appearance of the characters and their motion to better understand their effects on children's interactions. In our experiment, character appearance was a within-subjects variable and motion was a between-subjects variable. The characters displayed human behaviors, including facial expressions and mirroring. We also manipulated facial motion magnitude by exaggerating and damping the characters' spatial motion. Exaggerated facial motion magnitude has been associated with easier emotion recognition and perceptions of increased emotional intensity [3, 9, 22]. By exaggerating our characters' faces, we hoped to make them more expressive.

In contrast to previous research, we found that despite having strong preferences about confederate appearance, the children behaved similarly across conditions. Our participants behaved and conversed with animated and real people in similar ways, suggesting that appearance does not affect children's speech and that animated characters could be suitable substitutes for real people in conversational applications. Moreover, this research highlights the importance of using both self-report and behavioral measures when conducting experiments about design.

RELATED WORK

Children interact with animated characters on a frequent basis. Often, these characters are on television shows or in educational games, in which case the interactions are staged and the characters cannot respond to unexpected behaviors. Simulated interactions on television are important because they help engage children and improve learning [24, 26]. Since the successful launch of *Blue's Clues*, children's educational television programs have followed a similar format of characters looking directly at the viewers, asking questions, pausing, and then "acknowledging" viewers' responses (for a review, see [21]).

Although children will interact with animated characters, they do not always treat these characters as they would other people [4, 29]. For example, a prior study found that children used more gestures and words when conversing with adults than with computer characters [4]. As intelligent agents become more human-like, the hope is that children will interact with the agents as they do real people.

Children learn from a young age to pay attention to socially relevant information, including eye gaze, gestures, and emotional displays (for a review, see [2]). Children's comprehension of facial expressions develops over time (for a review, see [37]). When children are between two and three years old, facial expressions are categorized as positive or negative. The number of categories grows and the criteria for each category narrow as the children develop. Most 4- and 5-year-olds can recognize basic emotions on the face although their accuracy may not be very high. By the time children are 9 and 10 years old, they are almost as accurate as adults [25, 37]. Because exaggeration can make facial expressions easier to recognize [10, 22], we hypothesized that exaggerating the facial motion of animated characters may improve child-character interactions.

Although young children pay attention to people who are physically present, there has been a question as to how children view people who appear on screens. As adults, we understand that people who are not physically present may still provide useful information to us. For example, we listen and learn from newscasters about what traffic to avoid or how to prepare for the weather. Children, on the other hand, do not necessarily listen and learn from people on television, especially if there is no interaction. Troseth and colleagues [36] conducted several studies in which two-year-old children watched people on monitors or in the same room give useful hints for a game. Only the physically present people and the people on monitors who interacted with the children were able to get the children to use the hints. These results emphasize the importance of interaction if young children are to pay attention to animated characters.

Very little research has compared how children interact with animated characters and other people. Oviatt [29] conducted a study with ten 6- to 10-year-old participants. The researcher compared how children spoke with an animated character to an adult experimenter, and she found that children had fewer disfluencies in their speech with the characters than with the adult. Although participants were encouraged to ask questions in both interactions, the tasks were not identical. In the character condition, participants spoke with multiple animated animals to learn more about the different species. When participants spoke to the experimenter, they played a game of Twenty Questions where the children asked the questions.

In another experiment comparing child-computer interaction to child-human interaction, Black and colleagues [4] compared nine 4- to 7-year-olds' interactions with an animated human agent on a computer to those with an adult experimenter in person. The comparison tasks were more similar: there was a single animated human character who asked questions that were similar to the questions that the adult experimenter asked. The researchers found that children were less verbose and spoke slower when speaking with the agent rather than the adult; however, the researchers also noted that their agent used fewer words and spoke slower than the adult partner. This work supports the idea that children will emulate the speaking style of their conversational partners, as suggested by other researchers [16, 17]. Additionally, they reported that the children looked away from the adult more than from the animated human character. To build upon these two previous studies, we wanted to examine more precisely how children's patterns of engagement would differ between a human and an animated character by having the children take part in two nearly identical interactions using the same apparatus. We hypothesized that they would attend more to the character.

CONTRIBUTION

We designed our experiment to limit confounds and inconsistencies. We used two different confederates and corresponding animated characters to ensure that effects were not a result of a specific character's design. Additionally, comparison tasks were very similar, semi-structured conversations

(a) Video with AAM tracking (b) Normal cartoon (c) Damped cartoon (d) Exaggerated cartoon

Figure 1. Example frames of confederate video tracking and corresponding character being animated with different levels of facial motion. In this frame, the exaggeration/damping is most clearly shown in the mouth.

that were counterbalanced so that conversation topics were not confounded with participants' interaction partner. For both tasks, the interactive partner was an adult confederate who appeared on-screen. In one task, she was shown through video; in the other task, she was shown as a cartoon character. The confederate's cartoon character was human and was customized to her appearance. The confederate was blind to her appearance on the participant's screen. Also, we created questionnaires to determine the correspondence between subjective measures and participant behavior. Through this combination of control measures, we examined the precise effects of character appearance and motion on children's experiences. Our results indicate that children are resilient to the appearance and motion of their conversation partners, regardless of their personal preferences. Therefore, it should be possible to design conversational agents that elicit natural behavior from children.

METHOD

In this section, we explain the technique we used to animate a character with our confederate's facial motion in realtime. We also describe our equipment, study materials, participants, procedure, and measures.

Apparatus

Our goal is to create believable interactive animated characters, with motion that mimics the pacing, style, and facial gestures of humans; to that end, we opted to have two confederates "puppet" the characters. To ensure that the confederates were blind to the study condition, they were animated using a markerless, computer vision method for face tracking: active appearance models (AAMs) [13, 14, 27]. This method tracks a person's face in realtime, permitting it to be mapped onto a character's face without noticeable delay. We designed a desktop-like audiovisual telecommunications system so that research participants could interact with confederates who both appeared as themselves through video and as animated characters while using natural eye contact and speech.

Active Appearance Models

AAMs mathematically model the face shape and appearance of people and characters. To animate a character using human motion, an AAM can be customized to an individual, and a corresponding AAM can be created for the character. To create a custom AAM, images of an individual's face in different poses are manually labeled so that a mesh of 79 vertices fit to the individual's face. Once the face and appearance space have been learned from the training images, the individual's face can be tracked from new video in realtime. Animating a character requires a mapping from the individual's AAM to the character's AAM. When the individual moves his/her face, the 79 mesh vertices change position. The change in position is mapped to the character's mesh, and then the character's appearance is warped to illustrate movement (see Figure 1). We purposefully created characters that resembled the confederates so that the remapped motion would be as accurate as possible. For further information on the creation and use of AAMs, please see [13, 14, 27].

Given that each vertex changes position during movement, we followed previous procedures [5, 35] to multiply those changes in position by specific scale factors, thus exaggerating or damping the spatial movements across all features of the face. We selected damped and exaggerated scale factors based on a previous perceptual study in which the adult threshold of facial motion level sensitivity was determined [23]. We selected equally perceptible levels of damping (-20%) and exaggeration ($+25\%$). For example images of exaggerated and damped character motion, see Figure 1.

We used 2D AAMs for this research. Therefore, our characters always faced forward. We added rigid points around the tops of the characters' heads and damped the face border and nose points by 50% to ensure that the character would not warp excessively if the confederate turned her head. Body motion was not tracked, so the torsos of the characters moved rigidly with respect to a pivot located at their mouths. The confederates practiced extensively with the characters prior to the experiment to ensure that they did not generate movements that appeared unnatural or otherwise distracting when presented on the characters.

Telecommunications System

Our audiovisual telecommunications system is diagrammed in Figure 2. It was designed to maximize natural interactions: both people appear life size and can make eye contact as they would in person. Two setups were positioned in separate rooms with a control room in between. All video and audio data from each setup was relayed through the control room for video and audio processing before being presented to the other setup. Each setup consisted of a black box hous-

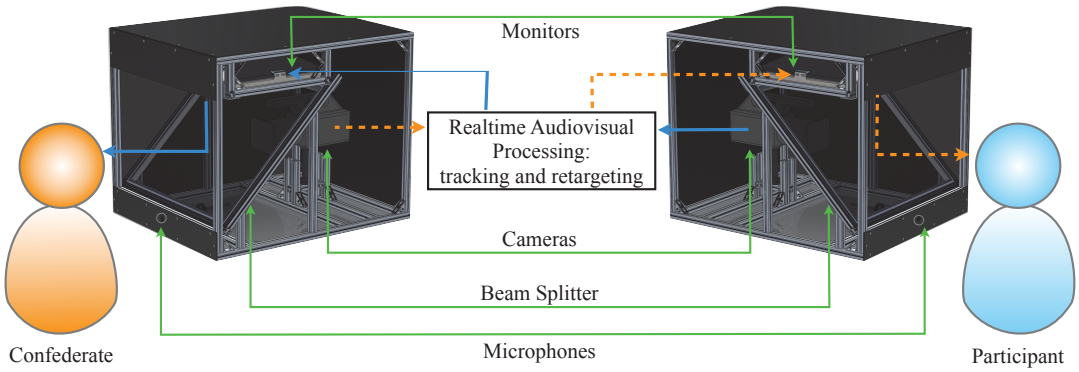

Figure 2. Diagram of the telecommunications system.

ing a monitor, camera, microphone, and beam splitter, and the box was positioned on a height-adjustable table to ensure that the camera was at eye level for the user. A speaker was also placed within the participant's setup, but headphones were used on the confederate's setup to avoid auditory feedback. We used a beam splitter made of reflective material between the user and the camera so that the camera was hidden directly in front of the user, allowing for eye contact between users, often an impossibility in audiovisual telecommunications. The monitor was placed above the beam splitter in order to project visual information directly in front of the user. A shotgun microphone was mounted below the beam splitter to capture audio.

A computer attached to each setup controlled the presentation of visual information. When animating a character, the participant's computer tracked the confederate, retargeted the motion to the animated character, and displayed the character on the participant's monitor. If the system was displaying a video feed, the computer was used to add a small amount of delay to the presentation of the video in order to replicate the delay induced by tracking and animating the confederate character. A sound mixer in the control room was used to add delay to the audio and ensure that the audio and video/animation remained in sync. Our measurements indicated that the delay inherent in our system is 100 ms for video and 166 ms for animation. Because the confederate always saw video of the participant, the delay for the confederate was 100 ms; however, because the participant saw both video and animated conditions we kept the participant's delay at 166 ms. Previous research with a similar system validated that these delays have a negligible effect on conversation between adults [34].

Storybook

In order to introduce the children to the task, we created storybooks for each of the confederates and possible task orders. Each storybook starred one of two confederates and described how she enjoyed playing pretend, sometimes as a cartoon character. She particularly enjoyed pretending to own a bakery. Depending on the order of tasks assigned to a particular participant, the storybook then described one of two tasks: designing a cake or an ice cream sundae. For the cake task, the children were told that the confederate had already created three cakes and needed the child's help to make a fourth, selecting the flavor of cake, the frosting, and a topping from a list. In the ice cream task, the children were asked to help design a fourth sundae by selecting the ice cream flavor, a sauce, and a topping from a list. After the child completed the first task, the experimenter read the second part of the story that described the other task. Participants were randomly assigned a task order and confederate.

Visual Aids

We used felt fabric to create facsimiles of all of the food components described in tasks so that the participants could see their creations at the end of each task and have a more concrete representation of the items for imagination and recall.

Questionnaires

To obtain participant feedback, we used a modified smileyometer with written labels. Smileyometers have been used frequently with children as they are understandable and reliable; however, we included written labels because they are better for older children [6, 30]. The traditional smileyometer includes a neutral midpoint; however we removed the midpoint as prior work [7] found that four response options and no neutral midpoint obtained the most reliable responses from children. Our participants were all familiar with the various types of smileys as they were used in the other experiments that participants completed on the same day. Unfortunately, eliciting truthful responses from 4- to 5-year-olds can be very difficult as they are susceptible to satisficing, and they will often select the most positive response [38]. When presenting the rating scale, the experimenter verbally stated each option while pointing to the corresponding smiley. The experimenter also verbally confirmed each of the participants' responses. After each task, the experimenter verbally asked the children four questions and offered possible answers while showing the questions and a rating scale:

1. How much did you like talking to [name] just now? (Really liked, kind of liked, didn't really like, did not like)

2. How much fun did you have talking to [name] just now? (Lots of fun, kind of fun, kind of bored, bored)

3. How nice was [name] to you just now? (Very nice, kind of nice, kind of mean, mean)

4. How much did you like talking to [name] about [topic]? (Really liked, kind of liked, didn't really like, did not like)

After both tasks, the children answered an additional three questions. Again, participants were verbally read the questions and possible responses as well as shown images of the responses. Participants were given response options in a random order:

1. Did you like talking to [name] more when she looked like a real person or when she looked like a cartoon character? (photo of confederate and cartoon character)
2. If you could speak to [name] again, would you want to see her as a real person or as a cartoon character? (photo of confederate and cartoon character)
3. Did you like talking about ice cream or cake more? (image of ice cream sundae visual aid and cake visual aid)

To measure participants' extroversion, we asked parents to answer a short, six-question survey. The six items were selected to represent the six facets of extroversion, as defined in the NEO-PI-R [15], a well-known and validated personality measure for adults. The items were selected from the M5-PS-35 [19, 32], a measure created and validated to assess preschool children's personality.

Participants

Children between the ages of four and ten years were recruited to take part in a series of short, unrelated experiments, including this study, that lasted a total of approximately 90 minutes. In total, 69 children (mean age = 7.17 years, standard deviation = 2.02 years, 36 boys and 33 girls) participated successfully in this paradigm. Five additional children were excluded from analyses due to technical or behavioral issues. See Figure 3 for participant breakdown. Participants were recruited using email lists and advertisements in local gathering places, and they were compensated for their time. The research was approved by our Institutional Review Board.

Procedure

Upon arrival at the experiment location, each child and his or her parent/guardian was met by one of the experimenter team. Parents/guardians completed the extroversion questionnaire in a separate room. Most parents remained in the separate room while their children completed the study; however, sixteen parents accompanied their children to the study room. They sat 12 feet to their children's left with the experimenter in between. Parents could not see the telecommunications screen, but they could hear the conversation. Parents of children who completed the study successfully sat silently during the study. The child was allowed to select one of many stickers to help the experimenter decorate the apparatus, in order to let him or her warm up and become accustomed to the apparatus and environment. Then, the experimenter seated the child facing the apparatus with the curtain still down and read the storybook to prepare the child to play a game of pretend with the confederate either as a video or an animated character. Upon completion of the first half of the storybook, the experimenter asked if the child was ready and then whether the confederate was ready. After hearing agreement from both, she raised the curtain so that the participant and confederate were able to see each other.

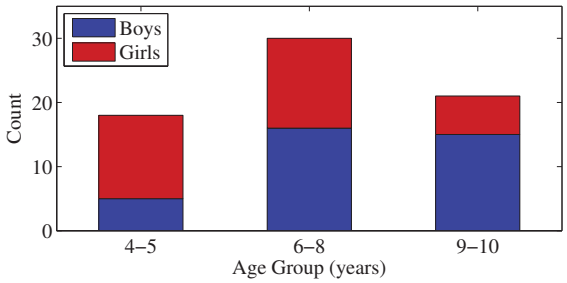

Figure 3. Participants by gender and age group.

Before starting the first task, the confederate engaged the participant in unstructured small talk, typically about the child's summer activities, age, and favorite school subjects, in order to make the child comfortable with the confederate and the apparatus. When the confederate believed that the child was comfortable, she began the first task. While the child made selections of food, the experimenter assembled the final product. When the child made his/her second selection, the confederate challenged the choice and offered her own suggestion. Upon completion of the first task, the experimenter closed the curtain so that the child could no longer see the confederate, showed the child the design, and completed the questionnaire with the child. Then, the experimenter verified that the confederate was ready to begin the next task and raised the curtain. For the second task, there was no small talk; the confederate asked the child if he or she was ready and then began the task. The confederate challenged the child on his/her second selection just as she had done in the first conversation. Again, the experimenter created the food design based on the child's selection, closed the curtain at the end of the task, showed the child the creation, and completed the questionnaire with the child. Finally, they completed the last three questions comparing both tasks and characters.

Measures

Independent variables were divided into experimental and participant variables. Experimental variables included appearance and motion manipulations. Participant variables included participant age, gender, and extroversion score. Dependent variables were split into conversation, gaze, gesture, and self-report measures.

We annotated the video recordings for child speech, gaze, and gesture using ELAN [28, 33], open-source software for annotating video and audio recordings. Each annotation has a start time, end time, and label or transcription. For each measure that involved annotated data, we had a primary annotator who annotated all of the data and a secondary annotator who annotated one third of each child's data. To evaluate interrater reliability, we calculated the percentage of aligned annotations and then calculated Cohen's Kappa (κ) for the aligned annotations. Percent alignment and κ are given below.

Experimental variables

Our main experimental variables included *confederate appearance* and *facial motion level*. Other experimental variables that we controlled in our analyses included *conversation topic*, *conversation order*, and *confederate*. The within-

subjects variables were *confederate appearance* with two levels (cartoon, video), *conversation topic* with two levels (cake, sundae), *conversation order* with two levels (first, second), and *confederate* with two levels (A, B). *Facial motion level* was a between-subjects variable with three levels (damped, unaltered, exaggerated). The experimental conditions were counterbalanced across participants.

Participant variables

Participant variables included *gender*, *age group*, and *extroversion score*. Participants were divided into three age groups (four to five years old, six to eight years old, nine to ten years old) based on theories of cognitive development [25, 37]. Parents completed a short questionnaire to assess their child's level of extroversion. The parents' responses were averaged across the six questionnaire items to create a single *extroversion score* (Cronbach's $\alpha = .6712$).

Conversation measures

Conversation measures included *conversation length*, *number of utterances*, *number of words*, and *confederate influence*. The primary annotator transcribed all participants' speech, and the secondary annotator marked the times when participants spoke. Pauses between utterances had to be at least 500 ms long. The annotators had 82% alignment on their annotations; we did not calculate Cohen's κ because the secondary annotator did not transcribe speech. From the transcriptions, we calculated the *number of utterances* and *number of words* that each participant used. *Conversation length* was measured between when the experimenter asked the confederate if she was ready and when the curtain covered the screen. Because the first conversation included unstructured small talk, the start of the first task occurred when the confederate asked if the child was ready to begin the task. For *confederate influence*, we looked at when participants changed their selection based on the confederate's challenge. If the participant changed his/her original selection to the confederate's selection, we scored the confederate as influential. If the participant stuck with his/her original selection, we scored the confederate as not influential.

Gaze measures

Gaze measures included the percentage of time participants spent looking at the screen, *percent on-screen*, and the average length of each on-screen gaze segment, *average gaze length*. While *percent on-screen* gives a rough estimate of how much of the conversation participants watched the screen, the *average gaze length* gives an estimate of how long participants sustained their gaze at the screen. Annotators marked when participants were looking on- and off-screen. The annotators had 86% alignment of annotations, and they agreed on the annotation labels with perfect reliability (Cohen's $\kappa = 1$).

Gesture measures

Gesture measures included the *total number of gestures* and the *number of nods*, *shakes*, and *shrugs* each participant used. Gestures were only considered if they were communicative. For example, if a participant moved his/her head up-and-down to indicate "yes," the motion was counted as a nod; however, if the participant was simply bouncing in his/her chair, the head movement was not considered a nod. Annotators marked and labeled the beginnings and ends of gestures. The two annotators had 83% alignment, and they agreed on the labels with perfect reliability (Cohen's $\kappa = 1$).

Self-report measures

Self-report measures included *conversation score*, *appearance preference*, and *topic preference*. Participants were asked four questions after each task and another three questions after completion of both tasks, as described in the Questionnaires section of this paper. The first four questions asked participants to rate their conversations. The last three questions asked participants to compare their conversations. The ratings from the first four questions were combined to create a *conversation score* with good reliability (Cronbach's $\alpha = .6879$). The two questions, asking participants to select between the cartoon and video confederate, were combined to create the measure of *appearance preference* with good reliability (Cronbach's $\alpha = .7178$). The last question was used as an indication of participants' *topic preference*.

RESULTS

We conducted several ANalyses Of VAriance (ANOVAs) to investigate possible effects from the independent variables. Due to a lack of variability in *extroversion scores* across participants and the fact that no relationship was found between *extroversion score* and the dependent measures, we excluded *extroversion score* from further analyses. We also excluded *number of shakes* and *shrugs* from our analyses, as the median number of times these gestures occurred during conversation were 1 and 0, respectively. The *number of nods* and the *total number of gestures* were kept in the analyses. Although participants had strong preferences for appearance and topic, they did not alter their behavior to reflect these preferences. Interestingly, confederate appearance only affected the number of words children used.

Preliminary analysis

To determine which control variables to include in our main analysis, we first conducted a repeated measures ANOVA with *conversation topic* and *conversation order* as within-subjects variables and *confederate* as a between-subjects variable. Only *conversation length* was affected by these variables. Confederate B had significantly longer conversations than Confederate A, $F(1, 65) = 43.91$, $p < .0001$. Because the conversations were semi-structured and there were no significant effects of confederate on the number of utterances or words, the difference in conversation length was likely due to a difference in the confederates' rates of speech. Conversation length was also affected by conversation order, $F(1, 65) = 14.51$, $p = .0003$; however, conversation order and confederate did not create a significant interaction.

To understand the difference in conversation length between the first and second conversation, we look at the significant interaction of conversation order and topic, $F(1, 65) = 5.35$, $p = .0239$. Participants' second conversation was only longer than their first conversation if the second topic was ice cream sundae. From our analysis of self-report measures, we know that participants preferred the sundae topic to cake. We also

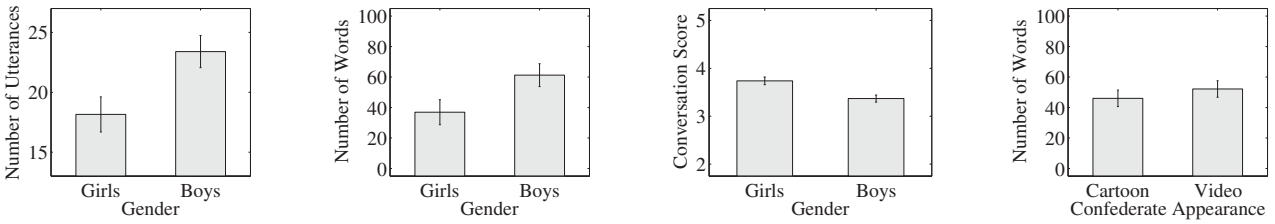

Figure 4. Main effects of gender on (a) number of utterances and (b) words, (c) conversation score, and of appearance on (d) number of words. All effects significant at $\alpha = .05$.

know that participants did not actually speak more about sundaes because we found no significant effect of topic on conversation length, nor did we find any differences in number of utterances or number of words. These results suggest that participants spent more time thinking about their responses if the second conversation was on their preferred topic. Participants were familiar with the conversation structure and the confederate by the second conversation so they may have felt less pressure to make their selections quickly. Because the control variables did not affect any of the behavioral measures except for conversation length, we did not include them in our main analysis.

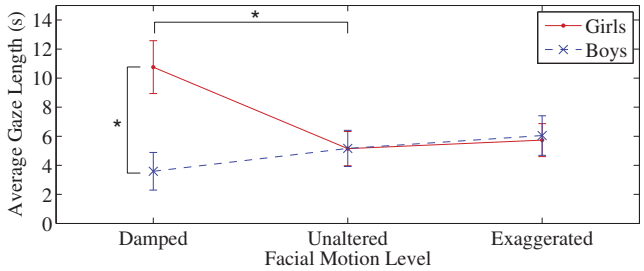

Figure 5. Interaction of facial motion level and gender on average gaze length. The * indicates significance at $\alpha = .05$.

Main analysis

We were most interested in how the experimental variables of *confederate appearance* and *facial motion level* would affect the dependent measures. Because many developmental changes occur between ages 4 and 10, we included the participant variables of *age group* and *gender* in our main analysis. We conducted a repeated measures ANOVA with *age group*, *gender*, and *facial motion level* as between-subjects variables and *confederate appearance* as the within-subjects variable. To analyze the data on appearance and topic preferences, we conducted a three-way ANOVA with age group, gender, and facial motion level as between-subjects variables.

Effects on conversation measures

We found that male participants had significantly more utterances than female participants, $F(1, 55) = 6.27$, $p = .0153$. Similarly, male participants also used more words than female participants, $F(1, 55) = 4.30$, $p = .0428$. On average, male participants used 25 more words than female participants. Although we had a disproportionate number of older boys, there were not significant interactions between gender and age for utterances or words, $F(2, 55) = 1.81$, $p = .1731$ and $F(2, 55) = .81$, $p = .4480$. Participants used significantly more words when speaking to the confederates by video than by cartoon character, $F(1, 63) = 5.50$, $p = .0222$, but the difference was six words, on average (see Figure 4). We found no significant effects of our experimental or participant variables on confederate influence.

Effects on gaze measures

We found a significant interaction between gender and facial motion levels with participants' average gaze length, $F(2, 55) = 3.89$, $p = .0264$ (see Figure 4). Female participants watched the damped cartoon character for longer periods of time than male participants, $F(1, 55) = 8.67$, $p = .0047$. Female participants also watched the damped cartoon character for longer periods of time than the unaltered and exaggerated cartoon characters, $F(1, 55) = 7.61$, $p = .0079$, and $F(1, 55) = 5.49$, $p = .0228$, respectively (see Figure 5). We found no significant effects on percent of on-screen gaze.

Effect on self-report measures

We found a significant effect of gender on conversation score, with female participants rating the conversations higher than male participants, $F(1, 55) = 10.43$, $p = .0021$. We also found a significant interaction of age group and gender with conversation score, $F(2, 55) = 3.19$, $p = .0488$. The data illustrates that males and females converged on their conversation scores as age group increased, and that the largest difference between scores occurred between the 4- to 5-year-old males and females. The 4- to 5-year-old males rated their conversations significantly lower than the 4- to 5-year-old females, $F(1, 55) = 14.66$, $p = .0003$ (see Figure 6).

We found no significant effects or interactions of age group, gender, and facial motion level on preferences. We did notice strong preferences across participants. We ran several Chi-squared tests to investigate these preferences. Participants who selected "both" instead of selecting a single topic or appearance were excluded from the analyses. Participants preferred the confederate in video to cartoon, $\chi^2(1, N = 54) = 21.41$, $p < .0001$. Participants also preferred speaking about sundaes to cakes, $\chi^2(1, N = 52) = 11.08$, $p = .0009$. If we include all participants and use "both" as a third category we still find a strong preference for video and sundae, $\chi^2(2, N = 69) = 29.30$, $p < .0001$ and $\chi^2(2, N = 69) = 14.87$, $p = .0006$, respectively.

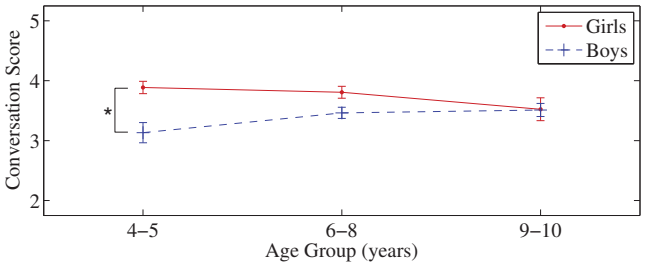

Figure 6. Interaction of age and gender on conversation score. The * indicates significance at $\alpha = .05$.

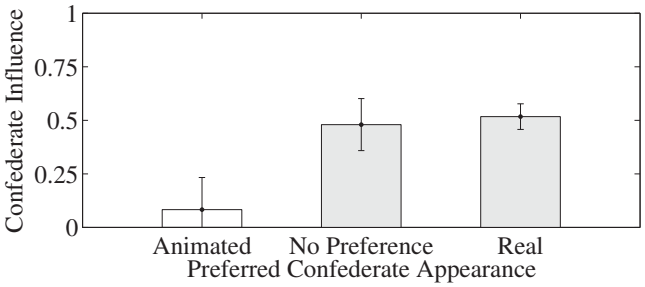

Figure 7. Main effect of participants' preference for confederate appearance on confederate influence.

Post-hoc analysis

Because participants had strong preferences for confederate appearance, we conducted a post-hoc analysis to ascertain whether or not participants' preferences influenced their behavior. We ran a repeated measures ANOVA with *age group*, *gender*, and *participant preference* as between-subjects variables and *confederate appearance* as a within-subjects variable. We found the same significant main effects of gender on number of utterances, number of words, and conversation score. We discovered a significant main effect of participant preference on confederate influence, $F(2, 53) = 3.39$, $p = .0413$ (see Figure 7). Participants who preferred the cartoon version of the confederate were less likely to be influenced by the confederate compared to the participants who had no preference or preferred the video version of the confederate, $F(1, 53) = 5.71$, $p = .0205$.

We also found a significant interaction between confederate appearance and participant preference with conversation score, $F(2, 63) = 3.69$, $p = .0305$ (see Figure 8). Participants who preferred the confederate in video rated their conversations with the animated confederate significantly lower than their conversations with the confederate in video, $F(1, 63) = 7.41$, $p = .0084$. In contrast, participants who preferred the animated confederate did not score their conversations significantly differently, $F(1, 63) = 2.22$, $p = .1409$, but this may be a reflection of the low number of participants (10) who preferred the animated confederate.

DISCUSSION

In our study, child participants conversed with an adult confederate who appeared as herself through video in one conversation and as a cartoon character in another conversation. Participants were aware that they spoke to the same confederate in both tasks even though she appeared different. We observed that some children were surprised that the cartoon character could see and respond to them. We also observed some children who were surprised by the confederate in the video condition, and some children who were not surprised by either condition. Statistically, the children did not converse, gaze, or gesture differently when speaking to the confederate through video and cartoon. The one exception was that participants used a few more words when in the video condition. Although our participants behaved similarly across conditions, they had a strong preference for the video condition.

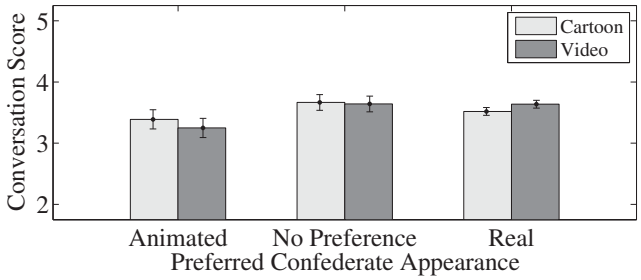

Figure 8. Significant interaction of participant preference and confederate appearance with conversation score.

We also altered the facial motion level of the cartoon character to see if it would influence participants' behavior. We found that facial motion level affected girls differently than boys, such that they had longer periods of on-screen gaze when the cartoon character's motion was damped. It is possible that the girls were more engaged with the task and therefore more attentive when the motion was damped because damped facial motion can be harder to understand [10, 22]. No other effects of facial motion level were found.

We found several effects of gender on participant behavior. Although female participants rated the tasks higher, male participants spoke more, both in number of utterances and words. The youngest girls also rated the tasks the highest, supporting the idea that they were the most engaged with the task.

Unlike prior studies [4, 29] that compared children's behaviors when conversing with an animated character to a person, our animated character was controlled in real time by the same person that participants spoke to through video. Prior studies used intelligent agents or Wizard of Oz techniques to control their animated characters, and comparison conversations were face-to-face. We limited confounds by designing our tasks to be as similar as possible. Participants had two conversations with the same confederate, engaged in two similar tasks (designing cakes and ice cream sundaes), and used the same apparatus to converse. The cartoon characters were even designed to have similar appearances to their respective confederates.

LIMITATIONS AND FUTURE WORK

There are some limitations to this work that we plan to address in the future. Our conversations were short and may not have allowed sufficient time for participants to adjust their

behaviors. Different types of tasks, such as ones that require more trust and disclosure, may also cause children to alter their behavior based on their preferences. Future work should examine how longer and/or different types of tasks (e.g., storytelling, problem solving) could elicit different behaviors based on children's preferences.

We used adult female confederates and similar-looking animated characters in our study. Future work could explore the effects of confederate gender, age, and animated character appearance. Many of the animated characters children interact with are other children or non-human (e.g. Dora from *Dora the Explorer* and Daniel from *Daniel Tiger's Neighborhood*); therefore, a comparison of children's behaviors during interactions with adult, child, and non-human animated characters would also be interesting and informative. Our characters were rendered in a more realistic-looking style than many popular cartoon characters, and children may prefer more familiar and simplistic rendering styles. An examination of how children react to a character with different rendering styles would help answer this question.

Our cartoon characters were not perfectly realistic in their motion or in their appearances. We used 2D AAMs and 2D characters, which limited how our confederates and the characters could move. Specifically, 2D AAMs cannot track faces properly when certain features are obfuscated, such as when an eye disappears from view due to a head turn. Similarly, the characters were incapable of head rotation; therefore, they could not nod or shake their heads. Due to these limitations, we requested that our confederates limit their head motion. Although our confederates kept their heads mostly still, they still made some small nods and shakes, which are naturally occurring movements. The lack of these small movements in the cartoon condition may have influenced participant preference for the video condition. In the future, different tracking and animation techniques could be tested to verify our findings.

Our findings suggest that operators/animators could adjust a character/avatar's motion to increase a child's attention and to suit a child's facial processing ability. In the future, we intend to investigate how we can customize character/avatar appearance and motion to improve conversations with children who have facial processing deficits and social and communicative disorders, like those on the autism spectrum. Because animated characters created for children are often child-like and non-human, we believe future work should investigate the use of different types of animated characters. Along similar lines, animated characters are used for more than just short conversational tasks; therefore, exploring whether children might behave differently with real people and characters during other types of tasks would be useful.

CONCLUSION

We ran an experiment in which children conversed with an adult partner who appeared as herself and as a cartoon avatar via videoconference. We found that despite having strong preferences for confederate appearance, the children behaved similarly between conditions. In our study, we also examined how an animated character's facial motion level might affect children. According to Cassell and Tartaro [11, 12], embodied conversational agents in social settings should strive to elicit behaviors from users that are indistinguishable from the behaviors users would exhibit when with other real people. Our results indicate that, with child users, the goal of eliciting natural behaviors should be possible without needing perfectly realistic-looking human characters and head motion.

ACKNOWLEDGMENTS

We thank our participants and their families. We are grateful to Jill Lehman, Rachel Browne, Brooke Kelly, Randall Hall, Emily Jensen, Tomas Simon Kreuz, Peter Carr, Iain Matthews, Mo Mahler, Jimmy Krahe, Cole Heiner, and Melanie Danver. This work was funded by Disney Research and award R03HD068816 from the Eunice Kennedy Shriver Institute of Child Health and Human Development of the National Institutes of Health.

REFERENCES

1. Andrist, S., Pejsa, T., Mutlu, B., and Gleicher, M. Designing effective gaze mechanisms for virtual agents. In *Proc. CHI 2012*, ACM Press (2012), 705–714.

2. Baldwin, D., and Tomasello, M. Word learning: A window on early pragmatic understanding. In *Proc. Child Language Research Forum*. Chicago, IL, 2001, 3–23.

3. Bartneck, C., and Reichenbach, J. Subtle emotional expressions of synthetic characters. *International Journal of Human-Computer Studies 62*, 2 (2005), 179–192.

4. Black, M., Chang, J., Chang, J., and Narayanan, S. Comparison of child-human and child-computer interactions based on manual annotations. In *Proc. Workshop on Child, Computer and Interaction 2009* (2009).

5. Boker, S. M., Cohn, J. F., Theobald, B.-J., Matthews, I., Brick, T. R., and Spies, J. R. Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars. *Phil. Trans. R. Soc. B 364*, 1535 (2009), 3485–3495.

6. Borgers, N., de Leeuw, E., and Hox, J. Children as respondents in survey research: Cognitive development and response quality 1. *Bulletin de Méthodologie Sociologique 66*, 1 (2000), 60–75.

7. Borgers, N., Hox, J., and Sikkel, D. Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity 38*, 1 (2004), 17–33.

8. Burleson, W., and Picard, R. W. Gender-specific approaches to developing emotionally intelligent learning companions. *IEEE Intelligent Systems 22*, 4 (2007), 62–69.

9. Calder, A. J., Rowland, D., Young, A. W., Nimmo-Smith, I., Keane, J., and Perrett, D. I. Caricaturing facial expressions. *Cognition 76*, 2 (2000), 105–146.
10. Calder, A. J., Young, A. W., Rowland, D., and Perrett, D. I. Computer-enhanced emotion in facial expressions. *Phil. Trans. R. Soc. B 264* (1997), 919–925.
11. Cassell, J. Embodied conversational agents: representation and intelligence in user interfaces. *AI Magazine 22*, 4 (2001), 67.
12. Cassell, J., and Tartaro, A. Intersubjectivity in human-agent interaction. *Interaction Studies 8*, 3 (2007), 391–410.
13. Cootes, T. F., Edwards, G. J., and Taylor, C. J. Active appearance models. *IEEE Trans. on Pattern Analysis and Machine Intelligence 23*, 6 (2001), 681–685.
14. Cootes, T. F., Wheeler, G., Walker, K., and Taylor, C. J. View-based active appearance models. *Image and Vision Computing 20*, 9-10 (2002), 657 – 664.
15. Costa, P. T., and MacCrae, R. R. *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual*. Psychological Assessment Resources, 1992.
16. Coulston, R., Oviatt, S., and Darves, C. Amplitude convergence in children's conversational speech with animated personas. In *Proc. International Conference on Spoken Language Processing*, vol. 4 (2002), 2689–2692.
17. Darves, C., Oviatt, S., and Coulston, R. Adaptation of users' spoken dialogue patterns in a conversational interface. In *Proc. International Conference on Spoken Language Processing*, vol. 1 (2002), 561–564.
18. de Melo, C. M., Carnevale, P., and Gratch, J. The impact of emotion displays in embodied agents on emergence of cooperation with people. *Presence 20*, 5 (2011), 449–465.
19. Grist, C. L., Socha, A., and McCord, D. M. The M5-PS-35: A five-factor personality questionnaire for preschool children. *Journal of Personality Assessment 94*, 3 (2012), 287–295.
20. Guadagno, R. E., Swinth, K. R., and Blascovich, J. Social evaluations of embodied agents and avatars. *Computers in Human Behavior 27*, 6 (2011), 2380–2385.
21. Hayes, D. *Anytime Playdate: Inside the Preschool Entertainment Boom, or How Television Became My Baby's Best Friend*. Free Press, New York, NY, 2008.
22. Hess, U., Blairy, S., and Kleck, R. E. The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior 21*, 4 (1997), 241–257.
23. Hyde, J., Carter, E. J., Kiesler, S., and Hodgins, J. K. Perceptual effects of damped and exaggerated facial motion in animated characters. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition 2013* (2013).
24. Kirkorian, H. L., Wartella, E. A., and Anderson, D. R. Media and young children's learning. *Children and Electronic Media 18*, 1 (2008), 39–61.
25. Knowles, A. D., and Nixon, M. C. Children's comprehension of expressive states depicted in a television cartoon. *Australian Journal of Psychology 41*, 1 (1989), 17–24.
26. Linebarger, D. L., and Walker, D. Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist 46*, X (2004), 1–21.
27. Matthews, I., and Baker, S. Active appearance models revisited. *International Journal of Computer Vision 60* (2004), 135–164.
28. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. ELAN. http://tla.mpi.nl/tools/tla-tools/elan.
29. Oviatt, S. Talking to thimble jellies: Children's conversational speech with animated characters. In *Proc. INTERSPEECH 2000* (2000), 877–880.
30. Read, J. C., and MacFarlane, S. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proc. IDC 2006*, ACM Press (2006), 81–88.
31. Rideout, V. J., Vandewater, E. A., and Wartella, E. A. Zero to six: Electronic media in the lives of infants, toddlers and preschoolers. Tech. rep., The Henry J. Kaiser Family Foundation, 2003.
32. Scheck, A. M., L. Alvin Malesky, J., Grist, C. L., and McCord, D. M. Personality in preschool children: Preliminary psychometrics of the M5-PS questionnaire. *American Journal of Psychological Research 6*, 1 (2010), 134–156.
33. Sloetjes, H., and Wittenburg, P. Annotation by category: ELAN and ISO DCR. In *Proc. LREC* (2008).
34. Tam, J., Carter, E., Kiesler, S., and Hodgins, J. Video increases the perception of naturalness during remote interactions with latency. In *Proc. CHI EA 2012*, ACM Press (2012), 2045–2050.
35. Theobald, B.-J., Matthews, I., Mangini, M., Spies, J. R., Brick, T. R., Cohn, J. F., and Boker, S. M. Mapping and manipulating facial expression. *Language and Speech 52* (2009), 369–386.
36. Troseth, G. L., Saylor, M. M., and Archer, A. H. Young children's use of video as a source of socially relevant information. *Child Development 77*, 3 (2006), 786–799.
37. Widen, S. C., and Russell, J. A. Children acquire emotion categories gradually. *Cognitive Development 23*, 2 (2008), 291–312.
38. Zaman, B., Abeele, V. V., and Grooff, D. D. Measuring product liking in preschool children: An evaluation of the Smileyometer and this or that methods. *International Journal of Child-Computer Interaction 1*, 1 (2013), 61–70.