J Autism Dev Disord (2014) 44:2475–2485
DOI 10.1007/s10803-014-2116-8

ORIGINAL PAPER

Are Children with Autism More Responsive to Animated Characters? A Study of Interactions with Humans and Human-Controlled Avatars

Elizabeth J. Carter · Diane L. Williams · Jessica K. Hodgins · Jill F. Lehman

Published online: 25 May 2014

Abstract Few direct comparisons have been made between the responsiveness of children with autism to computer-generated or animated characters and their responsiveness to humans. Twelve 4- to 8-year-old children with autism interacted with a human therapist; a human-controlled, interactive avatar in a theme park; a human actor speaking like the avatar; and cartoon characters who sought social responses. We found superior gestural and verbal responses to the therapist; intermediate response levels to the avatar and the actor; and poorest responses to the cartoon characters, although attention was equivalent across conditions. These results suggest that even avatars that provide live, responsive interactions are not superior to human therapists in eliciting verbal and non-verbal communication from children with autism in this age range.

Electronic supplementary material The online version of this article (doi:10.1007/s10803-014-2116-8) contains supplementary material, which is available to authorized users.

E. J. Carter (✉) · J. K. Hodgins
Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: liz.carter@aya.yale.edu

D. L. Williams
Speech-Language Pathology, Rangos School of Health Sciences, Duquesne University, 600 Forbes Avenue, Pittsburgh, PA 15282, USA

J. K. Hodgins · J. F. Lehman (✉)
Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: jill@kidaccess.com

Keywords Autism · Animated characters · Computer-assisted technology · Computer-based interactions · Communication · Avatars

Introduction

We were given the unique opportunity to assess the responsiveness of children with autism to *Turtle Talk with Crush*, a 12-min attraction that uses an interactive avatar at theme parks. In the presentation, an animated character named Crush [the surfer sea turtle from the movie *Finding Nemo* (Lasseter and Stanton 2003)] converses with audience members in a humorous way about what it is like to be a sea turtle versus a human. Throughout the experience, a variety of behaviors are elicited from the children in the audience: asking questions individually, answering questions both individually and as a group, pointing, raising their hands, taking turns, and imitating speech. Parents and bloggers have reported that their children with autism spectrum disorder (ASD) perform better in the attraction than in regular daily life (e.g., Jones 2008). Important elements of this environment were thought to be the non-human (or alternatively, cartoon) nature of the communication partner coupled with the interactive, reciprocal format.

The parent reports of their children's positive reactions to the interactive avatar Crush are consistent with assumptions that have been previously discussed in the ASD literature. Individuals with ASD are thought to respond better to computer-assisted technology (CAT) because it produces less anxiety than a typical human-to-human interaction (Hailpern 2008); it is more motivating (Moore and Calvert 2000; Tjus et al. 2001); it reduces the number of off-task behaviors (Chen and Bernard-Opitz

1993; Plienis and Romanczyk 1985); and it can focus the child's attention on the salient cues (Moore and Calvert 2000).

Based on assumptions about the positive value of CAT for children with ASD, interventions that use this technology have proliferated dramatically, with a recent review reporting that articles on the topic in peer-reviewed journals increased from 5 per year in the late 1990s to 35 to 40 per year from 2008 to 2011 (Ploog et al. 2013). These computer-based interventions (CBIs) have largely targeted behaviors in the domains of social skills and language (Ploog et al. 2013). Social skills training CBIs using simulated interpersonal scenarios have been reported to be successful in increasing rates of appropriate social behaviors in children with ASD (Bernard-Opitz et al. 2001; Beaumont and Sofronoff 2008; Milne et al. 2010). CBIs incorporating avatars also have been reported to improve language comprehension in this population (Bosseler and Massaro 2003; Williams et al. 2004; Grynszpan et al. 2008). These results suggest that CBIs may be useful for developing social and language skills in children with ASD.

The positive effects of CBIs appear to be consistent with assumptions about individuals with ASD preferring to interact with non-humans. However, despite the demonstrated benefits of CAT for teaching communication skills to children with ASD, only a small number of studies have directly evaluated whether instruction that employs CAT is superior to instruction implemented by a human teacher (Ramdoss et al. 2011). For example, in an extensive review of 45 papers that investigated the use of various CATs for children with ASD, Ploog et al. (2013) found that only eight included any sort of human-based intervention as a comparison. The successfulness of the CBI has generally been attributed to the characteristics of that intervention, specifically the use of CAT, without a direct comparison to traditional face-to-face therapy. As technology improves, it is likely that there will be a shift in using CAT from its current role of supplementing the work of a therapist into a role where it provides additional therapy hours with similar efficacy as an in-person therapist. Thus, direct comparisons will become more important.

Some studies, unrelated to the investigation of CBIs, have addressed whether aspects of behavior vary when children with ASD interact with a human versus a virtual or robotic partner. For example, one experiment compared attention by children and adolescents with ASD to four types of electronic screen media, including an animated video, a pre-recorded video of themselves, a video of a familiar individual using an immersive virtual reality (VR) system, or their own use of the VR system (Mineo et al. 2009). The children exhibited the maximum amount of visual attention when they were using the VR system themselves, but the most vocalizations when they were watching a familiar person using the VR system, although the effects were not particularly large. The authors did note an unexpectedly high level of visual attention to the animated video (Mineo et al. 2009). In the language domain, Tartaro and Cassell (2008) examined the use of virtual peers versus typically developing peers for building collaborative narratives with 7- to 11-year-old children with ASD. They found that the children with ASD had more contingent verbalizations with the virtual peer than with the human peer (Tartaro and Cassell 2008). However, they did not examine whether that was because the virtual peer was controlled by a human adult rather than a child (even though the physical appearance was that of a same-age peer) or because it was more predictable than the real, child peer.

Comparisons of the responsiveness of children with ASD to language presented in varying human and CAT formats will begin to address some of the unanswered questions and perhaps give some guidance as to the important elements for the design of language-learning interventions for children with ASD. The opportunity to observe children during interactions with the avatar in *Turtle Talk with Crush* provided us with an opportunity to explore some of the unanswered questions about the responsiveness of children with autism in reciprocal interactions while varying the important elements of humanness and contingency of response. The purpose of this study was to measure the nonverbal and verbal behavior of children with autism in four different conditions (1) during interaction with a highly experienced human therapist, (2) while in the audience watching the human-controlled Crush avatar (run and voiced live by highly trained professional actors), (3) while interacting with a former theme park actor who spoke in a Crush-like format and who had experience interacting with children but with no specific training in interacting with children with autism, and (4) while watching children's television programs that sought to elicit conversational and imitative behaviors. These four conditions allowed us to make comparisons to explore the effects of different variables. We could compare the children's performance when interacting with Crush and with the actor to determine what behavioral effects, if any, were due to the use of an exciting cartoon avatar rather than a human, independent of the type of speech pattern being used. Also, we could compare a trained therapist to other humans who were not specifically trained to work with children with ASD, both with and without a computer avatar. Finally, we could compare the children's performance under these conditions to that of non-human-controlled animation that explicitly invited interaction from the children. All of the conditions were tailored to be as similar as possible to the interaction possibilities provided during *Turtle Talk with Crush*, the element of the experiment that we could not alter, and they were designed to be appropriate for the age and ability range of the participants.

Method

Participants

Twelve children with autism (two girls) between the ages of 4.2 and 8.2 years old ($M = 5.9$, $SD = 1.2$) participated in this research study. Participants who already had a clinical diagnosis of autism were recruited through an online call for participation associated with the Autism Society of America's (ASA) 2011 National Conference and Exposition in Orlando, Florida, home of the theme park, as well as through email requests supported by the Autism Society of Greater Orlando. Only children who had not been to the theme park for at least 1 year were eligible for participation to reduce the familiarity of the children with the *Turtle Talk with Crush* attraction.

Inclusion criteria included scores at or above the cut-off range for autistic disorder on the *Autism Diagnostic Interview—Revised* (ADI-R; Lord et al. 1994), which was administered over the telephone with the child's guardian prior to full enrollment in the study. The prior administration of the ADI-R occurred to verify that the children had a clinical history consistent with a diagnosis of autism prior to inviting them to participate in the study in Orlando, FL. In addition, to ascertain that the children met research criteria for autism based on current behavioral presentation, the *Autism Diagnostic Observation Schedule* (ADOS; Lord et al. 2000), was administered at an initial onsite session at the hotel where the ASA Conference was being held. In addition, the children behaviorally presented with autism based on expert clinical opinion. Exclusionary criteria were a history of or current occurrence of tuberous sclerosis, Fragile X syndrome, fetal cytomegalovirus infection, prematurity, seizure disorders, birth asphyxia, or head injury.

To characterize the child's level of spoken language, a language sample was collected during standard sections of the ADOS protocol using the materials for that section. Sections of the ADOS that were used included the construction task, make-believe play, conversation, the demonstration task, description of a picture, birthday party, and bubble play. Criteria for the samples were that they include a minimum of 50 complete and intelligible utterances and a minimum of 100 words. The language samples were video-recorded and transcribed by research assistants trained to reliability using the *Systematic Analysis of Language Transcripts* (SALT; Miller and Chapman 2010). The samples ranged in length from 57 to 106 complete and intelligible utterances (group $M = 70$; $SD = 14.29$) and 115–442 words ($M = 261.50$; $SD = 109.73$); these ranges were consistent with differences in the level of generative language ability of the children. Mean length of utterance in morphemes (MLUm) was obtained for each child using the automated analysis program provided in SALT and ranged from 1.91 to 5.04 morphemes ($M = 3.22$; $SD = 1.06$). See Table 1 for participant characteristics. The ADI, ADOS, and language scores indicate a range of function levels within this group.

The research was approved by the institutional review board at our primary university. Participants were compensated for both the phone interviews and their in-person participation in the research.

Interaction Conditions

Therapist

This condition took place in a medium-sized room at the hotel near the registration area for the conference. The interaction was video-recorded using two cameras on tripods placed in opposite corners of the room. The parent remained in the room, sitting off to the side. In addition to the therapist, another experimenter was always present to ensure the interactions were recorded properly.

The therapist, a female, was a licensed, certified speech-language pathologist with a doctoral degree and more than 30 years of clinical experience with young children with a variety of developmental language problems, including ASD. Before administering the ADOS, the therapist engaged in a semi-structured interaction with each child following a script that was developed to contain similar interactive communication opportunities with the same speech acts and gestural communication as the ones provided during the *Turtle Talk with Crush* experience (e.g., verbal imitation, social gestures, requests for factual information, etc.). The therapist and child were seated on the floor facing each other. None of the children were known to the therapist before participating in this study. All of the children had met the therapist for the first time just minutes before the interaction was initiated. The interaction took place at the beginning of the session in order to ensure that familiarity with the therapist did not affect behaviors. At the initiation of the interaction, the therapist greeted the child, introduced herself, and asked his or her name and age. Then, she showed the child a toy animal and asked what the animal would like to eat, presenting food options and prompting the child to feed the toy animal. She asked more questions about the basic properties of the animal before saying that the animal needed a nap and asking the child what to do. Then, she prompted imitation behaviors for saying goodbye to the animal, waving, and putting it away. The child could also

Table 1 Participant characteristics

Participant	Age	Gender	ADOS communication	ADOS social	ADOS total	MLU	MLU group	One-on-one with Crush?
1	4.23	M	9	14	23	1.91	Low	No
2	4.7	M	4	10	14	4.37	High	Yes
3	4.78	M	7	11	18	3.56	High	Yes
4	5.09	M	4	9	13	4.27	High	Yes
5	5.18	M	5	10	15	5.04	High	Yes
6	5.25	M	8	11	19	2.51	Low	Supported failure
7	5.67	F	10	14	24	2.09	Low	No
8	6.41	F	10	14	24	1.99	Low	No
9	6.71	M	6	7	13	3.2	High	Yes
10	7.05	M	8	11	19	2.3	Low	No
11	7.24	M	7	11	18	4.02	High	Yes
12	8.19	M	4	9	13	3.42	High	Raised hand

have a snack item if he or she desired. They finished the interaction with a "high five." We created the script to ensure a similar number and types of bids for interaction in each of the conditions, using the Crush script (over which we had no control) as the point of comparison. The exchange lasted between 1.43 and 2.43 min, as measured from the time when the therapist started the script to when she completed the script ($M = 2.07$, $SD = .30$). The script is available in Supplementary Materials 1.

Cartoon Videos

After the ADOS administration was completed, each child sat alone in front of a big-screen television and watched segments of two children's television programs, *Blue's Clues* and *Dora the Explorer,* that have been reported to be related to greater vocabularies and higher expressive language scores in toddlers with typical development (Linebarger and Walker 2004). The clips included short, complete scenes that specifically sought to elicit imitative speech, gestures, and answers to questions. In *Blue's Clues,* the host asked questions (e.g., "What's missing?") that were followed by a pause in which children in the audience could answer the question. After the brief pause, prerecorded children's voices would provide the correct response (e.g., "Sprinkles!" when one cupcake on a plate lacked them), which the main character would then confirm was correct. *Dora the Explorer* segments included questions that could prompt both verbal and nonverbal responses (e.g., "Do you see a blue butterfly with yellow polka dots?"). After a pause, an arrow would appear and highlight the correct answer, and a character would confirm which answer was correct (e.g., "There it is!"). In other segments from the show, the audience would be commanded to do something specific (e.g., "Can you say 'salta'? Say, 'Salta!'"), and either the characters would pause to allow audience members to respond or they would do it at the same time. In this show, some commands were physical in nature, such as prompting the audience member to jump. Typically developing young children have been reported to actively participate and respond to these programs (Linebarger and Walker 2004), and some of our participants did as well, although they received no instruction from the experimenters to do so. In all, each child watched 6 min of clips from the two shows that included a total of 26 bids for a verbal response and 22 bids for a gestural response. Because of the design of the programs, there was significantly more time in between bids than there is in a typical one-on-one conversation; thus, this condition was longer in duration than some other conditions. One child was excluded from this condition because she reacted so negatively to the cartoons when they were first displayed that this activity could not be completed.

Interactive Avatar (Crush)

Upon arriving at the park, families were met at the customer service booth outside of the gate and given their tickets. An experimenter guided them through the theme park to the *Turtle Talk with Crush* exhibit in The Living Seas Pavilion. Typically, children sit on the floor in the front of the room while adults sit on rows of benches behind them. Each of our participants was seated as far forward as possible on the floor with a parent accompanying them if the child desired. A large screen showing an animated underwater scene was at the front of the room. At the beginning of the session, a human host explained the rules and the scene to the audience, pointed out the emergency exits, and introduced Crush. Then, the animated turtle swam onto the screen and began to interact with the audience. In order to be optimally interactive, Crush is tele-operated by a human actor who can see and hear the audience. Crush started the interaction by engaging the entire group, eliciting responses from all of the children at once. Then, he engaged in one-on-one conversations with one child and his or her parent, during which he explained that his 7-year-old son, Squirt, would ask him questions to which he did not know the answers. Crush continued engaging either the group or individuals in conversation for several minutes, periodically requesting gestures and specific statements. He used surfer-style language (e.g., "Dude!") and an Australian accent. When possible, he would specifically call on the research participant to converse if the child raised his or her hand. An experimenter ensured that the Crush actor knew which children were research participants, and only one research participant was present per performance. After approximately 10 min, another character, a fish named Dory, appeared in the scene and engaged the audience in conversations with questions, answers, and imitative statements. Dory was not tele-operated, running instead using a prerecorded script. However, she was still interactive with the audience and with Crush, eliciting similar behaviors from the group. Shortly after the interactions with Dory, Crush said good-bye to the crowd. In all, the experience lasted 12–15 min. Afterwards, an experimenter escorted each research participant and his or her family to a room in the back of the building for the next interaction condition. Two professional cameramen recorded the participants using HD video cameras.

We used a tele-operated character on the grounds that it is a best case scenario (given the current technology) for an animated character that is trying to elicit a social response. For example, Cassell and Tartaro (2007) proposed that a benchmark for an interactive conversational agent—useful for those that will eventually function autonomously—is the similarity of the responses it can elicit to those elicited by a human. Thus, the use of this character was the most level playing field that we could find given the current state of the art, and we elected to use it to reduce the maximum number of possible confounds between the humans and the animated character. Any differences should therefore be attributable to the character alone, not the use of an autonomous, not-fully-interactive system.

Human Actor

For the final condition, each participant met an actor in a separate room in The Living Seas Pavilion. The actor was an adult female who had previous experience working at the theme park and interacting with young children, but no therapeutic training. Upon the arrival of each child, the actor engaged in a scripted conversation that used similar terminology and style as Crush and included opportunities for elicitation of the same speech acts and gestural communication as the Crush interaction. The similarity in style and vocabulary was created in order to determine whether the participants' responses to Crush were simply based on his slow, relaxed manner of speech or his clever and unusual sayings or style of interaction, rather than his appearance. The script is available in Supplementary Materials 2. Each interaction lasted between 2.5 and 8 min ($M = 4.3$ min, $SD = 1.5$ min), depending on the effects of individual participant's engagement and attention on the speed of conversation. Duration was measured from when the actor started the script to when she completed the script. The conversation was designed to have a similar number of bids for interactive behaviors as the Therapist and Cartoon Video conditions. One professional cameraman recorded that interaction with an HD video camera.

Procedure

The order of conditions could not be varied due to the unique nature of this experiment. We needed to confirm the child's diagnosis of autism before the child participated in the segments of the experiment in the theme park. In addition, the Therapist condition needed to occur prior to the administration of the ADOS so that it was at a point in which the child was the least familiar with the therapist. The Cartoon Videos condition occurred after the administration of the ADOS, given that the children were acclimated to the setting and were more likely to attend to the videos. The families then went to the theme park for the other two experimental conditions within 72 h of the hotel session. Upon arrival at the park, they immediately went to *Turtle Talk with Crush* to ensure that we would be able to record this condition while the children were not fatigued and were in a good mood. Finally, they participated in the session with the human actor, after which the families were free to enjoy the park independently.

After the children completed the experiment, the parents filled out a questionnaire about whether and how often each child watched *Dora the Explorer*, *Blue's Clues*, and *Finding Nemo*; how much and what type of speech or applied behavior analysis (ABA) therapies he or she received; and how well they thought their child responded to the *Turtle Talk with Crush* presentation. All of the children were familiar with the programs and movies and received speech and/or ABA therapy.

Coding

Annotators who were naïve to the purpose of the study coded the videos for Attention, Gesture, and Verbal

Table 2 Example prompts and codes

Response type	Therapist	Avatar	Actor	Cartoon
Expected gesture	Great job! High five!	Give me a high fin!	Give me a high fin!	Up, up, up! Stand up!
Expected verbal	Say, "Hello, froggy."	Everyone, try: "Dude."	Say it like this: DUUUUUUUDE!!!!	Here comes another watermelon. Yell, "Salta!"
Expected—one of verbal/gesture	How old are you?	Can you give me a high fin?	Can you talk like a turtle?	Do you see a blue butterfly with yellow polka dots?
Afforded verbal	Look what I have!	Shout it out if you know it.	Glad to meet you.	–
Afforded gesture	Frogs say ribbit. [Makes frog hop and passes frog.]	Raise your flipper if you have a question.	–	–

expectation-response pairs. One annotator was responsible for coding expectations. In the case of Attention, expectations specified to whom the child should be attending as a function of the stimuli (e.g., therapist, Crush, Dora, actor) and were created for every 30 s of interaction. Eye direction was used to determine the object of attention, and the majority of the 30-s time period had to be spent on the appropriate target in order for the participant to be counted as "attending" during that interval. For gesture and verbal behavior, expectations were considered to be *expected, expected-one of,* or *afforded*, as a function of the prompt, and extended from the end of the prompt to the beginning of the next utterance. An *expected* code was used when the prompt was an explicit request for a gesture ("Close your eyes!") or verbal response ("Say 'Dude'!"). An *expected-one of* code was used when either a gesture *or* a verbal response would be considered adequate ("Do you know what frogs eat?"). An *afforded* code was used when the prompt was a conditional request, that is, the child was given an opportunity to respond but a response was not required ("Raise your hand if you have a question"). Sample prompts and codes are available in Table 2.

Two additional annotators independently scored the children's responses as a function of the expectation using the vocabulary *yes*, *no*, and *supported*. A *yes* indicated that the expected behavior occurred; a *no* indicated that the expected behavior did not occur; and *supported* indicated that the behavior occurred but external support was provided by the parent. For Attention expectations, a *no* was given if the child's attention wandered from the focus for more time than would be expected in a neurotypical child. This coding primarily related to the child's behavior during the Crush performance, a situation in which his or her behavior could be compared to the other children in the audience. Given the more free-form nature of this interaction as compared to the one-to-one interactions in the other conditions, we wanted to ensure that the behaviors that we expected were equivalent to the behaviors of typically developing children. The comparison to typically developing audience members was only used in cases where prompts were of a questionable nature. *Supported* was used if the child's attention was focused because a parent had redirected him or her. Gesture and verbal expectations received a *yes* or *no* scoring based on the presence or absence of the response, and *supported* if a parent re-prompted ("Say goodbye") or gave physical assistance (initiating a wave by lifting the hand). Inter-rater reliability across all conditions was considered moderate for Attention ($\kappa = .42$), due to differing tolerance for the duration of shifts in attention. Inter-rater reliability for both gesture and verbal coding was very good ($\kappa = .85$ and $\kappa = .84$, respectively). Disagreements between annotators were resolved by a third independent annotator, and the majority consensus was used in subsequent analysis.

Data Analysis

For additional analysis purposes, we created a weighting system to combine gestural and verbal measures. The system differentiated between responses that were *afforded*, *expected,* or *supported* in both domains as described in the section on Coding. We weighted verbalizations more heavily than gestures because we believed that verbalizations provided a greater challenge to our participants. An *afforded* response was weighted the most; it suggested a higher level of independent responding because an opportunity to respond was implied but not explicitly requested. *Supported* responses received the least weight because the child did not respond until receiving additional cuing from the parent. *Expected* responses were considered to be an intermediate level of response because the child had an explicit opportunity to respond. The points system was as follows: three points were allotted if the child provided an afforded verbal response, two points for an afforded gesture, and .5 points for a supported response of either type to an affordance. For expected responses, two points were allotted for verbalizations, one point for gestures, and .5 if either type was supported. If more than one type of response was expected (e.g., pointing and verbalizing), the child was assigned a separate set of points for each type of

Table 3 Participant responses

Participant	Attn.—Therapist	Attn.—Avatar	Attn.—Actor	Attn.—Cartoon	Comm.—Therapist	Comm.—Avatar	Comm.—Actor	Comm.—Cartoon
1	0	0.87	0.18	0.73	0.52	0.58	0.24	0.32
2	1	0.79	1	0.93	0.67	0.16	0.53	0
3	0.83	0.91	0.8	0.67	0.71	0.59	0.56	0.21
4	1	0.62	0.43	0.93	0.63	0.44	0.35	0.24
5	1	0.75	1	0.53	0.65	0.66	0.8	0
6	0.2	0.39	0	0.81	0.22	0.03	0.09	0
7	0.25	0.45	0	0.67	0.39	0.06	0.25	0.03
8	0.5	0.68	0	n/a	0.48	0.41	0.24	n/a
9	1	0.59	0.5	0.87	0.55	0.69	0.44	0.44
10	0.2	0.19	0.29	0.47	0.3	0.03	0.32	0
11	1	0.68	0.88	0.8	0.58	0.64	0.36	0.12
12	1	0.93	1	0.6	0.69	0.54	0.74	0.06

response. Zero points were assigned whenever the child failed to respond to an afforded or expected bid for a response.

Because each interaction had differing lengths and numbers of opportunities for gestures and verbal behavior, we calculated response measures as a proportion of the number of expectations for that condition for each participant, ranging from zero to one. Repeated-measures ANOVAs with Bonferroni-corrected pairwise comparisons among conditions were performed using SPSS software to examine attention and combined gestural and verbal responses across the four conditions. We also performed a principal component analysis (PCA) to examine correspondences across behaviors and conditions.

Results

Response rates for individual participants are given in Table 3.

Attention

No statistically significant differences were found among the interaction conditions for attention $F(3) = 1.17$, $p = .34$. Percentages of time that the participants spent paying appropriate attention [mean(SD)] were .67(.40) for the therapist, .67(.26) for the cartoon videos, .65(.22) for *Turtle Talk*, and .51(.41) for the actor. No pairwise comparisons were statistically significant.

Combined Gestural and Verbal Measures

A significant main effect of condition on gesture and verbal responses was found, $F(3) = 26.24$, $p < .0005$. The highest proportion of appropriate responses was to the therapist [.42(.13)], followed by the actor [.24(.10)], Crush [.23(.18)], and the cartoon videos [.07(.09)]. Pairwise comparisons were significant between the therapist and all other conditions, between Crush and the cartoon videos, and between the actor and the cartoon videos (all $p < .05$), but not between Crush and the actor ($p > .05$).

Principal Component Analysis

We conducted a PCA to identify factors to determine common sources of variance in 12 items: ADOS Communication and Social scores; age; mean lengths of utterance in morphemes (MLUm); and attention and gesture/verbal combined percentage scores for each of the four conditions. In this way, we could examine the relationships between the children's characteristics and their behavior during the experiment. In order to have metrics where positive values corresponded to improved behaviors, we created reverse scores for the ADOS by taking the maximum possible score for a participant given the module used for assessment and subtracting the actual score. One child was eliminated from these analyses due to refusal to participate in the Cartoon Video condition, as described earlier. Although we had a relatively small number of participants, we were able to find some strong effects. The analysis identified four factors with eigenvalues >1.0; combined, they explained 88 % of the total variance. (See Table 4 for eigenvalues and factor loadings.) The first component, reflecting social engagement with humans, included positive coefficients for ADOS Social reverse score, ADOS Communication reverse score, MLUm, and attention scores for the actor and the therapist, indicating that these measures were positively correlated with each other. A Cronbach's alpha calculation indicated high reliability of this factor ($\alpha = .82$). The second component, for interactive communication, had positive correlation coefficients for gesture/verbal responses to Crush, the

Table 4 Principal component analysis

Variable	Factor	Weight	Factor eigenvalue
ADOS-S	1	0.87	5.96
ADOS-C	1	0.88	5.96
MLUm	1	0.84	5.96
Attn.—Actor	1	0.74	5.96
Attn.—Therapist	1	0.81	5.96
Comm.—Avatar	2	0.72	1.9
Comm.—Therapist	2	0.82	1.9
Comm.—Actor	2	0.63	1.9
Attn.—Crush	2	0.93	1.9
Attn.—Cartoon	3	0.61	1.46
Comm.—Cartoon	3	0.94	1.46
Age	4	0.91	1.29

therapist, and the actor; as well as attention to Crush, with a highly reliable Cronbach's alpha ($\alpha = .84$). The third component was responsiveness to the cartoon videos, with positive coefficients for attention and gesture/verbal response to them and a lower Cronbach's alpha ($\alpha = .54$). Finally, age was the only factor on the fourth component, indicating that it did not interact strongly with the other variables.

Individual Variation

Even though the number of participants was relatively small, the children did divide into two distinct groups based on their generative language ability. This presented the opportunity to do some preliminary analysis of the children's responses in the four conditions based on these differing levels of ability. We examined correspondences with MLUm (used as an indicator of the child's spoken language level) by dividing the children into two groups: MLUm greater than versus less than 3.0. An MLUm of 3.0 was chosen to be consistent with the minimum criteria for the *Sentences* level as determined by a panel of experts in the language development of children with ASD (Tager-Flusberg et al. 2009). In our participant group, five children had MLUs at or below 2.51; the other seven children had MLUs of at least 3.20. Although these groups were too small for statistical analyses, we noted distinct patterns. In the high MLU group, all of the children raised their hands to speak with Crush, and all but one successfully interacted with him in a back-and-forth, one-on-one conversation. (The final child was not called on.) In the low MLU group, only one child raised his hand to ask a question, and he was unable to successfully participate after being called on. When considering attention to the communication partner in the four conditions, four of the children in the low MLU group had a greater number of 30-s intervals marked as attentive during the Cartoon Videos condition and the Interactive Avatar Crush than to either of the humans. The fifth refused to watch either the *Dora* or *Blue's Clues* cartoon videos. Members of the high MLU group showed the highest number of attentive intervals to the therapist, as well as higher levels of attention overall. The results for the high MLU group's attention to the actor did not show a consistent pattern relative to the other conditions.

Discussion

Although this study includes a relatively small number of participants, the results suggest that, while animated characters can engage and sustain the attention of young children with autism, animation itself does not necessarily increase the likelihood that these children will use appropriate and desirable communicative behaviors. There were no significant pairwise differences between the interactive avatar Crush and the human actor in their abilities to elicit attention, verbal behaviors, and gestures from the children. Moreover, an experienced human therapist elicited superior responses on gestural and verbal measures as compared to both the interactive avatar and the human actor. In addition, while the cartoon videos taken from children's television programming were able to capture an equivalent amount of attention from the participants as the other conditions, they elicited the lowest amounts of gestural and verbal responses. Together, these results suggest that, contrary to previous reports, there may not be an inherent aversive quality that renders humans less appealing than computer-generated avatars as interaction partners for children with autism.

In addition to comparing responsiveness of children with autism in different types of human/nonhuman communication partners, we also examined two important questions facing the use of CBI. The first is whether engagement necessarily corresponds with better behavioral performance in social interactions for children with autism. The argument has been made that children with ASD prefer to interact with computers and animated characters over people (Baron-Cohen et al. 2009), suggesting that children with ASD should pay more attention and be more responsive to nonhumans. However, our results cannot fully support that assumption, finding no differences amongst the children overall in attention between the animated and the human conditions. In fact, despite paying a good deal of attention to the cartoon videos taken from television programs, they consistently showed significantly reduced behaviors in response to those characters. While this does not necessarily mean that the children would not be able to learn from prerecorded animations, it suggests that attention alone is not enough to elicit desired behaviors. Comparing the children's responsiveness across the four conditions, it appears to be important that the communication condition include a contingently-responsive partner, whether that partner is human or animated. Requesting a response from the viewer, as occurred in the cartoon videos condition, was not sufficient. This suggests that a level of interactivity with contingent responses should be included in CBIs to achieve optimal responses from children with autism.

The second major question we addressed is whether animated characters can outperform humans at eliciting social and language behaviors from children with autism. There were no significant differences in attention or combined gesture and verbalization measures between using an actor in combination with an avatar (Crush) or in regular, human form. Moreover, an experienced therapist consistently elicited more gesture and verbal responses than any other condition. While it is possible that the actor who was tele-operating the avatar did not outperform the human actor because of other factors (e.g., a large group vs. one-on-one interaction), it remains clear that a trained individual can still elicit more positive conversational behaviors from a child with autism than an untrained individual, even when the children are given fewer overall opportunities to respond. That is, the children were more likely to respond to the bids for communication during the interaction with the trained therapist.

The findings from the current study are consistent with previous results suggesting that children with ASD exhibit more contingent verbalizations with a virtual character run by an experienced adult than when interacting with a typically developing child (Tartaro and Cassell 2008). The salient contrast may not be human versus animated human, but experienced communication partner (the adult human controlling the responsiveness of the avatar) versus inexperienced and less contingently responsive human child peer. Children with ASD may be most responsive to explicit requests for communication with immediate feedback after a response whether this is provided by a non-human CAT or by a human who is trained to provide this type of scaffolding.

It is still possible that the maximum number of positive conversational behaviors possible would be elicited by a trained therapist using an avatar, a possibility that should be explored in the future. Additionally, future research should examine whether one-on-one interactions with an interactive avatar, such as Crush, can elicit more positive behaviors than when the interaction occurs in a group setting. It is possible that more individual attention from the avatar could positively affect children's response levels. Unfortunately, we were unable to include such a condition in this study.

Furthermore, we did not examine the children's responses to the various individuals during therapeutic interactions, but focused instead on regular conversations. It is possible that the reduced cost of using an avatar, both in terms of time and finances, balances out any loss of impact from not using a live therapist. For example, avatars might be useful in situations where a therapist cannot reach an individual in order to facilitate online interactions. Also, if a virtual character could be preprogrammed with contingent responses without requiring human intervention, it might overcome the weaknesses of the cartoon videos in eliciting social behaviors. Thus, the children could use such software in addition to participating in a traditional therapy program (or in place of it, if such therapy is unavailable due to location or cost). It could be that differences in the effectiveness of a human partner versus a computer-generated partner would be reduced in interactions that are part of a program of treatment. The important point is that contingency of responding appears to be a key element for eliciting both verbal and gestural communication from the children with autism. Using a computer-generated character who does not respond contingently would be expected to reduce the cost-benefit tradeoff of the use of a CBI rather than a human therapist.

As previously noted in recent reviews of the investigations of CBIs with children with ASD (Ploog et al. 2013; Ramdoss et al. 2011), more direct comparisons between intervention programs using CAT and programs using human teachers and therapists are definitely needed. Although the participant group was relatively small, the findings from the current study challenge the basic underlying assumption that children with ASD would respond more positively to an electronically generated character than to a human therapist. They also suggest that measures of attention are insufficient to predict children's responses during various social interactions. Gaining the child's attention is probably not sufficient to promote learning of language and communication skills.

Our results are consistent with the *Social Gating Hypothesis* of language acquisition, which contends that social interaction is important for language learning (Kuhl 2007). According to this model, the social brain "gates" the domain-general computational and cognitive mechanisms that support language acquisition in typically developing children (Kuhl 2007). Furthermore, the reciprocal nature of social interaction is an essential component for promoting the acquisition of language (Kuhl 2007). This model of language development arose from experiments with typical infants in which children exposed to foreign-language material via standard television showed no learning whereas those who participated in social interaction with a human being during the language exposure did (Conboy and Kuhl 2011; Kuhl et al. 2003). As seen with children with typical development who demonstrated greater language learning in response to an actual human interaction than in response to a televised human, the children with autism in the current study produced more gestural and verbal communication when responding to an experienced human therapist than when interacting with cartoon videos of characters from children's television even though the segments selected explicitly invited responses from the children. Furthermore, the responsiveness of the children with autism varied based on the degree of social scaffolding that occurred during the interaction, either from the human therapist, from the human-controlled avatar (and the models of the other children in the group setting), or from the untrained human actor.

Limitations

The participant group in this study was small and of a purposely circumscribed age range so that the activities were developmentally appropriate; however, this limits the generalizability of the results. A strength of the participant group was that it included children with autism with both high and low levels of verbal skill, suggesting that the observations are not limited to children toward either the higher or lower end of the spectrum.

Another limitation of the current study is that, due to the nature of the interactions being studied, they were not equivalent as to setting or the presence of others. The potential confound of the presence of other children occurred for the interactive avatar (Crush). Based on the observed behaviors of the children participating in this study, the other children appeared to have more of a positive than a negative effect. The children with autism frequently responded with a brief delay after observing a response from the other children in the audience. Therefore, the other children in the audience provided explicit cues as to both when a response was expected and what an appropriate response might be.

Additionally, we were unable to randomize the order of conditions because of the nature of the experiment. We needed to confirm autism diagnoses with the therapist before visiting the theme park; thus, the Therapist and Cartoon Videos conditions came before the Crush and Actor conditions. We did not note any consistent increase or decrease in behavior quality over time across the children. Moreover, they were interacting with different individuals in every condition, reducing the possibility that familiarity with an experimenter would affect behavior.

Clinical Implications

Although this study is a preliminary one, it does provide some insight into the important elements of CAT for the development of social and language skills in children with ASD. Animation or non-human forms are not necessarily more effective at eliciting socially meaningful responses from children with ASD. Contingency of the responses provided by the communication partner, whether human or animated, should be considered when designing programs for use with this population of children. Whether a child with ASD has a communication partner that is another human or is computer-generated and/or tele-operated, the partner should engage the child's attention, give clear opportunities for expected or afforded responses, and respond contingently to the child's communication attempts.

Acknowledgments Financial support for this research was provided by Disney Research, Pittsburgh. We wish to thank Lori Georganna for facilitating all aspects of our interaction with the theme park, as well as Casey Lovoy, Dana Barvinchak, Mk Haley, Brooke Kelly, Chelsea Kogelschatz, Dayna Rubendall, and Breanna Fisher for their assistance in performing this research. We thank the Autism Society of America and the Autism Society of Greater Orlando for their help in recruiting.

References

Baron-Cohen, S., Golan, O., & Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions? *Philosophical Transactions of the Royal Society B, 364*, 3567–3574.

Beaumont, R., & Sofronoff, K. (2008). A multi-component social skills intervention for children with Asperger syndrome: The Junior Detective Training Program. *Journal of Child Psychology and Psychiatry, 49*(7), 743–753.

Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of Autism and Developmental Disorders, 31*(4), 377–384.

Bosseler, A., & Massaro, D. W. (2003). Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. *Journal of Autism and Developmental Disorders, 33*(6), 653–672.

Cassell, J., & Tartaro, A. (2007). Intersubjectivity in human–agent interaction. *Interaction Studies, 8*(3), 391–410.

Chen, S. H. A., & Bernard-Opitz, V. (1993). Comparison of personal and computer-assisted instruction for children with autism. *Mental Retardation, 31*(6), 368–376.

Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Developmental Science, 14*(2), 242–248.

Grynszpan, O., Martin, J. C., & Nadel, J. (2008). Multimedia interfaces for users with high functioning autism: An empirical investigation. *International Journal of Human–Computer Studies, 66*(8), 628–639.

Hailpern, J. (2008). *The spoken impact project: Using audio and visual feedback to impact vocalization in non-verbal children with autistic spectrum disorder*. Urbana, IL: University of Illinois at Urbana-Champaign.

Jones, V. (2008). Autism, talking turtles, and the magic of Disney. *Better Health*. Retrieved September 25, 2013 from http://getbetterhealth.com/autism-talking-turtles-and-the-magic-of-disney/2008.06.26.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science, 10*(1), 110–120.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, 100*(15), 9096–9101.

Lasseter, J., & Stanton, A. (2003). *Finding Nemo [Motion picture]*. USA: Walt Disney Pictures.

Linebarger, D. L., & Walker, D. (2004). Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist, 46*, 1–22.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H, Jr, Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(5), 659–685.

Miller, J. F., & Chapman, R. S. (2010). *Systematic analysis of language transcripts [Computer software]*. Madison: University of Wisconsin—Madison.

Milne, M., Luerssen, M. H., Lewis, T. W., Leibbrandt, R. E., & Powers, D. M. (2010). Development of a virtual agent based social tutor for children with autism spectrum disorders. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–9). IEEE.

Mineo, B. A., Ziegler, W., Gill, S., & Salkin, D. (2009). Engagement with electronic screen media among students with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(1), 172–187.

Moore, M., & Calvert, S. (2000). Brief report: Vocabulary acquisition for children with autism: Teacher or computer instruction. *Journal of Autism and Developmental Disorders, 30*(4), 359–362.

Plienis, A. J., & Romanczyk, R. G. (1985). Analyses of performance, behavior, and predictors for severely disturbed children: A comparison of adult vs. computer instruction. *Analysis and Intervention in Developmental Disabilities, 5*(4), 345–356.

Ploog, B. O., Scharf, A., Nelson, D., & Brooks, P. J. (2013). Use of computer-assisted technologies (CAT) to enhance social, communicative, and language development in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(2), 301–322.

Ramdoss, S., Lang, R., Mulloy, A., Franco, J., O'Reilly, M., Didden, R., et al. (2011). Use of computer-based interventions to teach communication skills to children with autism spectrum disorders: A systematic review. *Journal of Behavioral Education, 20*, 55–76.

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., et al. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research, 52*, 643–652.

Tartaro, A., & Cassell, J. (2008). Playing with virtual peers: Bootstrapping contingent discourse in children with autism. In *Proceedings of the 8th International Conference for the Learning Sciences-Volume 2* (pp. 382–389). International Society of the Learning Sciences.

Tjus, T., Heimann, M., & Nelson, K. E. (2001). Interaction patterns between children and their teachers when using a specific multimedia and communication strategy observations from children with autism and mixed intellectual disabilities. *Autism, 5*(2), 175–187.

Williams, J. H., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual–auditory integration during speech imitation in autism. *Research in Developmental Disabilities, 25*(6), 559–575.