

“Hey Google, Do Unicorns Exist?”: Conversational Agents as a Path to Answers to Children’s Questions

Silvia B. Lovato

Northwestern University
Evanston, IL, United States
slovato@u.northwestern.edu

Anne Marie Piper

Northwestern University
Evanston, IL, United States
ampiper@northwestern.edu

Ellen A. Wartella

Northwestern University
Evanston, IL, United States
ellen-wartella@northwestern.edu

ABSTRACT

Children are known to be curious and persistent question-askers. The pervasiveness of voice interfaces represents an opportunity for children who are not yet fluent readers to independently search the Internet by asking questions through conversational agents such as Amazon Alexa, Apple’s Siri, and the Google Assistant. Through a two-week, in-home deployment study involving 18 families (children aged 5-6 and their parents), we report on which questions children choose to ask the conversational agent, the answers the agent provided, challenges in use, and their perceptions of the technology. Based on our analysis, we identify several considerations for the design of voice-based conversational agents that aim to support young children’s question-asking behavior and subsequent development.

Author Keywords

Children; Question-Asking; Conversational Agents; Voice User Interfaces; Digital Assistants; Smart Speakers

CCS Concepts

Human-centered computing → Human computer interaction (HCI) → Sound-based input / output

INTRODUCTION

Young children are prolific and persistent questions-askers [9,42]. Children, before they are able to read and write, have historically depended on others for answers to questions they are unable to figure out on their own. Access to the Internet, through computers and touchscreen devices, puts searching within reach, although typing into search fields and reading search results can prove challenging for children [2,13,15]. The recent proliferation of intelligent conversational agents—particularly digital assistants such as Apple’s Siri, Amazon’s Alexa and the Google Assistant—make it possible for children who are old enough to articulate a question clearly, but not yet fluent in reading or writing, to independently search the Internet by

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
IDC '19, June 12–15, 2019, Boise, ID, USA

© 2019 Copyright is held by the owner/author(s).

Publication rights licensed to ACM.

ACM 978-1-4503-6690-8/19/06 \$15.00

<https://doi.org/10.1145/3311927.3323150>

pressing a button, or saying a wake word, and speaking their questions directly to a device.

Conversational agents are available in smartphones and tablets, which are present in the homes of 98% of American children [37], as well as through home-based smart speakers. Smart speakers in particular have experienced explosive growth recently, with industry reports [1,32] predicting that between 2020 and 2022 more than 50% of households in the United States will have one. Despite their prevalence, we know relatively little about how children use conversational agents to seek answers to their questions. While research has recently begun to study children’s use of conversational agents (e.g., [12,49]), these studies have so far been based on observations of children using voice interfaces during short sessions outside the home. Further, there is little research on the kinds of questions children choose to ask such interfaces. Beyond issues of speech transcription [22], there may be obstacles in children’s use of conversational agents to find and understand online information when used in a naturalistic setting. Despite their anecdotal appeal with pre- and emerging readers, such interfaces were not developed with children as their primary users and may not understand children’s requests easily or may respond with language that is too difficult for children to understand. Missing from the literature is a detailed understanding of in-home use of conversational agents, particularly as children use them together with their parents for an extended period of time, allowing the opportunity to freely experiment with and explore the agent.

We present the results of a field deployment in which 18 families with a child between the ages of 5 and 6 years old used a smart speaker (i.e., a Google Home Mini) for a two-week period. Our analysis reveals that even though children are understood correctly the majority of the time, only half of the answers they get address their questions directly. We include an analysis of the kinds of questions children chose to ask, the most commonly encountered difficulties, and children’s perceptions of the conversational agent after two weeks of use. The present study contributes one of the first naturalistic studies of how young children use an in-home voice-based conversational agent to ask questions and find information.

RELATED WORK

Question-Asking Behavior in Children

The field of developmental psychology has long studied children's question-asking behavior in the context of cognitive development. Children ask questions when they find inconsistencies in their understanding of the world around them [9] and to access information not available through first-hand experience, such as abstract concepts or absent referents [17]. Vygotsky identifies a child's social interactions as key to the child's development, where *more knowledgeable others* scaffold the child's understanding of concepts and processes [45]. Although the idea of more knowledgeable others is generally connected to teachers, parents and other mentors, it can also be represented by peers and other sources of information, including computers. Children's causal questions ("why" and "how" questions) have been found to be an important theory building mechanism that supports conceptual change in childhood [6]. Given the importance of question-asking in child development, the idea of a constantly available, voice-operated device with the ability to answer questions an attractive proposition.

The value of question-asking is dependent on the quality of the answers children receive and on whether they receive a relevant answer at all. Tizard & Hughes [42] studied the questions of young girls at home and at nursery school and found that children who appeared engaged and asked many questions at home hardly asked any at school. They suggest that one of the reasons for this difference might have to do with the perceived availability of the teachers as sources. Unlike parents and teachers, conversational agents in smart devices or home-based speakers are always available. However, we do not yet know how well they support children in this important search for answers. Research has found that children as young as 5 years do take past experience into account when choosing sources of information, both human and electronic [11,30], so young children who experience difficulty interacting with voice interfaces might be reluctant to use them again.

Drawing from this long history in developmental psychology on children's question-asking and learning from others, our study aims to understand how children explore, evaluate and decide whether to rely on a conversational agent as an information source of whom they can ask questions.

Computers as Sources of Information

Whether and how children choose to ask questions of an electronic source relates to how they understand these devices. An extensive review of the literature on children's conceptions of computers between 1968 and 2012 [38] found that among the ideas used by children to describe computers are *omniscient database* and *intelligent machine*. Eight- and 11-year-olds have been found to believe that computers contained the results of all possible mathematical calculations [44], showing that children do see computers and other Internet-connected devices as

sources of information, not just entertainment (video and game playing).

Early studies of children searching electronic databases for information have shown that children tended to use natural language long before it was a viable way to retrieve information (e. g., [26,40]). More recent work on children and Internet search using a traditional computer (e.g., [13,14]) has shown difficulties related to typing and spelling as well as challenges related to how to break down a complex query. Findings from a study on query reformulation using custom-made conversational agents found similar results [49]. An exploratory study focused specifically on children's perceptions of intelligent agents [12] showed that children perceived the agents as friendly and trustworthy. With regards to intelligence, while preschoolers offered mixed responses, the older children in the group, ages 6 to 10, mostly thought the agents were smarter than the children themselves, and that they largely related smarts to access to information, or the ability to answer questions about topics that were familiar to the children.

Interacting with Voice Interfaces

While earlier work [29] has found that younger children (age 4-7) mostly used computers and tablets to play games and watch videos, seldom attempting to search for information; the proliferation of conversational agents has the potential to change that understanding. If voice interfaces make answers accessible to pre- and emerging readers, young children may also come to think of connected devices as sources of information.

An emerging literature has begun to examine children's interactions with voice interfaces and conversational agents in particular. Early studies of children interacting with voice user interfaces focused on children aged 7 and older using specific applications, such as homework tutors (e.g., [46]), or tools meant to diagnose speech impairments (e.g. [25]). More recent work has begun to examine the practices of younger children. A study based on an analysis of YouTube videos found that children's interactions with Apple's Siri were mostly exploratory, where children attempt to understand the agent, followed closely by information-seeking questions [24]. More recently, preschoolers showed persistence when interacting with a non-responsive voice interface inside a tablet game [8]. Although a bug prevented the application from receiving sound input from the children, most of them insisted, repeating the utterance, or varying tone and pronunciation, until encouraged by an adult to give up. Yarosh and colleagues [49] used a Wizard of Oz study to investigate children's (5 to 12 years old) ability to ask questions that required reformulation, as well as children's preferences with regards to personalization (being addressed by the agent by name) and personification (the agent identifying itself by name). They found that many children needed assistance with queries that required reformulation, such as those involving comparisons. Also, while children

preferred personified interfaces to non-personified ones, personalization made no difference.

While Yarosh and colleagues [49] focus on the important issue of reformulation using a custom Wizard of Oz interface, our study looks more broadly at the types of questions children choose to ask when a smart speaker is available in their home for an extended time period, including the challenges they may find besides questions that require reformulation. Other studies on smart speaker use [35,36,39] only consider children tangentially, if at all. The current study builds on this growing body of work by analyzing more widely the types of questions children choose to ask of a conversational agent. Further, prior work has yet to examine naturalistic use of commercially available conversational agents as part of children's question-asking behaviors, a core contribution of the present study.

METHOD

Participants

We deployed smart speakers in family homes, focusing on children ages 5 and 6, along with one parent (N=18 child and parent pairs; 9 children identified as female; mean child age=5.97 years, SD=0.64). We focused on children aged 5 and 6 because this is a key period during which literacy emerges in most children [41]. This allows us to examine the use of this technology among pre- and emerging readers, for whom voice interfaces could increase independence in information access more dramatically.

Participants were recruited from a university database of families who had agreed to be contacted for research, through snowball sampling, and through fliers posted in public libraries, coffee shops and other community spaces. In order to participate, families had to be able to speak English, not have a smart speaker installed in their home at the time of data collection or any time prior, have a working home Internet connection (Wi-Fi), and be willing to have a smart speaker placed in their home for a period of two weeks. Twelve families identified as white, two as Asian, one as Native American, one as African American and two as other or mixed race. Additionally, we asked whether families were of Hispanic or Latino origin and two families identified as Latino. Average family income was about \$120,000 and average family size was 4.1 members. Parents reported an average of 5.9 years of post-secondary education. Three families reported speaking a language other than English at home 50% or more of the time.

Children were given a basic reading assessment before beginning the study (i.e., the Word Reading Test, form A [7]), which consists of five lists of ten words each, in increasing difficulty. It estimates reading grade level (varying from 0 to 10) based on word recognition. Average grade estimated was 1.27 (SD=1.36, range: 0-4, median 1.5). Half (9) of the children were in (or registered for, in the case of summer participants) kindergarten, 4 were in preschool and 5 were in first grade.

Procedure

We conducted a naturalistic study that included two home visits with each parent and child (together). During the first visit, after obtaining parental permission and consent and child assent, parents were asked to fill out a questionnaire with demographic information and children completed the reading assessment. At the end of the visit, a Google Home Mini device was installed in the home, tied to a unique Google account accessible only to the research team. We selected the Google Home Mini because the device provides log files that include raw sound files (the Amazon Echo's did not at the time we began the study) and because we wanted to avoid additional complexity associated with accidental purchase requests (i.e., Alexa might misunderstand a query and initiate a series of questions about making a purchase). Additionally, popular press reviews at the time (e.g., [28]) rated the Google Assistant as better at answering questions than Amazon's Alexa.

Eight identical speakers were used in the study. We marked each device with a sticker containing an ID number and a telephone number families could call if they encountered technical problems. Both the device and the account were left at their default settings (female voice, American accent). No third-party content accounts paid or free (e.g. Spotify, Pandora, Netflix) were made available, but devices could play music using Google Play's free service, which does not play specific tracks or artists, but does play related music stations. Families were told that the researchers would have access to a log containing their interactions. A researcher demonstrated basic functions of the device, such as volume controls, and conducted voice training with the child, which consisted of having the child follow prompts to say each of the wake words twice so that the child's voice would be identified in the usage logs. Participants were told that all family members could use the device without restriction, provided they were told that researchers would have access to their interactions. The child was then encouraged to ask any question they wished. If the child did not come up with a question of their own within a couple of minutes, they were encouraged to ask about the weather.

The research team monitored usage remotely, and families were contacted if no use was detected in more than three days to check for technical problems. We allowed families to keep the device for 2 to 3 three weeks, depending on their availability during that time period (i.e., to accommodate summer travel). When known, travel days were noted in our records of usage. Participants kept the device for an average of 14.28 days (min=14, max=22, median=14). At the end of that period, child and parent pairs were interviewed in their homes about their experience. Additionally, similar to Druga and colleagues [12] we collected likert-style ratings from the child using a six-point smiley-o-meter scale (three positive smileys and three negative frowns, representing -3 to +3). Children were first given a positive/negative forced choice (i.e., is it

Subject Area	Example Descriptive Codes (Subcodes)	Example Utterances	Total Instances
Science & Technology	plants, animals, engineering	“How are pillows made?” “What’s the fastest animal in the world?”	222 (24%)
Culture	history, television shows, fantasy, sports	“How old is Leo Messi?” “Do unicorns exist?”	195 (21%)
Practical	weather, recipes, directions	“What’s the weather today?” “Where’s the closest park?”	170 (18%)
Agent	agent, other agents (Siri, Alexa)	“How old are you?”	142 (15%)
Personal Information	user’s name, age, family information	“What’s my name?” “How old is (sibling’s name)?”	87 (9%)
Language	meaning, spelling, translations	“What does sassy mean?” “How do you spell Arizona?”	46 (5%)
Math	calculations	“What is 8 divided by 2?”	25 (3%)
Jokes	joke questions	“Why did the chicken cross the road?”	17 (2%)

Table 1: Subject areas for questions asked by focal child with total instances for each area.

friendly or not friendly?), and then asked to choose the smiley or frowny face that best represented the degree to which they felt about the device. We used this approach to ask whether the child thought the device was friendly, smart, alive, trustworthy, safe, and funny. Finally, parent and child pairs were compensated with a \$50 gift card. This amount was equivalent to the market value of the Google Home Mini and of the Amazon Echo Dot at the time, allowing the family to replace the research unit with one of their own if they desired. A different Google account was created for each device. Devices were factory reset between families.

Data Analysis

Given the naturalistic focus of this study, our main source of data were device logfiles capturing in-home usage. Usage logs from the devices were downloaded and entered into a spreadsheet for analysis. Each utterance included the date and time, transcription, raw audio file and response provided by the conversational agent. A total of 3683 interactions were recorded among all the families, with an average of 205 interactions per family (range: 61-421). Given our focus on question-asking behavior, we then analyzed all utterances to identify those that were questions (a total of 1,655 interactions) and removed all other utterances (e.g., commands such as volume controls, requests for music; interactions with actions (Google’s name for third party software such as game and story applications); and attempts at social interaction, such as “*Good morning*” or “*I’m home!*”). Although we had conducted voice training with the focal child in each family hoping that this would help identify which utterances were theirs, these were often mislabeled or unidentified. Therefore, we reviewed each question and manually identified the speaker in order to identify the focal child’s

questions and at the same time coded whether the child’s question was transcribed correctly. In the rare instances when more than one user appeared in the sound recording, we identified the person who asked the question as the user (e.g. if the focal child said the wake word but a sibling asked a question, the interaction was coded as the sibling’s). Truly collaborative questions in which the child participated were coded as the child’s. We then coded the subject of the focal child’s questions. We used an iterative process where two researchers first assigned descriptive codes to child questions (e.g. animals, sports, public figures) and then created larger categories to contain the original descriptions (see list of larger categories and descriptive codes in Table 1). This resulted in 38 initial question subject areas, which we further refined and grouped for a total of 8 subject areas. Each question was also coded based on whether the resulting answer addressed the question fully, partly, not at all, or was refused (i.e., the agent responded that it could not help with the question). For this coding, we were as literal as possible. If the response was vague, or meant to be humorous, but did not provide the information requested, it was coded as not addressing the question. Each usage log was double coded by two researchers. Discrepancies were discussed and reconciled.

At the end of the deployment, we conducted a debrief interview with each parent and child pair to capture their view of the experience using the smart speaker. We transcribed and analyzed interview data for emergent themes following a thematic analysis approach [3,4], which involved a process of inductive coding in order to identify patterns in parents and children’s descriptions of their experiences. Children’s ratings of the device using the

Subject Area	Focal Child	Parents	Siblings	Total
Science & Technology	222 (M=12.33, SD=16.96)	40 (M=2.22, SD=3.37)	57 (M=3.16, SD=2.25)	319
Culture	196 (M=10.83, SD=14.57)	60 (M=3.33, SD=6.31)	170 (M=9.44, SD=16.9)	426
Practical	170 (M=9.44, SD=8.15)	88 (M=4.89, SD=4.24)	84 (M=4.67, SD=8)	342
Agent	142 (M=7.89, SD=10.88)	15 (M=0.83, SD=1.54)	47 (M=2.61, SD=2.68)	204
Personal Information	87 (M=4.83, SD=15.41)	11 (M=0.61, SD=1.69)	25 (M=1.39, SD=4.51)	123
Language	46 (M=2.55, SD=3.58)	10 (M=0.55, SD=1.34)	24 (M=1.33, SD=1.94)	80
Math	25 (M=1.39, SD=2.25)	5 (M=0.28, SD=0.67)	33 (M=1.83, SD=4.15)	63
Jokes	17 (M=0.94, SD=1.73)	1 (M=0.05, SD=0.23)	2 (M=0.11, SD=0.32)	20
Total	905	230	442	1577

Table 2: Total questions (Mean, Standard Dev) by subject and type of family member

smiley-o-meter were entered into a spreadsheet for analysis.

FINDINGS

Device Usage Among Families

Throughout the in-home deployment, roughly 45% of interactions with the device were questions (1,655 out of a total 3,683 interactions). On average, each family asked 92 questions during the deployment (SD=64.47; range=20-234 questions). Focal children in each family were responsible for 56% of the questions asked (a total of 924 questions across all families, including questions for which the subject area was not clear to coders, such as questions that used made-up words), but on average asked 51.33 questions each (SD=40.33 range 8-146 questions). Most questions (77% of all questions) asked by focal children were asked during the first 7 days of use, accounting for a more intense exploratory period following placement of the device.

To understand whether household composition related to question asking behavior, such as through the need to share the device or modeling behavior of siblings, we analyzed family size and birth-order of the focal children. Families ranged in size from 3 to 5 members, counting only members able to initiate an interaction with the smart speaker (i.e., excluding infants), with the majority of families having 4 members (n=13), followed by families with 5 members (n=3) and 3 members (n=2). While larger families tended to ask more questions overall (M=144 questions for 3 members; M=209 questions for 4 members; M=224 questions for 5 members), there was no clear relationship between the mean number of questions asked by the focal child and family size (M=41.5 questions asked by children in 3-member families; M=57.66 in 4-member families; M=35.66 in 5-member families). However, the percent of total family questions that were asked by the focal child was, not surprisingly, larger in smaller families, with 90% in 3-member families, 54% in 4-member families

and 53% in 5-member families. That is, in smaller families, the focal child contributed a larger portion of the total questions asked of the conversational agent in that household.

Eleven of the 18 focal children had older siblings, 5 were the oldest and 2 were only children. Children with older siblings asked an average of 50 questions, or on average 46% of their family's questions. Children who are the oldest in their family asked an average of 58 questions, or 80% of the questions asked by their families. Only children asked an average of 41.5 questions, which amounted to 90% of the questions asked by their families. While children with older siblings had additional device use modeling, they still asked close to an average number of questions (mean=50 questions compared to 51.33, the mean number of questions by all focal children). This could perhaps be attributed to more often having to share the device.

Subject Matter of Questions

To understand the topics and domains of children's question-asking behavior, we coded fact-finding questions as belonging to one of eight main subject areas. Table 1 shows the main subject areas, along with examples of descriptive codes and sample utterances. Table 2 shows the break-down of questions by subject and by family member category (i.e. focal child, parents or siblings). Note that this count does not include visitors such as friends and relatives, whose questions added to the total count of questions asked.

Science, the category that includes such descriptive codes as animals, plants, nature and more accounted for 222 of focal children's questions (24%), the most of any category. It was followed closely by the culture category, which includes all pop culture and celebrity questions, accounting for 195 questions (21%). Questions that were of a practical nature, such as questions about the weather, local resources or how to prepare food accounted for 170 of all questions

(18%). The remaining subject areas that kids asked about represented a much smaller percentage of their total questions (language: 46 questions, 5%; math: 25 questions, 3%; jokes: 17 questions, 2%).

Two question categories are specifically about the technology: questions coded as “agent”, about the conversational agent itself, and questions coded as “personal information”, which are about the child and his or her surroundings and family and attempt to explore how much the agent knows about them. Of all the focal children’s queries, 142 questions (15%) asked the conversational agent questions as if it was a person (the “agent” in our code). This was not, however, evenly distributed. Two children account for half of those questions. Eighty-seven questions (9%) were about the child’s own personal information, suggesting they were trying to find out how much the conversational agent knew about them. One child, a 6-year-old girl, was responsible for 66 out of those 87 questions. She succeeded in introducing herself to the agent early in the 2-week period and taught the system the names of her family members. She then asked multiple times a day about those names (i.e. “What’s my name? What’s my dog’s name? etc.), as if to check if the device still remembered them.

Parents, perhaps not surprisingly, asked more practical questions than any other category, followed by culture questions (which include questions about sports, entertainment, and public figures). It is important to note that adults, knowing that researchers would have access to usage logs, might have been intimidated and asked fewer curiosity questions than they would have otherwise. Siblings asked more culture questions than other categories; this was followed by practical and science questions. Most children in our sample had older siblings (n=11), who were responsible for most sibling questions. However, siblings as young as 18 months asked questions.

Transcription and Answer Quality

One open question about conversational agents pertains to their effectiveness in transcribing children’s speech, understanding their questions, and then providing adequate answers to their queries. To assess this, we conducted an analysis of transcription fidelity and answer quality in order to identify breakdowns in children’s attempts to ask questions of the conversational agent. Our analysis shows that focal children’s questions were transcribed correctly an average of 89% of the time. Although not measured in a controlled environment, this performance appears to be an improvement over prior work that reported much lower transcription accuracy for children of the same age [22]. However, even with mostly correct transcripts, only 50% of children’s questions received a full answer, pointing to difficulties unrelated to their ability to pronounce words in a manner that the system could transcribe. To a small percentage of children’s questions, the conversational agent offered alternative questions to which it could provide answers (e.g. “I am not sure, but I can tell you the answer

to a similar question.”). Table 3 below shows the percent of the time that children’s questions were transcribed correctly and answered fully, partly, not at all, or refused.

Transcribed correctly	Answered			
	Fully	Partly	Not at all	Refused
820 (89%)	463 (50%)	73 (8%)	130 (14%)	201 (22%)

Table 3: Transcription Accuracy and Answer Quality

The majority of questions that were refused (i.e., agent responded that it was unable to answer the question) presented fluency, wording, or content challenges. For example, one child (age 5, female) asked, “How does the how do people who know what the weather is like how do they know?” Another (age 6, female) asked, “What the flowers inside of them?” Other refused questions resulted from wording or phrasing mistakes, such as “When was the nearest snow storm?” (age 6, male). Following this, an adult rephrased it as “the most recent snowstorm.” Another child (age 5, female) asked, “How far can English count?” as part of a sequence of questions about counting, possibly meaning to ask, “how far can one count in English?” Other questions were refused due to lack of context or details, such as “What the heck is that?” (age 5, male), “What’s the best-selling children’s author?” (age 6, male), and “How many Iowa games were there in the past?” (age 6, male). Similar to lacking context, other questions sought advice from the conversational agent, to which it was unable to respond. Examples of these questions include: “When is a good time to play?” (age 5, male) and “What should we bring to the beach today?” (age 5, female). Other refused questions were broad in nature or related to fantasy, such as “Where do people come from?” (age 6, female), “How is everything in the world made?” (age 5, female), “Can ghosts fly?” (age 5, male), and “Why do zombies eat people?” (age 5, female).

We also found that “why” and “how” questions, considered an important theory-building mechanism by developmental psychologists (e.g., [6]), were particularly hard for the system to answer. Consider the following exchanges between the conversational agent (CA) and a 6-year-old girl (C):

C: “Do squirrels like pumpkins?”

CA: “Here’s a summary from the website chicagotribune.com: Pumpkins can attract snacking squirrels. A squirrel stands on pumpkins carved into Halloween jack-o'-lanterns Oct. 25, 2004, in Washington. There’s this: Squirrels don’t even particularly like pumpkins.

C: “Why do squirrels eat pumpkins if they don’t like them?”

CA: “Sorry, I don’t know how to help with that.”

The same participant tried to engage the Google Assistant in an exchange about the lifespan of dog breeds, connecting a previous answer to a request for further explanation:

C: "How long does a golden retriever live?"

CA: "The lifespan of the Golden Retriever breed is typically between 10 and 12 years."

C (5 minutes later): "How long does a Labradoodle live?"

CA: "On the website coppercanyonlabradoodles.com, they say: Labradoodles are relatively long lived as a breed and an average life expectancy is between 12 and 15 years."

C: "Why do labradoodles live longer than golden retrievers?"

CA: "Sorry, I don't understand."

Another participant (age 5, female) asked, "how is paper made?" four times across three different days. The first time, she received instructions for making paper airplanes. The other three, she heard the following:

CA: "On the website wonderopolis.org, they say: To make paper from trees, the raw wood must first be turned into "pulp." Wood pulp is a watery "soup" of cellulose wood fibers, lignin, water, and the chemicals used during the pulping process."

Though this answer partly addresses the question (it was coded as a partial answer), it is an incomplete explanation of how paper is made (i.e., only covering how pulp is made). Moreover, the system repeated the same information multiple times rather than viewing this child's attempt at repetitive question-asking as a way of seeking further or different information.

Data from debrief interviews provides a more complete picture of the complexities of question-asking behavior and challenges children experienced in receiving answers. Our analysis of interviews alongside logfile data reveals three specific challenges with effectively receiving answers from the conversational agent.

First, children and their parents reported that responses were long or required interpretation. Some of the questions children asked were met with responses that might contain the answer, but not in a straight forward way. For example, a parent in one interview brought up an incident in which their child (age 6, male) asked, "Who is the second fastest person in the world? The conversational agent responded:

"Here is some information from the web that might possibly help. On the website YouTube.com they say Yohan Blake has been training intensive ahead of the Rio 2016 Olympic where he hopes to compete for the gold medal against Usain Bolt who has the current world record."

After replicating this exchange during the interview, we asked the participant what he thought this meant. He replied, "I have no idea." What parents meant by "requiring interpretation" could also be described as needing an alternate phrasing, requiring prior knowledge, or requiring inferential reasoning about subcomponents of the question. In this example, the child's lack of understanding of the response might come from the fact that understanding that Yohan Blake might be the second fastest person in the world requires the child to know (1) that having the "current world record" is the same as being the fastest and (2) that someone who is training in hopes to compete for the gold with the fastest person might be the second fastest.

A second challenge observed across our data is that answers that were meant to be funny to adults were often confusing to children. For example, a mother of a 5-year-old girl explained:

"I didn't like that when she asked about unicorns it said that [they are everywhere], it made it seem that unicorns are real, because I think it was more geared towards adults, so they find it funny. So, if a child is asking a practical question then, you know, it's going to give sort of a misleading answer."

Looking at the usage logs for further context, the child asked, "Do unicorns exist?" to which the agent responded, "I believe in unicorns. They sound something like this: [sound effects]. There are unicorns everywhere; even in the tech industry." For devices such as conversational agents that may be shared among multiple household members, tailoring responses based on who is asking the question seems like an essential step for future work.

A third challenge that emerged from our analysis is that follow up questions required that the user restate the information just mentioned in the previous question, and children often failed to do that. Although conversational agents remember some information over the course of multiple interactions (e.g., user's name), it appears that they do not currently integrate previous exchanges as context for future questions. For example (child age 5, male):

C: "You know how your freezer works?"

CA: "Here's a summary from the website greenlivingideas.com: Essentially, your fridge and freezer work by continually turning a refrigerant from a liquid to a gas, and then condensing that gas back into a liquid. That liquid is then pumped through a capillary which removes any moisture or contaminants before being sent to your fridge/freezer's evaporator coils."

C: "What's the name of the gas?"

CA: "Here's a summary from the website examples.yourdictionary.com: Oxygen is a pure gas because it is made of one type of item. Pure gases may also be compound molecules. Carbon dioxide

would be considered a pure gas, but it is also a compound molecule.”

Mother: “What’s the name of the *gods* in the freezer?” (gas was incorrectly transcribed as gods)

Both the child and the mother appear to expect the agent to understand a reference to a previous exchange, as a person would, but the agent does not yet support this sort of contextual understanding across sequences of children’s questions.

Conceptions of the Conversational Agent

As part of understanding children’s interactions with a conversational agent, we also aimed to understand their conceptions of how the agent works and what it “knows” about the world or about them. During the debrief interview, after having had the smart speaker at home for two weeks, children were asked several questions that shed light into how they conceive of the conversational agent (i.e., the smart speaker technology). First, they were asked how they referred to the device when talking to family or friends. The overwhelming majority said they simply called the speaker “Google” or “the Google,” which is not surprising since that is the name the device responds to (through the wake words are “Hey, Google” or “OK, Google”). The few children who referred to it differently used descriptive words (e.g. “the Google machine;” “the round gray thing”). Children were also asked to tell us how they would describe the smart speaker to a friend their age who had never seen one before. The descriptions were largely uniform (e.g., “It’s a speaker you can ask questions of.”), though this was likely influenced by the fact that we encouraged children to ask questions of the device when it was set up given the focus of our study.

More interestingly, we asked children how they thought the speaker knew the answers to the questions they asked. They were reassured that there was no right or wrong answer, that we were interested in how they thought about it and encouraged to guess if they didn’t know. Most children’s descriptions of how the speaker works included actions a person would perform, such as look up information on a smartphone. One child (age 5, male) explained, “I think she looks things up on her phone.” Another child (age 6, female) (C) explained to the researcher (R):

C: “Well, I think the Google, it just thinked and thinked and thinked and it got smarter. It is smart, it’s learning each day and every day. “

R: “And how does it learn? How do you think it learns?”

C: “By using its phone.”

R: “It has a phone?”

C: “Yeah, just you can’t, can’t see the hands.”

Is the device...?	Num. children agreed	Mean Rating	Median
Friendly	17	2.27 (SD=1.45)	3
Smart	11	0.66 (SD=2.25)	1.5
Trustworthy	15	1.72 (SD=1.52)	2
Alive	11	0.55 (SD=2.06)	1
Safe	15	1.94 (SD=1.83)	3
Funny	12	0.66 (SD=2.19)	1

Table 4: Children’s self-reported ratings of six dimensions of the device based on a six-point Likert-style scale (-3 to +3).

Another child (age 6, male), spoke of Google directly: “It asks another Google, Google tells it, and then I don’t know. Maybe it searches online and gives me an answer.” These responses echo prior work regarding children’s perceptions of other intelligent technologies [48]. Children have also been found to attribute mental states such as intelligence and feelings to social robots [21].

To complement interview data, we also asked children to rate the conversational agent along six dimensions using a smiley-o-meter, similar to Druga and colleagues [12]. Table 4 shows the average and median ratings given by children for each of the six dimensions. All children but one rated the device friendly to some degree. Referring to the usage logs, the one child (age 6, male) who gave it a -3 score on friendliness had made many attempts at social interactions with the conversational agent and complained about its unenthusiastic responses. Similarly, all but two children rated the device as safe. One of two participants who found it unsafe (male, age 5) was part of a family with three boys, the oldest of which had found a third-party action that made farting noises. This action was difficult to exit, continuing to make the noises after users had made unrelated requests; it made the device difficult to control.

After rating the device as smart or not smart, children were asked whether they thought the device was smarter than they are, or not as smart as they are. Half (n=9) of the children said the device was smarter than they were, 39% (n=7) said it was not as smart as them, and 11% (n=2) refused or were not asked this question. A common explanation for thinking the device was smarter than they were was that it knew more facts. This echoes findings from prior work [12]. Conversely, several of the children who said they themselves were smarter explained that, even though the device could look up information, they were better able to solve problems.

Children were also asked to explain why they thought the device was alive or not alive. Of the children who thought it was alive (11 out of 18), most said that it was alive because it could talk and/or it sounded like a person. The children who said it was not alive said it was a machine, like a robot.

DISCUSSION

Our findings show that children asked questions about a broad range of subjects. Questions about science and technology (24% of all questions) and culture (21%), which included topics such as entertainment, history and politics, were most frequently represented. These categories contain children's questions about the world around them, suggesting that they did think of the device as an information source for topics they were curious about. The two categories of science and technology and culture appeared more frequently than questions coded as practical (18%), which include the topic of the weather, questions about the agent itself (15%), and joke questions (2%) – note that jokes that were not in question format were coded as social interaction attempts and not included in the present analysis.

While the technology has evolved to transcribe children's questions correctly the majority of the time, the conversational agent in our study was still only able to answer about half of their inquiries. One reason for this may be the lack of commercial focus on use cases that are important to children, such as asking information- and explanation-seeking questions. Further, children may be significantly underrepresented in training data used in these systems, widening a gap between children's use practices and behaviors and those of adults.

Nevertheless, the fact that children chose to use the device to ask curiosity questions suggests that there is potential for this technology to play a role in children's self-directed learning. Self-directed learning, which is the ability to decide what to learn about, has been found to generate better information retention both in adults [27] and in younger children [34]. However, just as others have shown that children have difficulty reformulating questions to ask of a conversational agent [49], our analysis shows that they have challenges understanding complex and nested information within a conversational agent's response. Inferential reasoning and theory of mind (e.g., [47]) seem essential to the ability to reformulate questions and understand complex responses, making information seeking through conversational agents challenging for younger children who may not have yet developed these capacities.

The children in our sample did not have a smart speaker in their homes prior to participating in the study. Although most of them mentioned having interacted with Siri or the Google Assistant through tablets or parents' smartphones, this was the first time they interacted with a conversational agent without the benefit of a screen in which to view content. This being a relatively new technology to them, it is not surprising that they would think of the conversational agent as being alive and explain its behavior in human terms. Sherry Turkle [43] found that children in the 1980s thought of computer games that offered variation in how they performed (i.e., imperfectly vs. perfectly) as being alive. Conversational agents offer much more variation than a handheld game of tic-tac-toe, and they have evolved

to sound remarkably natural. This also raises user expectations about their abilities [31]: the more like a person a machine sounds, the more like a person it will be expected to behave, and the fewer concessions users will be willing to make. The children in our study expected the agent to carry out a conversation as a person would, a finding similar to prior work on how children understand intelligent devices (e.g., [48])

In television's earlier days, it was not uncommon for viewers, especially children, to believe that there were an actual set and small people acting out a scene inside the box and that the people on the screen could hear them [33]. It might be harder to find children who still subscribe to those ideas in today's screen-saturated world, just like children in a not-too-distant future might have fewer misconceptions about the nature of conversational agents.

Implications for Design

Several challenges appeared in our analysis of children asking questions of the Google Assistant that help inform subsequent design work. Here, we summarize the key design insights based on our empirical investigation.

Tailor answers to unique users. Some difficulties impacting children could be ameliorated if the system knew that a young child was speaking rather than an older sibling or adult. For example, responses that assume a humorous question, such as the existence of unicorns, could be answered differently to children than to adults. As we realized in the process of analyzing our data, the identification of children's voices is not yet reliable. Therefore, a stop gap solution while voice ID improves could be, as one parent suggested, allowing users to self-identify as they initiate an inquiry.

Simplify and decompose answers to children's questions. The quality of the responses offered, as well as their length and complexity, seem to be the result of a combination of the information available online and the algorithm used to select what information to include when providing responses. For example, in the exchange described above between a participant and the conversational agent about squirrels liking pumpkins, the answer included a sequence that seemed to be part of a photo caption ("A squirrel stands on pumpkins carved into Halloween jack-o'-lanterns Oct. 25, 2004, in Washington."), which is not needed to answer the question, "Do squirrels like pumpkins?" and adds length and complexity. The answer that was heard at the end would have been better on its own: "Here's the thing, squirrels don't even particularly like pumpkins." Other responses required children to perform higher level reasoning and draw on other worldly knowledge (e.g., understanding that training to race against the fastest person might mean someone is second fastest). Continued development in the ability to identify information that answers a question succinctly and directly will benefit all users, not just children. Given the challenges of achieving

this in real-time conversational AI systems, a crowd-powered approach may be a promising alternative [19].

Understand context through prior questions and responses. Children often fail to provide the necessary context for initial questions, as Yarosh and colleagues [49] also show. Beyond this, children (and adults) expect the agent to remember what was just said and be able to use that information as context for asking a follow-up question. That is, they see interaction as building across a sequence of questions and responses, and they expect these prior interactions to contribute to conversational grounding (i.e., a mutual understanding between two parties regarding what they are talking about) [5,10]. Such grounded, contextual understandings that take a sequence of interactions (e.g., multiple questions and responses) into account are not yet built into commercial conversational agents, as this is a difficult challenge for conversational AI systems in general [20]. An interim solution could be to either ask the child to clarify, as prior work suggests [49], or to confirm. For example, in the case described above where a participant wanted further information about the gas that makes the refrigerator work, when subsequently asked “what’s the name of the gas?” the agent could confirm if the user is indeed still talking about the same topic.

Adapt responses based on repetitive questioning. In our study, there were a few cases when children asked the same question over and over. This may be because they did not understand the answer or perhaps wanted more or different information on the same topic. If the conversational agent is able to identify and tailor responses to specific users, it could take repeat question-asking behavior into account for each individual as part of their own history. With this information, the system could then limit how many times it will repeat the same answer, after which it could either offer another option or suggest that the user ask the question in a different way. The open-ended fact-finding question-asking behavior we observed in our data is distinct from other uses of conversational agents, such as asking practical questions (e.g., What’s the weather today?) that are likely to be repeated daily. Thus, having the system take into account the *kind* of question asked in addition to frequency seems like an important next step in supporting children’s question-asking behavior.

Source curation and presentation. Another point worth noting is that, even though the Google Assistant mentions the source of many of its responses, reading out the URL before the answer to the question, none of the children remembered this when asked how they thought the device worked. This is perhaps not surprising in a media world dominated by apps on touchscreens, where few 5- and 6-year-olds have a need to understand URLs. However, if children are able to seek information independently through a conversational agent before they are able to understand where the information comes from, they might benefit from curation managed by parents and educators, who could be given access to source options to be drawn from when the

young children in their care are interacting with the technology. Alternatively, redesigning the presentation of source information could be a way to help children learn about where information comes from and source credibility at a young age.

Limitations and Future Research

One limitation of the present analysis is that our sample included predominantly middle- and upper-class families with highly educated parents. Therefore, we do not know how children of other socio-economic backgrounds and parental education levels might use this technology and what challenges they might encounter. For example, research has shown that children of lower SES react differently than wealthier, more educated counterparts when they receive dismissive or incomplete answers from adults [23,42], being less likely to insist or to come up with their own explanations. Therefore, future work on question-asking through conversational agents should consider a wider diversity of socio-economic backgrounds and parental education levels, as usage patterns and strategies may vary across different groups of children. Additionally, as conversational agents continue to expand their reach into homes, children might grow more habituated with how they work. Although our analysis uncovers initial question-asking behaviors and practices in situ, a longitudinal study of how children incorporate this technology into their repertoire of information sources over time could extend our understanding and inform future child-focused versions of conversational agents. Finally, conversational agents are now available in many languages and countries. Prior work has shown differences between cultures in children’s question-asking behavior [16,18]. Our sample includes only English-speaking American children. Future work should include children in other cultures and how their question-asking habits might impact their experiences trying to obtain answers from a conversational agent.

CONCLUSION

Our results show that 5- to 6-year-old children do choose to ask informational questions about the world around them when a conversational agent is available in their home, and that their questions are transcribed correctly the majority of the time. However, they do run into hurdles while doing this, both when asking questions and when interpreting answers. This points to a need for more research that investigates how best to support children’s inquiries, including the creation of tools that give parents and educators, who are familiar with the child, the ability to curate sources of information that can provide answers in a format child can understand, as well as request additional contextual information when children do not provide it.

SELECTION AND PARTICIPATION OF CHILDREN

In this work, children were selected to participate when a parent volunteered to enroll in the study by responding to a recruitment flyer, email or social media post. Upon arrival in the family home, it was explained to children that the study was about how children interact with the smart

speaker and that the researchers would have access to their interactions, that we would administer a short reading assessment and that we would record a video of our conversations with them, but that the video was solely for use by the research team and would not be shared online or with any other audience. If they agreed to take part, they were asked to cosign the permission/consent form along with their parent. The study protocol and the permission/consent document were approved by the university's Institutional Review Board.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the families who participated in the study. Without their willingness, this work would not have been possible. We would also like to acknowledge Su Min You for her assistance with recruitment and coding usage logs.

REFERENCES

- Giselle Abramovich. 2018. Study Finds Consumers Are Embracing Voice Services. Retrieved September 2018 from <https://www.cmo.com/features/articles/2018/9/7/adobe-2018-consumer-voice-survey.html#gs.6DCN6JNj>
- Dania Bilal. 2002. Children's use of the Yahoo!igans! Web search engine. III. Cognitive and physical behaviors on fully self-generated search tasks. *Journal of the American Society for information science and technology* 53, 13: 1170-1183.
- Richard E. Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development*. Sage.
- Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. In *Qualitative Research in Psychology*, 3, 2: 77-101.
- Susan E. Brennan. 1998. The grounding problem in conversations with and through computers. In *Social and cognitive approaches to interpersonal communication*. 201-225.
- Maureen A. Callanan and Lisa M. Oakes. 1992. Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development*, 7, 2: 213-233.
- Jeanne S. Chall, Florence G. Roswell, Mary E. Curtis, and John Strucker. 2003. *Quick adult reading inventory*. Continental Press, Inc.
- Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why Doesn't It Work? Voice-Driven Interfaces and Young Children's Communication Repair Strategies. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, pp. 337-348.
- Michelle M. Chouinard, Paul L. Harris, and Michael P. Maratsos. 2007. Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*: i-129.
- Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991: 127-149.
- Judith H. Danovitch and Reem Alzahabi. 2013. Children show selective trust in technological informants. *Journal of Cognition and Development* 14, 3: 499-513.
- Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. Hey Google is it OK if I eat you?: Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children*, pp. 595-600.
- Allison Druin, Elizabeth Foss, Leshell Hatley, Evan Golub, Mona Leigh Guha, Jerry Fails, and Hilary Hutchinson. 2009. How children search the internet with keyword interfaces. In *Proceedings of the 8th International conference on interaction design and children*, pp. 89-96.
- Allison Druin, Elizabeth Foss, Hilary Hutchinson, Evan Golub, and Leshell Hatley. 2010. Children's roles using keyword search interfaces at home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'10)*, pp. 413-422.
- Elizabeth Foss and Allison Druin. 2014. Children's Internet Search: Using Roles to Understand Children's Search Behavior. *Synthesis Lectures on information concepts, retrieval, and services*, 6, 2: 1-106.
- Mary Gauvain, Robert L. Munroe, and Heidi Beebe. 2013. Children's questions in cross-cultural perspective: A four-culture study. *Journal of Cross-Cultural Psychology* 44, 7: 1148-1165.
- Paul Harris. 2012. *Trusting What You're Told: How Children Learn From Others*. Harvard University Press.
- Ronald W. Henderson and Angela B. Garcia. 1973. The effects of parent training program on the question-asking behavior of Mexican-American children. *American Educational Research Journal* 10, 3: 193-201.
- Ting-Hao Kenneth Huang, Joseph Chee Chang, and Jeffrey P. Bigham. 2018. Evorus: A Crowd-powered Conversational Assistant Built to Automate Itself Over Time. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 295. ACM.
- IBM. 2018. How conversation (with context) will usher in the AI future. Retrieved December 2018 from <https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-conversation.html>
- Peter H. Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Nathan G. Freier, Rachel L. Severson, Brian T. Gill, Jolina H. Ruckert, and Solace Shen. 2012. "Robovie, you'll have to go into the closet now": Children's social

- and moral relationships with a humanoid robot. *Developmental psychology* 48, 2: 303.
22. James Kennedy, Séverin Lemaignan, Caroline Montassier, Pauline Lavalade, Bahar Irfan, Fotios Papadopoulos, Emmanuel Senft, and Tony Belpaeme. 2017. Child speech recognition in human-robot interaction: evaluations and recommendations. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 82-90.
 23. Katelyn E. Kurkul and Kathleen H. Corriveau. 2018. Question, explanation, follow-up: A mechanism for learning from others. *Child Development* 89, 1: 280-294.
 24. Silvia Lovato and Anne Marie Piper. 2015. "Siri, is this you?": Understanding young children's interactions with voice input systems. *Proceedings of the 14th International Conference on Interaction Design and Children*, pp. 335-338.
 25. Andreas Maier, Flonan Hönig, Stefan Steidl, Elmar Nöth, Stefanie Horndasch, Elisabeth Sauerhöfer, Oliver Kratz, and Gunther Moll. 2011. An automatic version of a reading disorder test. *ACM Transactions on Speech and Language Processing (TSLP)* 7, 4: 17
 26. Gary Marchionini. 1989. "Information-seeking strategies of novices using a full-text electronic encyclopedia." *Journal of the American Society for Information Science* 40, 1: 54-66.
 27. Douglas Markant, Sarah DuBrow, Lila Davachi, and Todd M. Gureckis. 2014. Deconstructing the effect of self-directed study on episodic memory. *Memory & Cognition* 42, 8: 1211-1224.
 28. Jay McGregor. 2017. Amazon Echo vs. Google Home Black Friday: Which is Best? Retrieved January 2018 from <https://www.forbes.com/sites/jaymcgregor/2017/11/15/black-friday-amazon-alexa-vs-google-home/#415774512cb7>
 29. Susan McKenney and Joke Voogt. 2010. Technology and young children: How 4–7 year olds perceive their own use of computers. *Computers in Human Behavior* 26, 4: 656-664.
 30. Candice M. Mills, Cristine H. Legare, Meredith G. Grant and Asheley R. Landrum. 2011. Determining who to question, what to ask, and how much information to ask for: The development of inquiry in young children. *Journal of Experimental Child Psychology*. 110: 539-560.
 31. Clifford Ivar Nass and Scott Brave. 2005. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT Press.
 32. Nielsen. 2018. (Smart) Speaking My Language: Despite Their Vast Capabilities, Smart Speakers Are All About The Music. Retrieved December 2018 from <https://www.nielsen.com/us/en/insights/news/2018/smart-speaking-my-language-despite-their-vast-capabilities-smart-speakers-all-about-the-music.html>
 33. Peter Nikken and Allerd L. Peeters. 1988. Children's perceptions of television reality. *Journal of Broadcasting & Electronic Media* 32, 4: 441-452.
 34. Eric Partridge, Matthew G. McGovern, Amanda Yung, and Celeste Kidd. 2015. Young children's self-directed information gathering on touchscreens. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society, Austin, TX. Cognitive Science Society*.
 35. Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. Alexa is my new BFF: Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, 2853–2859.
 36. Aung Pyae and Tapani N. Joelsson. 2018. Investigating the usability and user experiences of voice user interface: a case of Google home smart speaker. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pp. 127-131
 37. Victoria Rideout. 2017. The Common Sense census: Media use by kids age zero to eight. Robb, M., (ed.) *Common Sense Media*
 38. Michael T. Rucker and Niels Pinkwart. 2016. Review and discussion of children's conceptions of computers. *Journal of Science Education and Technology* 25, 2: 274-283.
 39. Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. Hey Alexa, What's Up?: A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*, pp. 857-868.
 40. Paul Solomon. 1993. Children's Information Retrieval Behavior: A Case Analysis of an OPAC. *Journal of the American Society for information Science* 44 (5): 245-264.
 41. William H. Teale and Elizabeth Sulzby. 1986. Emergent Literacy: Writing and Reading. *Writing Research: Multidisciplinary Inquiries into the Nature of Writing Series*. Ablex Publishing Corporation,
 42. Barbara Tizard and Michael Hughes. 1984. *Young Children Learning*. John Wiley and Sons.
 43. Sherry Turkle. 1984. *The Second Self: Computers and the Human Spirit*. MIT Press.
 44. Mike Van Duuren, Barbara Dossett, and Dawn Robinson. 1998. Gauging Children's Understanding of Artificially Intelligent Objects: A Presentation of "Counterfactuals" *International Journal of Behavioral Development* 22(4), 871-889.

45. Lev S. Vygotsky and Michael Cole. 1978. *Mind in society: The development of higher mental process*. Harvard University Press.
46. Wayne Ward Ronald Cole, Daniel Bolaños, Cindy Buchenroth-Martin, Edward Svirsky, Sarel Van Vuuren, Timothy Weston, Jing Zheng, and Lee Becker. 2011. My science tutor: A conversational multimedia virtual tutor for elementary school science. *ACM Transactions on Speech and Language Processing (TSLP)*7, 4 :18.
47. Henry M. Wellman and David Liu. 2004. Scaling of theory-of-mind tasks. *Child development* 75, 2: 523-541.
48. Julia Woodward, Zari McFadden, Nicole Shiver, Amir Ben-hayon, Jason C. Yip, and Lisa Anthony. 2018. Using Co-Design to Examine How Children Conceptualize Intelligent Interfaces. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 575.
49. Svetlana Yarosh, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, Ye Yuan, and A. J. Brush. 2018. Children asking questions: speech interface reformulations and personification preferences. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, pp. 300-312.