

EMPIRICAL ARTICLE

WILEY

“Alexa, let me ask you something different” Children's adaptive information search with voice assistants

Cansu Oranç  | Azzurra Ruggeri 

Max Planck Institute for Human Development,
Berlin, Germany

Correspondence

Cansu Oranç, Max Planck Institute for Human
Development, MPRG iSearch - Information
Search, Ecological and Active Learning
Research with Children, Lentzeallee
94, 14195, Berlin, Germany.
Email: oranç@mpib-berlin.mpg.de

Abstract

In this study, we investigated how children interact with voice assistants, particularly focusing on what kinds of questions they ask and how they react to the responses obtained. We recorded 3- to 10-year-old children's ($N = 43$) spontaneous interactions with Amazon Alexa, and analyzed the questions they asked, as well as how they adjusted their information search based on the responses received. Our results confirm previous work in showing that children's questions are mostly information-seeking, yet the type of questions children ask also depends on their age and familiarity with voice assistants. For example, children who are younger and less familiar with voice assistants are more likely to ask questions about themselves and their environment (e.g., “What is my sister's name?”). We also show for the first time that, even though all children are sensitive to the relevance and accuracy of voice assistants' responses to a certain extent, older children are more likely to change the topic and type of the questions asked upon receiving irrelevant or uninformative responses. This study shows that, with age and familiarity, children become more sensitive to the behavior, informativeness, and constraints of artificial agents, growing into adaptive and sophisticated technology users.

KEYWORDS

children, information search, question-asking, voice assistants

1 | INTRODUCTION

In the new digital era, voice assistants (VAs) are becoming ubiquitous in children's daily lives. Parents consult voice assistants such as Apple Siri and Amazon Alexa on topics such as weather conditions for a family trip, or directions to a birthday party. Although withdrawn due to privacy controversies, the toy company Mattel developed the idea of Aristotle, a voice-activated device specifically designed for monitoring and supporting children's daily lives and development. VAs and AI-powered educational toys (e.g., CogiToys Dino that operates on IBM's question-answering system Watson) are now being introduced

in formal and informal learning settings, answering children's questions in classrooms and at home (Lovato & Piper, 2015; Terzopoulos & Satratzemi, 2020). These new tools and practices raise important theoretical questions: What do children seek in their interactions with VAs, and what do they expect? How do they steer their interactions with VAs, adjusting their inquiry behavior based on the responses received, as well as on their previous experience with similar devices? Despite our children growing up increasingly surrounded by VAs, we still do not know much about the impact these interactive technologies have on children from a cognitive and social perspective. In this study, we address these gaps by thoroughly examining 3- to 10-year-old

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Human Behavior and Emerging Technologies* published by Wiley Periodicals LLC.

children's interactions with VAs, and the characteristics of their inquiry behavior.

2 | LEARNING BY ASKING QUESTIONS

Asking questions is a powerful learning tool that allows young learners to be precise about the information they want, inquire about absent objects or events, and address abstract concepts (Jones et al., 2020). Although the ability to ask effective questions develops from age 4 to adulthood (e.g., Herwig, 1982; Mosher & Hornsby, 1966; Ruggeri et al., 2016; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015), from a very early age, children are able to tailor their questions to maximize their information gain (Ruggeri et al., 2017). Children ask many questions, for a variety of reasons (Chouinard et al., 2007), yet they do not just fire off a question and leave it there. They critically evaluate the answers they receive from other people, act and react upon how informative they think the answer is. For instance, if satisfied, children tend to end their search for information or ask a follow-up question to receive additional information but if not, they persist and repeat their original question (Chouinard et al., 2007; Frazier et al., 2009, 2016). Although preschoolers are better at this than toddlers, it is not before the elementary school years that children become sensitive to the level of informativeness of a response (Ronfard et al., 2018). In other words, children carry on their investigations to receive satisfactory information and with age, they become better at doing so. What do they do when the respondent is an artificial agent instead of a human, such as a voice assistant?

3 | ASKING QUESTIONS TO VOICE ASSISTANTS

Children can now verbally search for information by asking questions of VAs, and engaging in conversations with them in natural language. This is a novel and exciting opportunity, especially for young children who do not yet have reading and writing skills to query the web (Lovato et al., 2019), and for those with difficulties in reading and writing, such as children with dyslexia. Limited research provides some insight into the interaction between children and VAs. Based on a survey with parents and an analysis of YouTube videos, Lovato and Piper (2015) found that children use VAs mostly to seek information (e.g., "Do whales sleep?"). They also seek to understand the assistant as an agent (e.g., "What is your favorite food?"), to have fun with it, either by requesting jokes or asking silly questions, and, to a lesser extent, to give commands, for example to send text messages. A follow-up study (Lovato et al., 2019) consolidated these findings by recording 5- and 6-year-old children's interactions with a VA at home, and showed that most of their questions were seeking factual information (e.g., about science, culture, daily and local information), followed by personal questions about the VA (e.g., its age), and about the children themselves (e.g., their name). These two studies reveal that children ask information-seeking and agent-related questions to VAs, but what do they receive in return?

The same study by Lovato et al. (2019) suggests that children receive an answer to their questions to VAs only half of the time. Children's questions were often only partly answered or not answered at all, mostly due to the phrasing issues, lack of crucial details, or because the questions sought something beyond the VA's abilities such as its advice. Children often try to repair such miscommunications they experience with VAs (Beneteau et al., 2019; Cheng et al., 2018; Yuan et al., 2019). For instance, Yuan et al. (2019) showed that 5 to 12-year-olds often try to reformulate their questions to receive a satisfactory answer, for example, by adding context or expanding the pronouns in the original question, with mixed success. Importantly, older children and adults were more likely to employ effective, successful reformulation strategies compared to young children. These studies together suggest that children consider VAs as a source of information, yet they often struggle in communicating their queries.

4 | PRESENT STUDY

The above-summarized research has mainly focused on children's questions, yet there are two other crucial components of an interaction between two agents: the answers received, and the reactions to those answers. Children's reactions to others' answers were suggested to be an indicator of the purpose of their initial questions. If the question was intended to acquire knowledge, children should persist, repeating or reformulating the question until they receive an informative response, whereas if the question was only meant to keep the conversational engagement going, any answer could be considered "satisfactory" (Frazier et al., 2009). Research on preschoolers' interactions with adults found support for the former, showing that children indeed ask questions to obtain answers and learn (Frazier et al., 2009; Kurkul & Corriveau, 2018). As suggested by Danovitch (2019), studying how children's success in finding information via Internet-based devices influences their subsequent information search will help us to understand children's self-directed learning in a digital world.

To our knowledge, this is the first study examining the *adaptiveness* of children's inquiry behavior with VAs. That is, it is the first study to explore whether and how children adapt the type of questions asked to the answers obtained by the virtual conversational partner. In contrast to previous research, this study addresses children's interactions with VAs in their entirety, as it thoroughly examines not only their questions, but also their reactions to the responses they receive. It also goes beyond prior work by adopting a more quantitative approach in analyzing the collected data, and targeting a wider age range.

5 | METHOD

5.1 | Participants

To shed light onto children's interaction with VAs across different developmental periods that were also studied in previous research (e.g., Lovato et al., 2019; Yuan et al., 2019), we aimed to include

preschoolers (age 3–5) and elementary school-aged children (6–10), who have been found to crucially differentiate in their question-asking behavior (e.g., the efficiency and adaptiveness of their questions; see Ronfard et al., 2018). Unlike most previous research, we included children already from age 3, as their verbal language proficiency can be sufficient to communicate with VAs (Lovato et al., 2019).

Fifty-four German-speaking children between the ages of 3 and 10 were recruited at the Berlin Zoo in Germany. Of these, seven children did not attempt to ask any questions, and four children tried but failed to communicate with Alexa (i.e., their questions were either incomprehensible or they did not go beyond asking if Alexa could hear them). Hence, the final sample included 43 children (Age range: 42–128 months, $M = 90$ months, $SD = 25.2$ months).

Institutional Review Board approval was obtained by the Max Planck Institute for Human Development in Berlin (protocol: N-2020-05a), and parents gave informed consent to allow their children to participate in the study.

5.2 | Procedure

The study was conducted in a designated testing area at the zoo. Children were seated at a table, opposite to the experimenter, with Alexa placed in front of them. We used Alexa Echo Dot third generation for the first 23 participants, but then switched to Alexa Echo Dot second generation due to technical issues with the previous device.

The experiment consisted of two phases. In the Q&A phase, the experimenter introduced Alexa to the child as “Alexa here is a voice assistant. You can ask her questions and she will answer” and asked “Alexa, can you hear me?” to demonstrate how it worked. The experimenter then asked the child if they would like to talk to Alexa, and invited them to ask questions if the answer was positive. If the child hesitated to initiate the conversation, the experimenter encouraged them by saying “Why don't you just try and say something?” When the child seemed to be hesitant to continue the conversation, the experimenter asked if they had anything else to ask, and if not, they proceeded to the next phase. In the *Abilities interview* phase, children were asked a series of open questions aimed at eliciting their understanding and assumptions about Alexa (e.g., “Do you think Alexa can get older?”). The complete list of questions can be found in Appendix. Following the two phases, parents were asked to complete an *Experience questionnaire* in which they were asked whether they had a smart speaker at home (i.e., a stand-alone device with an integrated VA, such as Amazon Echo or Google Home), how often they and their children used VAs on a 5-point scale (from *never* to *at least once a day*), and for what purposes. Children were gifted stickers at the end of the study to thank them for their participation.

5.3 | Data coding

The coding scheme was developed based on previous work. In particular, to code children's questions, we used previous research examining

children's interactions with VAs (e.g., Lovato & Piper, 2015). Alexa's responses were categorized based on analyst firms' benchmark studies comparing the intelligence of different VAs (e.g., Cognilytica, 2018). The coding scheme for children's follow-up behavior was developed upon data collection, as we realized that the existing coding schemes from previous work investigating child–adult interactions (e.g., Frazier et al., 2009) did not apply to our dataset. The first author studied the transcriptions to detect the change patterns. A first research assistant then coded the entire dataset based on the criteria presented below. A second research assistant independently coded 30% of the data. Inter-rater reliability was calculated using Cohen's kappa. Disagreements were resolved by a third rater.

5.3.1 | Q&A phase

Children's interactions with Alexa—their questions, as well as the response received—were recorded, transcribed, and then coded based on a predetermined coding scheme.

Children's questions

Children's utterances were coded as “questions” whenever they explicitly inquired about a piece of information or demanded an answer. In this sense, *questions* included both actual questions (e.g., “What do gorillas eat?”) and commands/prompts (e.g., “Tell me a joke”). Thus, “statements” were excluded from the analyses, even when they clearly targeted Alexa (e.g., “Alexa, I got new shoes”).

Children's questions were coded into four categories: (a) *Information-seeking* questions, targeting factual information (e.g., “Why do babies cry?,” “How is the weather today?”); (b) *Agent-related* questions, inquiring about Alexa (e.g., “What is your favorite color?,” “How are you?”); (c) *Asker-related* questions, inquiring about children's own life and environment (e.g., “What is my favorite food?,” “What is my sister's name?”); (d) *Entertainment-seeking* questions, asking Alexa to perform an action for entertainment purposes (e.g., “Tell me a joke”, “Can you play a song?”). Inter-rater reliability was very high (Cohen's $\kappa = .898$, 95% CI [0.82–0.976], $p < .001$).

Voice assistant's responses

Alexa's responses were coded into four categories: (a) *Proper*, that is, relevant and correct answers; (b) *Improper*, that is, irrelevant answers; (c) *Unknowing*, when Alexa stated that it did not know the answer; (d) *Demanding*, that is, answers that were convoluted or generally complicated to understand and required children to make some kind of an inference to obtain the answer, such as the answer “I've been available since November 6, 2014” to the question “How old are you?” Inter-rater reliability was very high (Cohen's $\kappa = .826$, 95% CI [0.726–0.926], $p < .001$).

Children's follow-up behavior

For each question-response pair, we coded whether the following question was different (coded as 1) or not (coded as 0) across four dimensions: (a) *Question type*, a change of category, for example, asking an information-seeking question after having previously asked an asker-related question; (b) *General content*, that is, a radical change of

the topic of conversation, for example, asking a question about nature after having previously asked a question about music; (c) *Specific content*, that is, a change of specific topic within the same general theme, for example, asking a question about animals after having previously asked about plants; (d) *Question function and purpose*, independently of the questions' content, for example, asking a *what* question after having previously asked a *how* question. This detailed coding allowed us to capture fine-grained transitions in children's inquiry behavior. Note that, although these dimensions are in principle independent, there were some more and less obvious dependencies. For example, a General-content change always entailed a change of Specific-content as well. Similarly, a Question-type change was often accompanied by a change in Question-function. Inter-rater reliability was overall fairly high: (question type change: Cohen's $\kappa = .968$, 95% CI [0.905–1.031], $p < .001$; general content change: Cohen's $\kappa = .669$, 95% CI [0.508–0.83], $p < .001$; specific content change: Cohen's $\kappa = .857$, 95% CI [0.747–0.967], $p < .001$; question function change: Cohen's $\kappa = .787$, 95% CI [0.648–0.926], $p < .001$).

5.3.2 | Abilities interview phase

Children's answers to the question "What is Alexa good for?" were coded as referring to Alexa being a source of information (e.g., "It can answer your questions") or not (e.g., "It can play music"). All other questions were coded on a scale ranging from 1 (not having the ability in question; e.g., "not smart") to 3 (having the ability in question; e.g., "is smart"), where 2 referred to indecisive responses (e.g., "sometimes smart").

5.3.3 | Children's previous experience with voice assistants

Based on the results of parents' questionnaires, we extracted an experience score that combined information about families' smart speaker ownership and parents' report of children's frequency of use of VAs. The scores ranged from -1 (i.e., the child does not have a smart speaker at home, and never uses it) to 4 (i.e., the child uses VAs at least once a day). Note that children who were reported to have experience with VAs (i.e., those who scored 1 to 4) may not necessarily have a smart speaker at their own home, as they may be using it at their grandparents' house or may be using a VA embedded in another device (e.g., Siri on an iPhone). This scoring allowed us to differentiate children who do not have any exposure to VAs from those who do not use VAs themselves, but might be familiar with the device as they have a smart speaker at home.

6 | RESULTS

All analyses were conducted using the software Jamovi that runs on R (The Jamovi Project, 2020). Because the parents of 13 children did not complete the Experience questionnaire, we first report a full set of analyses using only age (in months) as predictor ($N = 43$). We then performed the analyses again including only the subset of children

whose parents completed the Experience questionnaire ($N = 30$, Age range: 42–128 months, $M = 93.37$ months, $SD = 25.36$ months), and adding children's experience with VAs as a second predictor.

6.1 | Overview of the sample characteristics

Overall, children in our sample were not very experienced with voice assistants (Range: -1 to 4, $M = 0.97$, $SD = 1.92$; see Figure 1). Descriptive statistics of children's answers to the abilities interview questions are presented in Table 1. We performed Kendall's Tau-b correlations between children's age (in months) and experience, and their responses to the abilities interview questions. We only found a significant negative association between children's age and their attribution of social capacity to Alexa, $\tau_b = -.416$, $p = .003$. All remaining p values were above .05.

6.2 | Overview of children's questions and the responses they received

Children asked an average of 6.91 questions ($SD = 5.91$, Range: 1–26 questions). Figure 2 illustrates the distribution of the number of

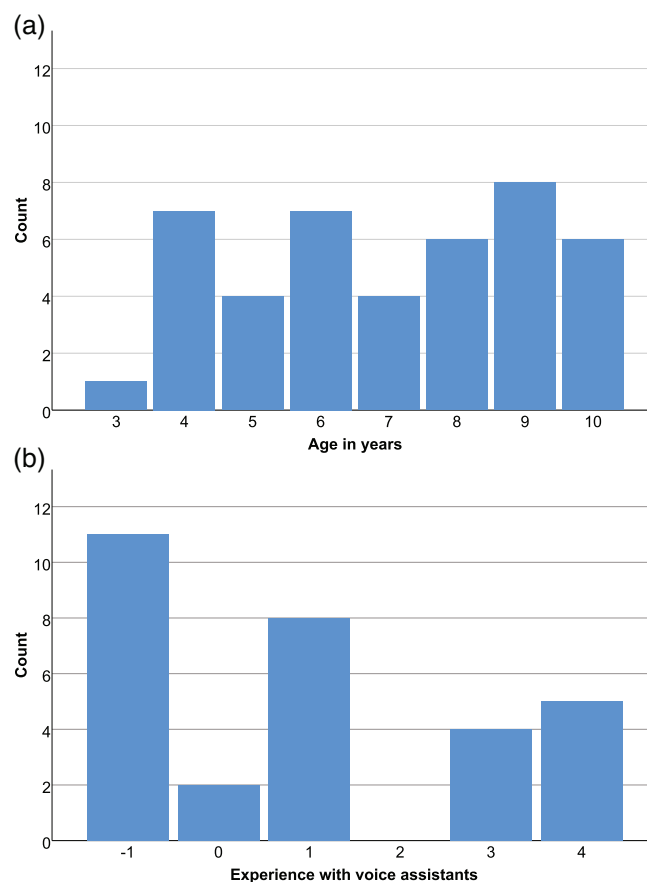


FIGURE 1 Histograms of the sample's (a) age and (b) experience with voice assistants (-1 = the child does not have a smart speaker at home, and never uses voice assistants; 4 = the child uses voice assistants at least once a day)

TABLE 1 Descriptive statistics of children's answers in abilities interview (range: 1–3)

	<i>n</i> ^a	Mean	SD
Truthfulness	39	2.85	0.43
Intellectual capacity	39	2.59	0.72
Cognitive ability	36	2.25	0.87
Biological property	38	1.87	0.99
Social capacity	34	1.82	0.97
Emotional capacity	38	1.63	0.91

^aNote that not all children answered all the abilities interview questions.

questions they asked. Overall, 8.54% of these questions were only aimed at confirming whether Alexa was working, and did not have any content (i.e., empty questions such as “Alexa, can you hear me?”). The linear regression analysis assessing the proportion of content questions (over all questions asked, including empty ones) revealed that age (in months) was a significant predictor, $F(1, 41) = 6.73, p = .013, R^2 = .141$: Younger children asked a higher proportion of empty questions compared to older children ($\beta = 0.376, 95\% \text{ CI } [0.083-0.668], t = 2.594, p = .013$). Empty questions were excluded from subsequent analyses.

We performed a linear regression analysis to assess the proportion of questions to which children received a response, and found that age was a significant predictor, $F(1, 41) = 15.109, p < .001, R^2 = .269$: Overall, older children were more likely to receive a response to their questions compared to younger children ($\beta = 0.519, 95\% \text{ CI } [0.249-0.789], t = 3.887, p < .001$).

When experience was included in the model, in addition to age, the model was again significant, $F(2, 27) = 10.877, p < .001, R^2 = .446$: Both older children ($\beta = 0.631, 95\% \text{ CI } [0.323-0.939], t = 4.207, p < .001$) and children with more experience ($\beta = 0.477, 95\% \text{ CI } [0.169-0.785], t = 3.18, p = .004$) were more likely to receive a response to their questions. Adding an Age \times Experience interaction did not improve the model fit ($p = .125$).

6.3 | Children's questions

We first examined the initial questions children asked, which would have not been influenced by their following interaction with Alexa. Among children's first questions, 56% were information-seeking, 30% agent related, 7% asker related, and 7% entertainment-seeking. These proportions were significantly different ($\chi^2(3) = 27.977, p < .001$). In particular, the proportion of information-seeking first questions was significantly higher than that of agent-related ($z = 2.396, p = .016$), asker-related ($z = 4.879, p < .001$), and entertainment-seeking questions ($z = 4.879, p < .001$). The proportion of agent-related questions was also significantly higher than of asker-related and entertainment-seeking questions ($z = 2.771, p = .005$ for both). There was no difference between the proportions of asker-related and entertainment-seeking questions ($z = 0, p > .999$).

We then calculated for each child the overall proportion of questions asked belonging to each of the four question types. On average, children's questions were 54% information-seeking, 31% agent-related, 8% entertainment-seeking, and 7% asker-related. A Friedman test revealed a significant difference in question types, $\chi^2(3) = 34.192, p < .001$. Post-hoc Durbin–Canover tests revealed a significant difference between each type, all p values $< .009$, except for the difference between entertainment-seeking and asker-related questions ($p = .567$). Therefore, children asked information-seeking questions the most, followed by agent-related questions, and they asked entertainment and asker-related questions equally the least.

To study the effects of age and experience on the type of questions children asked, we ran a series of linear regression models for each type of question, first with age as the only predictor, and then with age, experience, and their interaction. None of the models was significant (all p values $> .05$), with the only exception being the model predicting the proportion of asker-related questions with age, experience, and their interaction, $F(3, 26) = 3.561, p = .028, R^2 = .291$, in which the interaction Age \times Experience was significant ($\beta = 0.361, 95\% \text{ CI } [0.025-0.697], t = 2.206, p = .036$): Among older children, who asked lower proportions of asker-related questions overall, experience did not make much of a difference. However, among younger children, those who were less experienced asked a higher proportion of asker-related questions compared to those who were more experienced.

We further explored the relationship between the abilities children attribute to Alexa and the types of questions they asked, yet Kendall's Tau-b analyses did not reveal any significant associations (all p values $> .05$).

6.4 | Voice assistant's responses

For each child, we calculated the proportions of responses they received by type. On average, 52% of the responses received were proper, 22% unknowing, 17% improper, and 9% demanding, which were overall significantly different according to the Friedman test, $\chi^2(3) = 37.747, p < .001$. Post-hoc Durbin–Canover tests revealed that proportions of all categories were significantly different from each other (all p values $< .023$), except for the difference between improper and unknowing responses ($p = .955$). That is, most responses children received were proper, followed by a similar proportion of improper and unknowing responses, and with demanding responses being the least frequent. As before, we ran a series of linear regression models for each type of response, first with age as the only predictor, and then with age, experience, and their interaction of age and experience. None of the models were significant (all model p values $> .05$). The distributions of the proportion of proper responses children received across age can be seen in Figure 3.

We next analyzed the relationship between the type of questions children asked, and the type of responses they received in return. As the proportion of information-seeking questions increased, proportion of demanding responses decreased ($r = -.38, p = .012$), suggesting

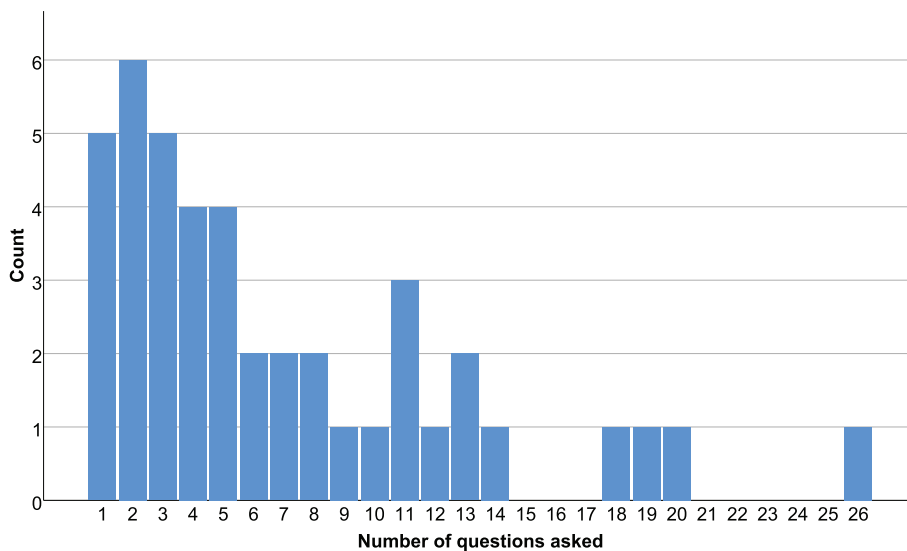


FIGURE 2 Histogram of the total number of questions children asked

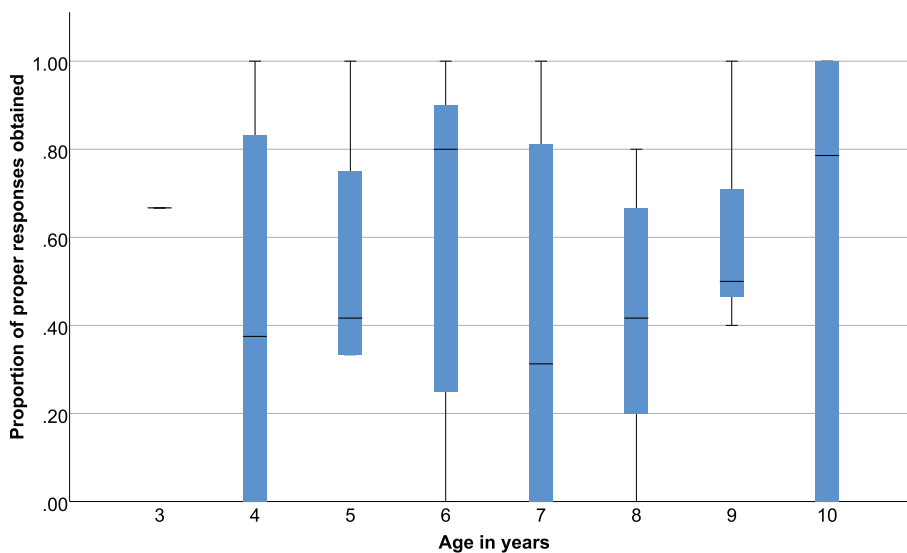


FIGURE 3 Boxplot of the proportion of proper responses children received from the voice assistant across age groups

that Alexa gave more direct answers instead of witty answers to information-seeking questions. On the contrary, as the proportion of agent-related questions increased, so did the proportion of demanding responses ($r = .543$, $p < .001$), suggesting that Alexa tried to be witty and gave indirect answers to personal questions. Last, as the proportion of asker-related questions increased, the proportion of proper responses decreased ($r = -.396$, $p = .009$) and unknowing responses increased ($r = .409$, $p = .006$). Not surprisingly, Alexa was not able to answer questions about children's lives and environments.

6.5 | Children's follow-up behavior

For the following analyses, in line with previous work (e.g., Frazier et al., 2009; Kurkul & Corriveau, 2018), we considered each question-response interaction individually. We excluded from the analyses:

(a) interactions where the question did not receive a response from Alexa, (b) interactions where responses included a suggestion (e.g., "You could ask me..."), as in those cases it was not the child but Alexa reorienting the conversation, and (c) children's last questions. As explained above, children's reactions were coded as binary (0 = no change, 1 = change) for each of the four types considered (i.e., change in question type, general content, specific content, question function). We then performed a series of logistic generalized linear mixed models (GLMM) for each type of reaction, to test whether age and experience with VAs would predict change, allowing for a random intercept for each child.

First, we analyzed how children proceeded in general, without taking the type of response they received into account, entering age as the only predictor into GLMMs. This revealed a significant effect of age only for changes in function, $\chi^2(1) = 3.874$, $p = .049$: Older children were more likely to change the function of their questions throughout the interaction (OR = 1.74, 95% CI [1.002-3.021],

$z = 1.968, p = .049$). GLMMs for question type, general and specific content changes were not predicted by age (all p values $>.05$).

To study the effects of experience in addition to age, we ran the same GLMMs for each type of reaction first with age and experience as predictors, then with age, experience and the interaction of age and experience. None of the models with age and experience revealed significant effects (all p values $>.05$). When added, the interaction term was significant for general and specific content changes, yet these models were overfitted and there was not any random variance, so we ran the same models using logistic regression without the random intercepts and reproduced the same effects. The logistic regression models predicting change in general content ($\chi^2(3) = 10.06, p = .018$) and specific content were significant ($\chi^2(3) = 13.637, p = .003$). In particular, the interaction of age and experience was significant both for general content (OR = 0.979, 95% CI [0.962–0.997], $z = -2.295, p = .022$) and for specific content changes (OR = 0.982, 95% CI [0.967–0.997], $z = -2.301, p = .021$). More specifically, experience did not make much of a difference among younger children, who tended to stay on-topic, as did older and more experienced children. However, older and less experienced children were more likely to change topic across questions.

Second, we analyzed how children reacted to the responses they received. To that end, we dummy-coded the type of Alexa's responses, that is, we coded proper responses as 1 versus non-proper responses, coding together improper, demanding, and unknowing responses as 0. We initially ran separate GLMMs for each type of reaction with age and response type (proper vs. non-proper) as predictors, which did not reveal any significant effects (all p values $>.05$). We then fitted the same models by further adding the interaction between age and response type as predictors to study how children's reactions to Alexa's responses change by age. This revealed a significant interaction effect on children's change of question type, $\chi^2(1) = 4.994, p = .025$: Compared to younger children, older children were more likely to change the type of their next question after they received a non-proper response, whereas there was no age difference for proper responses, when all children tended to keep asking the same type of question (OR = 0.247, 95% CI [0.072–0.842], $z = -2.235, p = .025$).

We then ran the same GLMMs by adding experience, hence predicting type of reaction by experience in addition to age and response type. This revealed a significant effect of response type on children's change of function, $\chi^2(1) = 8.038, p = .005$. Controlling for their age and experience, children were more likely to change the function of their next question upon receiving a non-proper response than a proper response (OR = 0.117, 95% CI [0.026–0.515], $z = -2.835, p = .005$).

Finally, to assess if the effect of response type depends on the children's age, we ran the same GLMMs by further adding an interaction term for age and response type, hence predicting type of reaction by age, response type, experience, and the interaction of age and response type. This consolidated our earlier findings for question type, showing that even controlling for experience, there is a significant interaction between age and response type ($\chi^2(1) = 4.018, p = .045$),

and that older children were more likely to change the type of their next question upon receiving a non-proper response compared to younger children (OR = 0.11, 95% CI [0.013–0.952], $z = -2.004, p = .045$). We also found a similar pattern for the changes children make in the general content of their questions, $\chi^2(1) = 6.014, p = .014$: Similar to question type, older children were also more likely to change the general content of their next questions upon receiving a non-proper response (OR = 0.175, 95% CI [0.044–0.705], $z = -2.452, p = .014$).

7 | DISCUSSION

This study examined the adaptiveness of children's inquiry behavior with VAs, exploring whether and how 3- to 10-year-old children adapt the type of questions they ask of Amazon Alexa based on the answers previously obtained. In line with previous research (e.g., Lovato et al., 2019; Lovato & Piper, 2015), our results showed that, regardless of age, the majority of questions children asked VAs were seeking factual information, which suggests that they perceive these devices as a powerful source of information. However, we found that the type of questions children asked also depended on their age and familiarity with VAs. Children who are younger and less familiar with VAs were more likely to ask questions about themselves and their environment (e.g., "What is my sister's name?"). Our results also demonstrated that with age, children were more likely to adapt their inquiry behavior to the responses received, changing the topic and type of the questions asked upon receiving irrelevant or uninformative responses. This suggests that, as they get older, children become more sensitive to the informativeness of artificial agents, growing into more sophisticated VA users. Below, we discuss our main findings in more detail and provide suggestions for future research.

First, we found that children ask information-seeking questions more often than other types of questions. This pattern is consistent with previous studies (Lovato et al., 2019; Lovato & Piper, 2015), and we further demonstrated that children use VAs to seek factual information more than for other purposes, and that this finding holds for children of a broader age range than previously suggested. Regardless of their age, children seem to perceive VAs mainly as a source of information. This stands in stark contrast with children's usage of and intuitions about other digital devices such as tablets and smartphones, which they deem more as a source of entertainment than information (Eisen & Lillard, 2017). This may not be too surprising; touchscreen devices are currently more multifunctional than VAs, and therefore can be used for other purposes than searching for information such as calling other people, watching videos, playing games or taking photos. However, children younger than 8 years of age have also been found to be generally hesitant in trusting Internet-based sources for learning (Danovitch, 2019). In this sense, although VAs are relatively easy to control and be understood by children, as they operate by receiving verbal questions and commands, do children *trust* the information received by VAs more than other Internet-based sources? Do they remember, filter, and use the information received in a different way?

Children are selective and critical in trusting digital sources such as the Internet and social robots (e.g., Brink & Wellman, 2020; Wang et al., 2019), and our finding that children seek factual information from VAs does not necessarily mean that they trust the information to apply to their daily lives. Furthermore, the prevalence of information-seeking questions in our data may have been influenced by the experimental setting: children may use VAs for a wider variety of purposes at home (Beneteau et al., 2020), as opposed to a zoo where listening to music or turning on the lights would make no sense. Further naturalistic yet quantitative research is needed to examine if our findings are echoed in children's daily and home-based use of VAs.

We found that older children and younger children with previous VA experience asked fewer asker-related questions (i.e., about children's own life and environment) compared to younger, less experienced children. Previous studies observed that preschoolers often ask such questions to VAs at home, and they attributed this tendency to young children's developing theory of mind skills, or more generally, to their difficulty in understanding what one can and cannot know and answer (Lovato & Piper, 2019; Sciuto et al., 2018). Indeed, as they get older, children begin to differentiate between the types of questions that can or cannot be answered by machines (Danovitch & Keil, 2008). Younger children who are familiar with VAs, and older children with more general experience with technology, may already be able to grasp the constraints and limitations of these devices, whereas those who are unfamiliar with them may want to explore and probe their limits (Druga et al., 2017; Yuan et al., 2019).

Second, we found that younger children, whether experienced or not, and older children with more experience, were more likely to stick to the same themes and topics when asking questions. In contrast, older yet less experienced children were more likely to change topic during the interaction. On the one hand, younger children may just have been confined to a small selection of conversational topics by their limited conceptual knowledge (Ronfard et al., 2018). On the other hand, among the older children, with more general knowledge, those with experience might have already known what Alexa can or cannot answer, deciding to stick with those topics they felt could offer a meaningful and successful interaction. In contrast, those who were not as experienced might have asked questions about a wider variety of topics to try out and explore the device. This might also explain why we found that, compared to younger children, older children were more likely to vary the purpose of their questions, for example following a *why* question with a *what* question. In general, older children may be more prone to test the diversity and boundaries of VAs' knowledge just because their wider background allows them to explore further.

Third, we found that children were sensitive overall to the responses Alexa gave, and older children were even more sensitive than younger children. Children were more likely to ask a question with a different purpose upon receiving a non-proper response—irrelevant, inadequate, or complicated—than after a proper response. Yet, older children were also more likely, in response to a non-proper response, to change the type of question (e.g., following an

information-seeking question with a personal question) and the question topic (e.g., following a question about nature with a question about language).

Previous research suggests that when children receive unsatisfactory responses to their explanation-seeking, causal questions (e.g., *why* and *how* questions), they tend to persist and repeat their original questions, hence staying on-topic, to obtain the desired information (Frazier et al., 2009). In contrast, older children in our study changed topic, along with the question type, upon receiving an unsatisfactory response, possibly because the vast majority of their questions were fact-seeking (e.g., *what*, *when*, or *where* questions), and not explanation-seeking (note that in our dataset only 16 of the questions children asked were explanation-seeking). In our study, older children, more sensitive and critical about the informativeness of a response compared to younger children (Ronfard et al., 2018), might have concluded that "if Alexa doesn't know, she doesn't know," and therefore changed the course of their interaction. This is also in line with previous research showing that, in contrast to causal questions, children tended *not* to react to the answers received by adults to fact-based questions, whether satisfactory or not (Kurul & Corriveau, 2018). Future work may investigate if children's reactions to VAs' responses for fact-based versus causal questions differ. Lovato et al. (2019) found that VAs are most often not able to answer children's causal questions, yet they did not examine how children react in such instances. Furthermore, as mentioned earlier, children are selective in trusting digital tools, and they rely on accurate computers and robots over inaccurate ones from a young age (Brink & Wellman, 2020; Danovitch & Alzhabi, 2013). Our finding that older children veer off course upon receiving irrelevant and uninformative responses may have implications for their overall trust in VAs as a source of information.

Our work contributes to the growing literature trying to better comprehend children's understanding of and interactions with VAs, highlighting the unique potential of VAs for children's learning, and raising several questions that need to be answered before concluding that these devices are effective for educational purposes. First, what intuitions and assumptions do children have about the capacities and characteristics of VAs, what do they think VAs can do or feel, what do they explicitly expect from their interactions with them, and how do these assumptions and expectations influence their interactions? There is a considerable amount of research on children's perceptions of VAs (e.g., Druga et al., 2017; Festerling & Siraj, 2020; Xu & Warschauer, 2020), yet we do not yet know if and how these assumptions are linked with their learning and interaction behavior, with VAs and with interactive technology more generally, or how children's experience changes their assumptions about these devices. In this study, we have collected exploratory data on children's assumptions about Alexa's abilities, but we did not find any significant relationship with children's inquiry behavior, possibly due to the many missing data points. It is especially important to further study this relationship, possibly in a longitudinal study and using behavioral rather than self-report measures, as this has the potential to inform the development of effective interventions aimed at fostering successful educational experiences with

VAs. Second, future studies should look more closely at the impact of children's familiarity with VAs on their inquiry patterns and adaptive information search. Our data revealed that experience with VAs does indeed have an impact on children's interactions with these devices, but the effects of familiarity with technology more generally, and with VAs in particular, on children's overall learning habits, effectiveness and success are still to be examined. In this sense, interventional and cross-cultural studies can provide crucial insights, and because technological globalization is rapidly closing the "technology gap" between societies, this research question has to be urgently addressed. Last but not least, it is crucial to contextualize children's interaction with VAs within the broader social learning framework, comparing their learning and inquiry behavior with different virtual (VAs and social robots) and physical agents (peers, parents and teachers). Technological advancements create new sources from whom children can learn: when they have a question, in addition to asking their parents or friends, children can now google it on a computer, use a mobile app on a tablet, ask Alexa, or even a robot. We must understand when, why, and how children decide to interact with VAs in this broader landscape.

To conclude, children are avid question askers. Whereas adults around them may not be always available or eager to answer their questions, voice assistants offer an attractive alternative to quench children's thirst for information, as they are becoming ubiquitous in their daily lives. Understanding children's interactions with artificial agents will help us reap the benefits and mitigate potential dangers of children's interaction with Artificial Intelligence, and inform the development of effective conversational agents for children.

ACKNOWLEDGMENT

We thank Luise Mehner for her invaluable support with data collection and coding, and the Berlin Zoo for providing us with a safe and welcoming testing environment. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/hbe2.270>.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

ORCID

Cansu Oranç  <https://orcid.org/0000-0002-2341-2038>

Azzurra Ruggeri  <https://orcid.org/0000-0002-0839-1929>

REFERENCES

- Beneteau, E., Boone, A., Wu, Y., Kientz, J. A., Yip, J., & Hiniker, A. (2020). Parenting with Alexa: Exploring the introduction of smart speakers on family dynamics. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3313831.3376344>
- Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019). Communication breakdowns between families and Alexa. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3290605.3300473>
- Brink, K. A., & Wellman, H. M. (2020). Robot teachers for children? Young children trust robots depending on their perceived accuracy and agency. *Developmental Psychology*, 56(7), 1268–1277. <https://doi.org/10.1037/dev0000884>
- Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (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. <https://doi.org/10.1145/3202185.3202749>
- Chouinard, M. M., Harris, P. L., & Maratsos, M. P. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, 72, 1–112.
- Cognilytica. (2018). Voice Assistant Benchmark 1.0 Report. <https://www.cognilytica.com/2018/07/31/report-voice-assistant-benchmark-1-0-2018/>
- Danovitch, J. H. (2019). Growing up with Google: How children's understanding and use of internet-based devices relates to cognitive development. *Human Behavior and Emerging Technologies*, 1(2), 81–90. <https://doi.org/10.1002/hbe2.142>
- Danovitch, J. H., & Alzahabi, R. (2013). Children show selective trust in technological informants. *Journal of Cognition and Development*, 14(3), 499–513. <https://doi.org/10.1080/15248372.2012.689391>
- Danovitch, J. H., & Keil, F. C. (2008). Young humans: The role of emotions in children's evaluation of moral reasoning abilities. *Developmental Science*, 11(1), 33–39. <https://doi.org/10.1111/j.1467-7687.2007.00657.x>
- Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google is it ok if I eat you?" Initial explorations in child-agent interaction. Proceedings of the 2017 Conference on Interaction Design and Children. <https://doi.org/10.1145/3078072.3084330>
- Eisen, S., & Lillard, A. S. (2017). Young children's thinking about touchscreens versus other media in the US. *Journal of Children and Media*, 11, 167–179. <https://doi.org/10.1080/17482798.2016.1254095>
- Festerling, J., & Siraj, I. (2020). Alexa, what are you? Exploring primary school children's ontological perceptions of digital voice assistants in open interactions. *Human Development*, 64(1), 26–43. <https://doi.org/10.1159/000508499>
- Frazier, B. N., Gelman, S. A., & Wellman, H. M. (2009). Preschoolers' search for explanatory information within adult-child conversation. *Child Development*, 80(6), 1592–1611. <https://doi.org/10.1111/j.1467-8624.2009.01356.x>
- Frazier, B. N., Gelman, S. A., & Wellman, H. M. (2016). Young children prefer and remember satisfying explanations. *Journal of Cognition and Development*, 17(5), 718–736. <https://doi.org/10.1080/15248372.2015.1098649>
- Herwig, J. E. (1982). Effects of age, stimuli, and category recognition factors in children's inquiry behavior. *Journal of Experimental Child Psychology*, 33(2), 196–206. [https://doi.org/10.1016/0022-0965\(82\)90015-7](https://doi.org/10.1016/0022-0965(82)90015-7)
- The Jamovi Project. (2020). *jamovi* (Version 1.2) [Computer Software]. <https://www.jamovi.org>
- Jones, A., Swaboda, N., & Ruggeri, A. (2020). Developmental changes in question asking. In L. Butler, S. Ronfard, & K. Corriveau (Eds.), *The questioning child: Insights from psychology & education*. Cambridge University Press.
- Kurkul, K. E., & Corriveau, K. H. (2018). Question, explanation, follow-up: A mechanism for learning from others? *Child Development*, 89(1), 280–294. <https://doi.org/10.1111/cdev.12726>

- Lovato, S., & Piper, A. M. (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. <https://doi.org/10.1145/2771839.2771910>
- Lovato, S. B., & Piper, A. M. (2019). Young children and voice search: What we know from human-computer interaction research. *Frontiers in Psychology*, 10, 1–5. <https://doi.org/10.3389/fpsyg.2019.00008>
- Lovato, S., Piper, A. M., & Wartella, E. A. (2019). "Hey Google, do unicorns exist?" Conversational agents as a path to answers to children's questions. Proceedings of the 18th ACM International Conference on Interaction Design and Children. <https://doi.org/10.1145/3311927.3323150>
- Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. In J. S. Bruner, R. R. Oliver, & P. M. Greenfield (Eds.), *Studies in cognitive growth*. Wiley.
- Ronfard, S., Zambrana, I. M., Hermansen, T. K., & Kelemen, D. (2018). Question-asking in childhood: A review of the literature and a framework for understanding its development. *Developmental Review*, 49, 101–120. <https://doi.org/10.1016/j.dr.2018.05.002>
- Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, 143, 203–216. <https://doi.org/10.1016/j.cognition.2015.07.004>
- Ruggeri, A., & Feufel, M. A. (2015). How middle-level objects facilitate asking efficient questions in a categorization task. *Frontiers in Psychology*, 6(918), 1–13. <https://doi.org/10.3389/fpsyg.2015.00918>
- Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, 52(12), 2159–2173. <https://doi.org/10.1037/dev0000240>
- Ruggeri, A., Sim, Z. L., & Xu, F. (2017). "Why is Toma late to school again?" Preschoolers identify the most informative questions. *Developmental Psychology*, 53(9), 1620–1632. <https://doi.org/10.1037/dev0000340>
- Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "Hey Alexa, what's up?" Studies of in-home conversational agent usage. Proceedings of the 2018 Designing Interactive Systems Conference. <https://doi.org/10.1145/3196709.3196772>
- Terzopoulos, G., & Satratzemi, M. (2020). Voice assistants and smart speakers in everyday life and in education. *Informatics in Education*, 19(3), 473–490. <https://doi.org/10.15388/infedu.2020.21>
- Wang, F., Tong, Y., & Danovitch, J. (2019). Who do I believe? Children's epistemic trust in internet, teacher, and peer informants. *Cognitive Development*, 50, 248–260. <https://doi.org/10.1016/j.cogdev.2019.05.006>
- Xu, Y., & Warschauer, M. (2020). What are you talking to?: Understanding Children's perceptions of conversational agents. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3313831.3376416>
- Yuan, Y., Thompson, S., Watson, K., Chase, A., Senthilkumar, A., Brush, A. B., & Yarosh, S. (2019). Speech interface reformulations and voice assistant personification preferences of children and parents. *International Journal of Child-Computer Interaction*, 21, 77–88. <https://doi.org/10.1016/j.ijcci.2019.04.005>

AUTHOR BIOGRAPHIES



Cansu Oranç is a postdoctoral researcher at the Max Planck Research Group iSearch (Information Search, Ecological and Active Learning Research with Children) at the Max Planck Institute for Human Development in Berlin (MPIB). She received her Ph.D. in Psychology from Koç University in Istanbul, Turkey. Her research focuses on children's learning with technology. She is specifically interested in children's intuitions about and interactions with intelligent artificial agents, such as social robots and voice assistants, in formal and informal learning settings.



Azzurra Ruggeri is interested in how children actively seek information and explore to learn about the world. She leads the Max Planck Research Group iSearch (Information Search, Ecological and Active Learning Research with Children) at the Max Planck Institute for Human Development in Berlin (MPIB). Since

2017, she is also Professor of Cognitive and Developmental Psychology at the Technical University of Munich, and faculty of the International Max Planck Research School (IMPRS) on the Life Course. She worked at the MPIB from 2009 to 2013 as predoctoral fellow of the Max Planck Graduate program IMPRS Uncertainty, and in 2012 obtained a PhD in Psychology at the Humboldt University in Berlin, and a PhD in Cognitive Science at the University of Siena, Italy. As a Marie Curie Outgoing postdoctoral fellow, she spent two years at the University of California, Berkeley.

How to cite this article: Oranç, C., & Ruggeri, A. (2021).

"Alexa, let me ask you something different" Children's adaptive information search with voice assistants. *Human Behavior and Emerging Technologies*, 3(4), 595–605. <https://doi.org/10.1002/hbe2.270>

APPENDIX**Abilities Interview Questions**

Measured aspect	Question
Function	What is Alexa good for?
Intellectual capacity	Do you think Alexa is smart or not?
Truthfulness	Do you think Alexa tells the truth or not?
Emotional capacity	Do you think Alexa can feel happy or sad or not?
Social capacity	Do you think Alexa can be your friend or not?
Biological property	Do you think Alexa can get older or not?
Cognitive ability	Do you think Alexa is good at solving problems or not?